



T.C.
DÜZCE ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

**TIBBİ ARAŞTIRMALARDA DESTEK VEKTÖR
MAKİNELERİNİN KULLANIMI**

Özge AKŞEHİRLİ
YÜKSEK LİSANS TEZİ

BİYOİSTATİSTİK ve TIBBİ BİLİŞİM ANABİLİM DALI

DANIŞMAN
Prof. Dr. Handan ANKARALI

Düzce, 2012



T.C.
DÜZCE ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

**TIBBİ ARAŞTIRMALARDA DESTEK VEKTÖR
MAKİNELERİNİN KULLANIMI**

Özge AKŞEHİRLİ
YÜKSEK LİSANS TEZİ

BİYOİSTATİSTİK ve TIBBİ BİLİŞİM ANABİLİM DALI

DANIŞMAN
Prof. Dr. Handan ANKARALI

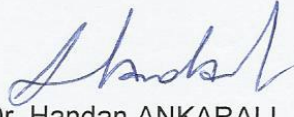
Düzce, 2012

KABUL VE ONAY

Biyostatistik ve Tıbbi Bilişim ABD Yüksek Lisans Programı Çerçevesinde yürütülmüş olan
“Tıbbi Araştırmalarda Destek Vektör Makinelerinin Kullanımı”
adlı çalışma, aşağıdaki jüri tarafından Yüksek Lisans Tezi olarak kabul edilmiştir.

Tarihi: 08 / 05 / 2012

TEZ SINAV JÜRİSİ

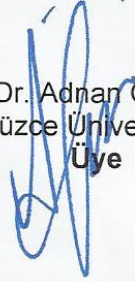


Prof. Dr. Handan ANKARALI
Düzce Üniversitesi
Başkan

Doç. Dr. Atilla Senih MAYDA
Düzce Üniversitesi
Üye



Doç. Dr. Adnan ÖZÇETİN
Düzce Üniversitesi
Üye



Yukarıdaki Tez, Yönetim Kurulunun 14 / 06 / 2012 tarih ve 28 sayılı kararı ile kabul edilmiştir.

Yrd. Doç. Dr. Talat BAHÇEBAŞI
Sağlık Bilimleri Enstitü Müdürü



BEYAN

Bu tez çalışmasının kendi çalışmam olduğunu, tezin planlanmasından yazımına kadar bütün aşamalarda etik dışı davranışımın olmadığını, bu tezdeki bütün bilgileri akademik ve etik kurallar içinde elde ettiğimi, bu tez çalışmasıyla elde edilmeyen bütün bilgi ve yorumlara kaynak gösterdiğimi ve bu kaynakları da kaynaklar listesine aldığımı, yine bu tezin çalışılması ve yazımı sırasında patent ve telif haklarını ihlal edici bir davranışımın olmadığı beyan ederim.

09.06.2012


Özge AKŞEHİRLİ

TEŐEKKÖR

Yüksek lisans eğitimin süresince bilgi ve tecrübesiyle bana yol gösteren, katkı ve manevi desteğini esirgemeyen tez danışmanım, değerli hocam Prof. Dr. Handan ANKARALI'ya,

Tezimin düzenlenmesinde ve hazırlanmasında çok büyük desteđi olan sevgili arkadaşım Arş Gör. Özge KAPISIZ'a, tez çalışmalarım boyunca manevi destekleri ile yanımda olan arkadaşlarım Öğr. Gör. S. Seda BAPOĞLU DÜMENCİ'ye, Selva KÜTÜK'e ve çalışma arkadaşım Duygu AYDIN'a,

Çalışmamın her aşamasında beni gönülden destekleyen sevgili aileme ve her türlü desteđiyle daima yanımda olan eşim Aras AKŐEHİRLİ'ye

Teşekkür ederim

Arş. Gör. Özge AKŐEHİRLİ

İÇİNDEKİLER

KISALTMALAR.....	İV
SİMGELER	V
ÖZET	1
ABSTRACT	2
1.GİRİŞ VE AMAÇ	3
2. GENEL BİLGİLER.....	7
2.1. Veri Madenciliği.....	7
2.1.1. Veri madenciliğinin tanımı ve temel bilgiler	7
2.1.2 Makine öğrenmesi	9
2.1.2.1. Makine öğrenmesi türleri	12
2.1.3. Veri madenciliği yöntemlerine genel bakış	14
2.2. İstatistiksel Öğrenme Teorisi (Vapnik-Chervonenkis Teorisi)	23
2.2.1. Beklenen risk ve deneysel risk	26
2.2.2. Vapnik-Chervonenkis (VC) boyutu.....	28
2.2.2.1. Aşırı uyum (overfitting) ve eksik uyum (underfitting) kavramları	31
2.2.3. Yapısal risk minimizasyonu	33
2.3. Destek Vektör Makineleri	35
2.3.1. Destek vektör makinelerinin tanımı	35
2.3.2. Destek vektör makinelerinin tarihçesi	38
2.3.3. Destek vektör makinelerinin uygulama alanları.....	39
2.3.4. Destek vektör sınıflaması.....	39
2.3.4.1. Doğrusal destek vektör makineleri ile sınıflama	39
2.3.4.2. Doğrusal olmayan destek vektör makineleri ile sınıflama.....	51

2.3.4.3. Çekirdek fonksiyonlar (kernel functions)	58
2.3.4.4. DVM’de kullanılan çekirdek fonksiyonları	61
2.3.4.5. Hiperparametrelerin seçimi	63
2.3.5. Destek vektör regresyonu	72
2.3.6. Destek vektör makinelerinin avantajları	74
2.3.7. Destek vektör makinelerinin dezavantajları	74
3. GEREÇ VE YÖNTEM.....	75
3.1. Veriler	75
3.2. Yöntem.....	79
4. BULGULAR	85
5. TARTIŞMA VE SONUÇ	118
6. KAYNAKLAR.....	122

KISALTMALAR

DVM	Destek Vektör Makinesi
DVS	Destek Vektör Sınıflaması
DVR	Destek Vektör Regresyonu
EM	Expectation Maximization
VTBK	Veri Tabanlarında Bilgi Keşfi
RF	Random Forest
YSA	Yapay Sinir Ağları
GA	Genetik Algoritmalar
VC	Vapnik-Chervonenkis
DRM	Deneysel Risk Minimizasyonu
YRM	Yapısal Risk Minimizasyonu
OKT	Optik Karakter Tanıma
DV	Destek Vektörleri
KKT	Karush-Kuhn-Tucker
RTF	Radyal Tabanlı Fonksiyon
GYA	Gece Yeme Anketi
BSQ	Vücut Şekli Anketi
SCL-90	Semptom Tarama Listesi
GYS	Gece Yeme Sendromu

SİMGELER

x_i	Giriş Verileri
y_i	Giriş Verilerine Uyan Çıkış Etiketleri
$f(x, \alpha)$	Karar Fonksiyonu
$L(y, f(x, \alpha))$	Kayıp Fonksiyon
$R(\alpha)$	Beklenen Risk
$P(x, y)$	Olasılık Dağılımı
R_{amp}	Ampirik Risk
l	Eğitim Verisi
h	VC Boyutu
n	Eğitim Verilerinin Sayısı
$G(n)$	Yükselme Fonksiyonu
$\Phi(h, n, \delta)$	VC Güvenilirliği
$L(x, \alpha)$	Langrangian Fonksiyonu
α_i	Langrange Çarpanları
L_P	Primal Lagrange İfadesi
L_d	Dual Lagrange İfadesi
w	Ağırlık Terimi
b	Bias Terimi
M	Marjin (Herhangi Bir Nokta ile Hiperdüzlem Arasındaki Mesafe)
C	Hata Maliyeti
ξ_i	Gevşek Değişken
$\Phi(x)$	Haritalama (Dönüştürme) Fonksiyonu
$K(x_i, x)$	Çekirdek Fonksiyon

ŞEKİLLER LİSTESİ

Şekil 2.1: Veri Madenciliğinin İşleyişi	7
Şekil 2.2: Veri Madenciliğinin Diğer Disiplinlerle İlişkisi.....	9
Şekil 2.3: Eğitici Öğrenme Yönteminin Yapısı	13
Şekil 2.4: Eğitici Öğrenme Yönteminin Yapısı.....	14
Şekil 2.5: Veri Madenciliği Metotları	15
Şekil 2.6: Sınıflamanın Görselleştirilmesi	16
Şekil 2.7: Karar Ağacı Şeması	18
Şekil 2.8: K-En Yakın Komşuluğu Yönteminde, $k=3$ için Yeni Verinin Sınıflandırılması	21
Şekil 2.9: Hataların Modellenmesi	25
Şekil 2.10: Beklenen ve Deneysel Risk	28
Şekil 2.11: R^2 Uzayında Tanımlanmış A, B, C Noktaları için VC Boyutu Örneği.....	29
Şekil 2.12: R^2 Uzayında Dört Noktanın Ayrımı	30
Şekil 2.13: Yükselme Fonksiyonu Grafiği	31
Şekil 2.14: Üç Farklı Model Kullanılarak Oluşturulmuş Bir İkili Sınıflandırma Örneği	31
Şekil 2.15: Aşırı Uyumun Meydana Gelme Sebepleri	32
Şekil 2.16: Beklenen Risk Fonksiyonunun Yapısı	35
Şekil 2.17: Doğrusal Olarak Ayrılabilen İki Sınıflı Sınıflama Problemi.....	40
Şekil 2.18: Verilerin Tamamının Ayrılabilmediği Durum İçin Doğrusal Ayrıcı Hiperdüzlem.....	41
Şekil 2.19: R^2 'de Noktaların Görünümü.....	45
Şekil 2.20: Yeşil Üçgenlerle Gösterilen Destek Vektörler	46
Şekil 2.21: Bulunan $\alpha_1 = -8$, $\alpha_2 = 6.5$ ve $\alpha_3 = -1.5$ Değerleri İçin Ayrıcı Hiperdüzlem.	47
Şekil 2.22: Verilerin Tamamının Doğrusal Olarak Ayrılmadığı Durum	48
Şekil 2.23: Verilerin Tamamının Doğrusal Olarak Ayrılmadığı Durum İçin Optimum Ayrıcı Hiperdüzlem	49
Şekil 2.24: Doğrusal Olarak Ayrılmayan Verilerin Farklı Boyutlardaki Uzaylara Aktarılması.....	52
Şekil 2.25: Doğrusal Olarak Ayrılmayan Verilerin, Doğrusal Olmayan Haritalama Fonksiyonları ile Daha Yüksek Boyutlu Belirleyici Nitelik Uzayına Taşınması	52

Şekil 2.26: R^2 'de Doğrusal Olarak Ayrılamayan Noktaların Görünümü	54
Şekil 2.27: Belirleyici Nitelik Uzayına Taşınan Verilerin Görünüşü	55
Şekil 2.28: Yeşil Üçgenlerle Gösterilen Destek Vektörler	56
Şekil 2.29: Bulunan $\alpha_1 = -11$, $\alpha_2 = 4.25$ Değerleri için Ayırıcı Hiperdüzlem.....	57
Şekil 2.30: Verilerin Çekirdek Fonksiyonu ile Belirleyici Nitelik Uzayına Taşınması .	60
Şekil 2.31: Örnek 3 için Karar Fonksiyonu	69
Şekil 2.32: Verilen XOR Verilerinin Analitik Düzlemde Gösterimi.....	70
Şekil 2. 33: Linear ϵ Kayıp Fonksiyonu	73
Şekil 4. 1: On Yedi Değişken Yardımıyla Kurulan Lojistik Regresyon Modelinin	
Yükselme ve Kazanç Grafikleri	90
Şekil 4.2: Değişkenlerin Önem Dereceleri	91
Şekil 4.3: On Yedi Değişken Yardımıyla Kurulan Lojistik Regresyon Modelinin	
Yükselme ve Kazanç Grafikleri	93
Şekil 4 4: On Yedi Belirleyici Nitelik Kullanılarak Elde Edilmiş Eğitim ve Test Verileri	
için Kesim Noktası Grafikleri	98
Şekil 4.5: On Yedi Belirleyici Nitelik Kullanılarak Elde Edilmiş Eğitim ve Test Verileri	
için Olasılık Düzenleme Grafikleri	100
Şekil 4.6: On Yedi Belirleyici Nitelik Kullanılarak Elde Edilmiş Eğitim ve Test Verileri	
için Duyarlılık ve Seçicilik Grafikleri	102
Şekil 4.7: On Yedi Belirleyici Nitelik Kullanılarak Elde Edilmiş Eğitim ve Test Verileri	
için Roc Eğrisi Grafikleri	102
Şekil 4.8: DVM Yöntemi ile 17 Belirleyici Nitelik Kullanılarak Elde Edilen Yükselme	
ve Kazanç Grafikleri	103
Şekil 4.9:DVM Yönteminde Kullanılan 17 Belirleyici Nitelikten Önemli Bulunanlar	104
Şekil 4.10: Tek Değişkenli (Univariate) Analiz, Lojistik Regresyon ve DVM Yöntemi	
ile Anlamlı Bulunan Değişkenler	105
Şekil 4.11: Üç Belirleyici Nitelik Kullanılarak Elde Edilmiş Eğitim ve Test Verileri için	
Kesim Noktası Grafikleri	109
Şekil 4.12: Üç Belirleyici Nitelik Kullanılarak Elde Edilmiş Eğitim ve Test Verileri için	
Olasılık Düzenleme Grafikleri	110
Şekil 4.13: Üç Belirleyici Nitelik Kullanılarak Elde Edilmiş Eğitim ve Test Verileri için	
Duyarlılık ve Seçicilik Grafikleri	112

Şekil 4.14: Üç Belirleyici Nitelik Kullanılarak Elde Edilmiş Eğitim ve Test Verileri için ROC Eğrisi Grafikleri	112
Şekil 4.15: Üç Belirleyici Nitelik Kullanılarak Elde Edilmiş Yükselme ve Kazanç Grafikleri	113

TABLolar LİSTESİ

Tablo 2.1: Makine Öğrenmesi Yönteminin Uygulama Alanları	11
Tablo 2.2: Yaygın Olarak Kullanılan Çekirdek Fonksiyonlar.....	62
Tablo 2.3: Bilinmeyen C Parametresi için Maliyet Matrisi.....	64
Tablo 2.4: Maliyet Matrisi ve Verilen Bir Model Matrisinin Harf ile İfade Edilmesi ...	64
Tablo 2.5: Maliyet Matrisine Ait Değerler	65
Tablo 2.6: M_1 Modeline Ait Matris	65
Tablo 2.7: M_2 Modeline Ait Matris	66
Tablo 3.1: Modelde Kullanılan Değişkenlerin Listesi ve Özellikleri.....	79
Tablo 4.1: Sayısal Özelliklere Ait Tanımlayıcı İstatistikler	85
Tablo 4.2: Kategorik Özelliklere Ait Tanımlayıcı İstatistikler.....	86
Tablo 4.3: Gece Yeme Sendromunu Etkileyen Faktörlerin Lojistik Regresyon ile İncelenmesi	87
Tablo 4. 4: On Yedi Değişken İçeren Lojistik Regresyon Modelinin Sınıflama Başarıları	88
Tablo 4.5: On Yedi Değişken Yardımıyla Kurulan Lojistik Regresyon Modelinin Yükselme ve Kazanç Değerleri	89
Tablo 4.6: Lojistik Regresyon Modeline Alınan Değişkenlerin Önem Dereceleri	91
Tablo 4.7: Üç Değişken İçeren Lojistik Regresyon Modelinin Sınıflama Başarıları.....	92
Tablo 4.8: Üç Değişken Yardımıyla Kurulan Lojistik Regresyon Modelinin Yükselme ve Kazanç Değerleri.....	93
Tablo 4.9: On Yedi Belirleyici Nitelik Yardımıyla DVM Yönteminden Tahmin Edilen Optimum Parametre Değerleri.....	94
Tablo 4.10: On Yedi Belirleyici Nitelik Kullanılarak DVM Yöntemi ile Elde Edilmiş Kesim Noktaları	96
Tablo 4.11: On Yedi Belirleyici Nitelik Kullanılarak DVM Yöntemi ile Elde Edilmiş Olasılık Düzenlemesi	99
Tablo 4.12: On Yedi Belirleyici Nitelik Kullanılarak DVM Yöntemi ile Elde Edilmiş Eğitim ve Test Verilerinde Modelin Sınıflama Başarıları	101
Tablo 4.13: DVM Yöntemi ile 17 Belirleyici Nitelik Kullanılarak Elde Edilmiş Yükselme ve Kazanç Değerleri	103

Tablo 4.14: On Yedi Belirleyici Nitelik Kullanıldığında DVM ile Elde Edilmiş, Değişkenlerin Önem Dereceleri.....	104
Tablo 4.15: Üç Belirleyici Nitelik Kullanıldığı Durumda Parametre Tahmini Sürecinde Elde Edilen Değerler.....	106
Tablo 4.16: Üç Belirleyici Nitelik Kullanıldığında DVM Yöntemi İle Elde Edilen Kesim Noktaları.....	107
Tablo 4.17: Üç Belirleyici Nitelik Kullanıldığında Elde Edilen Olasılık Düzenlemesi.....	110
Tablo 4.18: Önemli Bulunan Üç Belirleyici Nitelik Kullanılarak DVM Yöntemi ile Elde Edilmiş Eğitim ve Test Verilerinde Modelin Sınıflama Başarıları.....	111
Tablo 4.19: DVM Yöntemi ile 3 Belirleyici Nitelik Kullanılarak Elde Edilen Yükseltme ve Kazanç Değerleri.....	113
Tablo 4.20: İncelenen Model Performanslarının Karşılaştırması.....	114
Tablo 4.21: Eğitim ve test verilerinde doğrusal çekirdek fonksiyon ve on yedi belirleyici nitelik kullanılarak elde edilen modelin tanı başarıları.....	115
Tablo 4.22: Eğitim ve Test Verilerinde Polinomial Çekirdek Fonksiyon ve On Yedi Belirleyici Nitelik Kullanılarak Elde Edilen Modelin Tanı Başarıları.....	115
Tablo 4.23: Eğitim ve Test Verilerinde Sigmoid Çekirdek Fonksiyon ve On Yedi Belirleyici Nitelik Kullanılarak Elde Edilen Modelin Tanı Başarıları.....	116
Tablo 4.24: Kullanılan 4 Farklı Çekirdek Fonksiyonu için Analiz Süreleri.....	116
Tablo 4.25: DVM Yönteminde 17 Belirleyici Nitelik ve 4 Farklı Çekirdek Fonksiyon Kullanılarak ile Elde Edilen Tahmin Başarı Göstergelerinin Karşılaştırması.....	117

ÖZET

TIBBİ ARAŞTIRMALARDA DESTEK VEKTÖR MAKİNELERİNİN KULLANIMI

Özge AKŞEHİRLİ

Yüksek Lisans Bitirme Tezi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı

Tez Danışmanı Prof. Dr. Handan ANKARALI

Mayıs 2012, 131 sayfa

Günümüz teknolojisinde bilgisayar kapasite ve hızlarının artmasına paralel olarak, çeşitli sektörlerde, depolanan verinin kapasitesi ve karmaşıklığı da artmaktadır. Bu verilerin uygun ileri tekniklerle işlenmesi sonucunda ise yaşam koşullarında baş döndürücü teknolojik değişimler gözlenmektedir. Büyük boyutlardaki verilerden anlamlı bilgiler çıkarmak amacıyla özellikle son yıllarda yoğun kullanım alanı bulan ve gelecekte ise klasik istatistik yöntemlerin yerini alacak olan *veri madenciliği yöntemleri* tercih edilmektedir. *Destek vektör makineleri (DVM)*, bu yöntemlerden birisi olup günümüzde sınıflandırma, tahmin veya yapı tanıma amaçlarıyla en çok mühendislik uygulamalarında kullanılmaktadır. Bu tez çalışmasında, veri madenciliği kavramlarının tanımlanması, DVM'nin teorik temellerinin ayrıntılı olarak açıklanması ve bu yöntemin tanı koyma amacı güden tıbbi araştırmalarda kullanımı amaçlanmıştır. Bu amaca yönelik olarak çalışmada, ilk olarak veri madenciliğinden ve veri madenciliği yöntemlerinden kısaca bahsedilmiş, daha sonra destek vektör makinelerine kapsamlı olarak değinilmiştir. Veri madenciliği yöntemlerinden biri olan DVM, veriyi sınıflandırmak veya tahmin yapmak amacıyla kullanılan, eğitici bir makine öğrenmesi yöntemidir. DVM'nin temelini, verilerin bir düzlem veya hiperdüzlem ile ayrılarak sınıflandırılması işlemi oluşturmaktadır. DVM bu işlemi, iki sınıf arasındaki marjini maksimum yaparak gerçekleştirir. Bu şekilde veri eğitildikten sonra, DVM yeni gelen veriyi doğru sınıflamayı amaçlamaktadır. Tıpta DVM özellikle, kanser morfolojisinde, tedavi başarısının ve ilgili genin belirlenmesinde, çeşitli hastalıkların teşhisinde kullanılmaktadır. Araştırmanın uygulama bölümünde, Zonguldak Karaelmas Üniversitesi Tıp Fakültesi psikiyatri polikliniğine gece yeme sendromu şikayetiyle 1-31 Ocak 2011 tarihleri arasında ayaktan başvuran ve çalışmaya katılmayı kabul eden 433 hastaya ilişkin bilgiler kullanılmıştır. Bu veriler kullanılarak, tek değişkenli analiz, lojistik regresyon analizi ve DVM yöntemleri ile tanı koymada etkili olan değişkenler incelenmiş ve sadece GYA_puan değişkeni üç yaklaşımla da etkili bulunmuştur. Ayrıca lojistik regresyon ve DVM ile doğru sınıflama başarıları incelendiğinde, her iki yöntemin de benzer sonuçlar verdiği görülmüş ancak DVM'nin üstünlükleri tartışılmıştır. Ayrıca, uygulamalarda en sık adı geçen doğrusal, polinomial, sigmoid ve radyal tabanlı fonksiyon (RTF) performansları da karşılaştırılmıştır. Bu karşılaştırmalarda 4 farklı çekirdek fonksiyonunun sınıflama performanslarının benzer sonuçlar verdiği görülmüş ancak çeşitli yönlerden RTF'nin üstünlüğünden bahsedilmiştir.

Anahtar Kelimeler: Çekirdek fonksiyonlar, Destek vektör makineleri, Sınıflama, Veri madenciliği

ABSTRACT

USING OF SUPPORT VECTOR MACHINES IN MEDICAL RESEARCH

Özge AKŞEHİRLİ

Master of Thesis, Department of Biostatistics and Medical Informatics

Supervisor: Prof. Dr. Handan ANKARALI

May 2012, 131 pages

In today's technology, with increase of capacity and speed of computers, in various sectors, capacity and complexity of the stored data also increases. As a result of processing of these data with appropriate advanced techniques, the dizzying technological changes have been observed in living conditions. For the purpose of extracting meaningful information from huge amounts of data, *data mining methods* that found intensive usage especially in recent years and in the future will replace with the traditional statistical methods are preferred. *Support vector machine (SVM)* is one of these methods and today it is used for classification, estimation or pattern recognition mostly in engineering applications. In this study it was aimed to define concepts of data mining, explain theoretical foundations of SVM particularly and use of this method for diagnosing in medical research. For this purpose, firstly data mining and data mining methods were mentioned briefly, then support vector machines method was described comprehensively. SVM, which is one of data mining methods is a machine learning tool that uses supervised learning to classify or predict the data. Basic idea behind SVM is to classify the data by dividing them with a plane or hyperplane. SVM achieves this procedure by doing maximum the margin between two classes. After training of the data, SVM aims to classify the new data correctly. In medicine, SVM is used for cancer morphology, identifying success of treatment and related gene, diagnosing various diseases. In application of the study, informations about 433 patient who were refer to the outpatient department of Zonguldak Karaelmas University Faculty of Medicine between date of 1-31 January 2011 for complaints of night eating syndrome were used. Using these data, variables that is effective in diagnosing were examined with univariate analysis, logistic regression and SVM methods and only GYA_puan variable found to be effective in the three approaches. Also when the classification performances of logistic regression and SVM were examined, it was seen that both of the methods gave similar results but superiority of SVM was discussed. In addition, performances of linear, polynomial, sigmoid and radial basis function (RBF) mentioned most common in applications were compared. In these comparisons, it was seen that the classification performances of 4 different kernel function gave similar results, however superiority of RBF in various ways was mentioned.

Keywords: Kernel functions, Support vector machines, Classification, Data mining

1.GİRİŞ ve AMAÇ

Sağlık alanında çok çeşitli araştırmalar yapılmakta olup bu alan önemli bir veri kaynağıdır. Bu verilerden elde edilen bilgiler, çeşitli hastalıkların sınıflandırılması, tanımlanması, tanı ve tedavisinde ve aynı zamanda hastalıklara karşı koruyucu önlemler almada kullanılır. Büyük boyutlardaki bu karmaşık verilerden faydalı bilgilerin ortaya çıkarılması için klasik istatistik yöntemler yetersiz kalmaktadır. Bu nedenle özellikle bilgisayar yazılım ve donanımının hızla geliştiği son 20 yılda büyük boyutlu verilerin kayıt altına alınması ve saklanması olanaklı hale gelmiş ve bu verilerden gizli kalmış faydalı bilgileri ortaya çıkarabilecek yeni istatistik yöntemler geliştirilmeye başlanmıştır^{1,2,3,4}.

Gizli kalmış ilişkileri ortaya çıkarmak için kullanılan yöntemlere genel olarak *makine öğrenmesi yöntemleri* adı verilmektedir. Makine öğrenmesi metotları, geçmişteki verileri kullanarak veriye en uygun modeli bulmaya çalışırlar ve yeni gelen verileri de bu modele göre analiz ederler⁵. Makine öğrenmesi uygulamalarından biri olan veri madenciliği de, istatistiksel yöntemler ile çeşitli bilgisayar algoritmalarını kullanarak, veri tabanlarındaki veriden gerekli bilgi keşfini sağlamak için geliştirilmiş yöntemlerden birisidir. Başka bir deyişle, büyük miktarda verinin incelenip, onun içinden işe yarayan bilginin (modelin) elde edilmesi işlemine *veri madenciliği* denilmektedir¹.

Veri madenciliği, çok büyük veri tabanlarındaki ya da veri ambarlarındaki veriler arasında bulunan ilişkiler, örüntüler (pattern), değişiklikler, sapma ve eğilimler, belirli yapılar gibi çeşitli bilgilerin ortaya çıkarılması ve keşfi işlemidir⁶. Veri madenciliği birçok disiplini birleştiren disiplinler arası bir yöntemdir ve günümüzde temel olarak, finans sektörü, haberleşme sektörü, sağlık sektörü ve devlet uygulamaları alanlarında kullanılmaktadır. Bu yöntemlerden bazıları karar ağaçları, sinir ağları, regresyon yöntemleri, genetik algoritmalar ve destek vektör makineleri şeklinde adlandırılmaktadır.

Sağlık alanında veri madenciliği yöntemlerinin kullanım alanlarından birkaç örnek aşağıdaki gibi verilebilir⁷:

- Hastalıkları etkileyen faktörlerin ortaya çıkartılması

- Hastalıklara erken teşhis koyularak sağlığın korunması ve doğru tedavi yöntemlerinin seçilmesi
- Sağlık hizmetlerindeki kalitenin artırılması ve geleceğe dönük doğru sağlık politikalarının oluşturulması
- Koruyucu hekimliğin yaygınlaştırılması ve sağlık harcamalarının düşürülmesi
- Salgın hastalıkların tespit edilmesi için gerekli önlemlerin alınması
- Sağlık harcamalarındaki hileli işlemlerin ortaya çıkartılması, maliyetlerin düşürülmesi
- İlaç geliştirici firmaların, sağlık veritabanlarından yararlanarak doğru ilaçları geliştirmesi

Veri madenciliği yöntemlerinden biri olan destek vektör makineleri (DVM), veriyi sınıflandırmak (DVS) veya tahmin yapmak (DVR) amacıyla kullanılan, eğiticili (supervised) bir makine öğrenmesi yöntemidir⁸. DVM yöntemi, yapay sinir ağları yöntemleri ile yakından ilgilidir ve hatta *sigmoid çekirdek fonksiyonunun* kullanıldığı bir DVM modeli, iki katmanlı perceptron yapay sinir ağı (two-layer, perceptron neural network) modeline eşdeğerdir⁹. Birçok veri madenciliği yöntemlerinde olduğu gibi sınıflama veya tahmin modeli geliştirilmeden önce, DVM’de de veri seti eğitim (training) ve test verileri olarak ikiye ayrılır. Eğitim verisi eğitilerek en iyi sınıflama/tahmin yapan model bulunmaya çalışılır. Bulunan bu modelin geçerliliği test verisi üzerinde araştırılır. Her iki veriden elde edilen sınıflama veya tahmin başarısı birbirine benzer ve yüksek ise modelin başarılı (geçerli) bir model olduğuna karar verilir.

DVM’ler ilk olarak 1960’larda ortaya atılmıştır. DVM’nin dayandığı teorisinin 1970’lerde geliştirilmeye başlanmasıyla DVM’lere ilgi artmış, ancak ilk başarılı çalışmaları 1990’lı yıllarda gerçekleştirilmiştir. Bununla birlikte DVM yöntemi, matematikçilerin ve yapay zeka ile ilgilenen bilim adamlarının ilgi odağı olmuştur¹⁰. 1990’ların sonlarında, DVM’lerle ilgili ilk detaylı çalışma Scholkopf ve Burges (1998) tarafından yapılmıştır. Daha sonra ise, 2005 yılında Wang ve Kecman, 2006 yılında Tan, Steinbach ve Kumar tarafından bu yöntem kullanılarak çeşitli çalışmalar yapılmıştır³.

Günümüzde, DVM'nin birçok dünya problemine uyarlanabilir olması, DVM yöntemine olan ilgiyi arttırmakta ve bununla birlikte, bu yöntemle yapılan çalışmalar her alanda ağırlık kazanmaktadır. DVM'nin kullanım alanları incelendiğinde, görüntü ve metin sınıflandırma, nesne tanıma, el yazısı tanıma, ses tanıma ve yüz tanıma gibi çeşitli örüntü tanıma uygulamalarında sıkça kullanıldığı görülür¹¹. DVM, aynı zamanda biyolojik uygulamalarda da yükselen bir başarı göstermektedirler¹². Tıpta DVM özellikle, kanser morfolojisinde, tedavi başarısının ve ilgili genin belirlenmesinde, çeşitli hastalıkların teşhisinde kullanılmaktadır^{13,14}. DVM ilk uygulamalarından itibaren genellikle mühendislik alanında kullanılmıştır. Sağlık alanında yapılan uygulamalara bakıldığında, bu uygulamaların da yine mühendislik temelli kişiler tarafından yapıldığı görülmektedir.

DVM pratikte daha çok sınıflama amacıyla kullanılmaktadır ve sağlık alanı araştırmalarında bir tanı yöntemi olarak tercih edilebilir. Sınıflama genel anlamıyla; iki veya daha fazla grubu (sınıfı) çeşitli özellikler bakımından ayırt edebilme olarak tanımlanabilir. DVM yardımıyla sınıflamada, en az iki grup sahip oldukları özellikler bakımından doğrusal veya doğrusal olmayan modellerle ayırt edilebilmektedir. Sağlık araştırmalarında genellikle sağlıklı kontrol ve hasta grubu olarak iki grup ayırımı yapılmak istenir^{3,4}. Bu tez çalışmasında ağırlıklı olarak, doğrusal ve doğrusal olmayan DVM modelleri yardımıyla iki grubun sınıflandırılması üzerinde durulacaktır.

DVM'nin temelini, verilerin bir düzlem veya hiperdüzlem ile ayrılarak sınıflandırılması işlemi oluşturmaktadır. Yani, iki sınıfa ait verileri ayırabilecek en uygun düzlemi veya hiperdüzlemi belirlemektir. Doğrusal olarak ayrılabilen verileri, ait oldukları boyutta bir düzlem ile ayırabilmek mümkünken; doğrusal olarak ayrılamayan verilerin ait oldukları boyuttan daha yüksek boyutlu bir uzaya taşınarak, burada bir hiperdüzlem ile ayırmak mümkün olacaktır. DVM, doğrusal olarak ayrılabilen veriler söz konusu olduğunda, verileri ayırabilecek sonsuz sayıdaki doğru içerisinden marjini en yüksek yapacak olan doğruyu seçmeyi hedeflemektedir. Doğrusal olarak ayrılamayan verilerin olduğu durumda ise, bir haritalama yöntemi ile orijinal veriyi daha yüksek boyutlu bir uzaya taşır ve burada verileri sınıflandırmak için optimum olabilecek doğrusal ayırıcı hiperdüzlemi bulmaya çalışır^{12,15,16}.

Bu tez çalışmasında, genel bilgiler başlığı altındaki ilk bölümde; veri madenciliğinin tanımı yapılmış, veri madenciliğinin temellerinden, uygulama alanlarından, teknik

kavramlarından ve çeşitli veri madenciliği yöntemlerinin sınıflandırmasından bahsedilmiştir.

İkinci bölümde; makine öğrenmesinden, makine öğrenmesi türlerinden, DVM'nin temelini oluşturan istatistiksel öğrenme teorisinden ve istatistiksel öğrenme teorisinde önemli yeri olan Vapnik-Chervonenkis teorisi ile yapısal risk minimizasyonundan bahsedilmiştir.

Üçüncü bölümde; veri madenciliği yöntemlerinden biri olan ve bu tez çalışmasının konusunu oluşturan DVM'nin teorik temelleri detaylı olarak tanımlanmış, doğrusal olarak ayrılabilen ve ayrılamayan durumlarda destek vektör sınıflandırmasının işleyişinden, doğrusal olarak ayrılamayan verilerin sınıflandırılmasında önemli yeri olan çekirdek (kernel) fonksiyonlarından ve DVM'de kullanılan çeşitlerinden kapsamlı olarak bahsedilmiştir. Bu bölümün sonunda, destek vektör regresyonu özet olarak anlatılmıştır.

Gereç ve yöntem bölümünde, kullanılan veri seti ayrıntılı olarak açıklanmış ve veri analizi için uygulanan yöntemler nedenleri ile tanımlanmıştır. Ayrıca DVM yönteminin uygulaması için kullanılan DTREG programı adım adım tanıtılmıştır.

Bulgular bölümünde, kullanılan yöntemler yardımıyla elde edilen sonuçlara yer verilmiştir. Bu sonuçlara ait sayısal ve görsel değerler tablo ve grafikler halinde gösterilmiştir.

Son olarak tartışma bölümünde ise, elde edilen bulgular tartışılmış ve tezin amacı doğrultusunda bulguların literatüre katkısı vurgulanmıştır.

Bu tez çalışmasının amacı, DVM yönteminin teorik özelliklerini tanımlamak ve sağlık alanında yaygın olarak kullanımına katkı sağlamaktır.

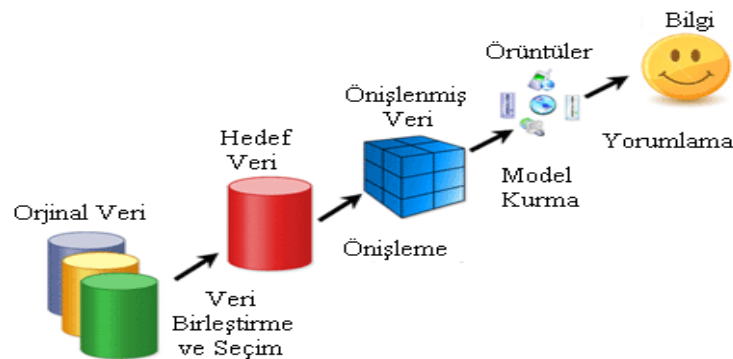
2. GENEL BİLGİLER

2.1. Veri Madenciliği

2.1.1. Veri madenciliğinin tanımı ve temel bilgiler

Gün geçtikçe, uydu verileri, tıbbi veriler, alışveriş verileri, otomasyon verileri gibi çeşitli veri tiplerindeki veri miktarlarının hızla artmasıyla, bu verilerin toplanması ve saklanması gibi problemler ortaya çıkmıştır⁶. Toplanan veri miktarının artması verinin karmaşıklığının da artmasına neden olmuştur. Günümüzde teknolojinin ilerlemesiyle, çok büyük boyutlu veriler bilgisayar ortamında veya çeşitli elektronik ortamlarda saklanabilir hale gelmiştir¹⁷. Ancak, önemli olan veriyi elde etmek ve depolamak değil, elde edilen verileri anlamak ve çözümlenektir. Toplanan veriler içinde gerekli ve faydalı bilgiyi elde etmeyi sağlayacak verilere ulaşmak istenmektedir. Veri tabanlarındaki veri sadece saklandığından, bu veriler tek başlarına değersizdirler. Ancak bu veriler belli bir amaç doğrultusunda işlendiği zaman anlamlı hale gelmektedir. Bu noktada, ham veriyi bilgiye veya anlamlı hale dönüştürme işini yapabilen *veri madenciliği* çözüm olarak düşünülebilir¹⁸.

Veri madenciliği, çok büyük veri setleri içinden faydalı bilgiye ulaşabilmek için geliştirilmiş bir yöntemler grubudur ve hipotezler üzerinden modeller kurmak yerine, varlığı bilinen fakat kesin olmayan örüntü yapılarını araştırır^{3,19,20}. Veri madenciliğinde amaç; bilgi elde edilmesi zor olan büyük veri yığınlarını analiz ederek anlamlı, gizli ve faydalı olabilecek bilgileri ortaya çıkarmak; bu bilgileri kullanarak bir model oluşturup, yeni gelecek bir veri hakkında yorum yapmayı ve bu veri hakkında tahminde bulunmayı sağlamaktır²¹. Şekil 2.1’de veri madenciliğinin görsel bir ifadesi görülmektedir.



Şekil 2.1: Veri Madenciliğinin İşleyişi

Veri madenciliği, birçok analiz aracı kullanımıyla veri içerisinde örüntü ve ilişkileri keşfederek, bunları doğru tahminler yapmak için kullanan bir süreçtir¹. Başka bir deyişle, veri madenciliği, büyük ölçekli veriler arasından bilgiye ulaşma, veriyi madenleme işlemidir²². Bunun yanında, geleceğe dönük kararlar alma aşamasında fikir de vermektedir. Ayrıca veri madenciliği Guidici tarafından, “*büyük miktardaki veriden, veri tabanı sahibine açık ve faydalı bilgi sağlamak amacı ile başlangıçta bilinmeyen örüntü ve ilişkileri keşfetmek için kullanılan seçme, keşfetme ve modelleme süreci*” olarak tanımlanmıştır³.

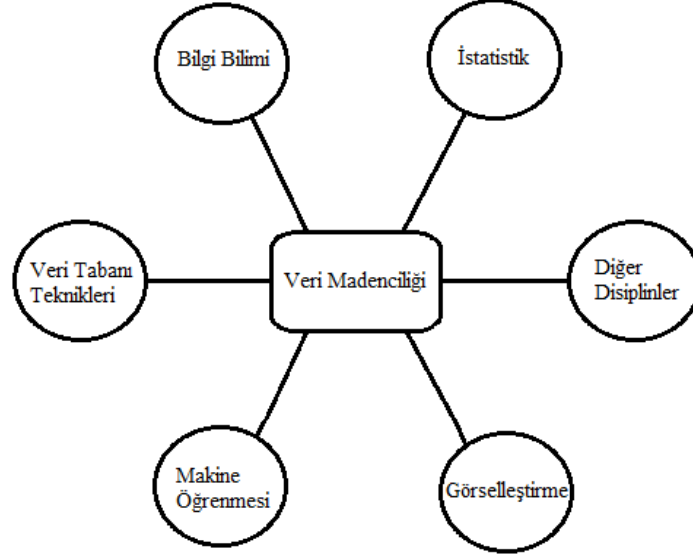
Başka bir tanım; “*veri madenciliği, varlığı bilinmeyen ilişkileri bulmak ve veri sahibine hem anlaşılabilir hem de faydalı olacak yollarla veriyi özetlemek için, gözlenmiş verilerin analiz edilmesi*” olarak verilebilir³.

Veri madenciliğinde kullanılacak olan verinin de bazı özelliklere sahip olması gerekmektedir. Bu özellikler de veri ambarı (data warehouse) ile sağlanmaktadır. Veri ambarları basit olarak, veri madenciliği işleminin yapılacağı verilerin toplandığı özel veri tabanları şeklinde ifade edilebilir. Veri ambarlarının oluşturulması işlemi verinin çeşitli kaynaklardan toplanarak, veriler içerisindeki uyumsuzluklar ve hatalardan arındırılmasından ibarettir⁶⁻²³.

Veri madenciliği teknikleri ile ilgili olarak 1950’li yıllarda, matematikçiler mantık ve bilgisayar bilimleri alanlarında çalışarak yapay zekâ ve makine öğrenmesi kavramlarını geliştirmişlerdir. Daha sonra, 1960’lı yıllarda istatistikçilerin yeni bir algoritma keşfetmesiyle, regresyon analizi, en büyük olabilirlik kestirim, sinir ağları vb. metotlar veri madenciliğinin ilk adımlarını oluşturmuştur⁶. 1970, 1980, 1990’lı yıllarda yeni programlama dilleri ve yeni bilgisayar tekniklerinin geliştirilmesiyle veri madenciliği; genetik algoritmalar, EM (Expectation Maximization) algoritmaları, karar ağaçları ve DVM gibi çeşitli yöntemleri de içine almıştır. 1990 yılıyla beraber veri tabanlarında bilgi keşfi (VTBK)’nin ilk adımları oluşturulmuş ve büyük veritabanları için veri ambarları geliştirilmiştir. Ayrıca, yeni teknolojilerle beraber veri madenciliği geliştirilerek yaygın olarak kullanılmaya başlanmıştır²³.

Veri madenciliği, disiplinler arası bir alandır ve istatistik, veri tabanları, yapay zeka, makine öğrenmesi, örüntü tanıma, bilgi toplama, veri görselleştirme, paralel ve dağıtık hesaplama gibi birçok disiplinden yardım alır^{18,24}. Veri madenciliğinde kullanılan yöntemlerin birçoğu iki ayrı araştırma dalı olan istatistik ve makine öğrenme olarak

bilinmektedir¹⁸. Şekil 2.2’de veri madenciliğinin çeşitli disiplinlerle olan ilişkisi gösterilmektedir.



Şekil 2.2: Veri Madenciliğinin Diğer Disiplinlerle İlişkisi

Günümüzde veri madenciliği; bankacılık, pazarlama, sigortacılık, telekomünikasyon, borsa, sağlık, endüstri, ilaç geliştirme, astronomi, tıp, üretim kontrolü, biyoloji, genetik ve mühendislik gibi birçok dalda uygulanmaktadır^{220,22}. Son zamanlarda, veri madenciliğine olan ilgi artmaktadır ve dünyada geniş bir uygulama alanına sahip olan bu konu, ülkemizde de gün geçtikçe önem kazanmaktadır³.

2.1.2 Makine öğrenmesi

Bilgisayar teknolojilerindeki gelişmeyle birlikte, günümüzde büyük miktarlardaki veriyi saklamanın ve işlemenin mümkün olmasının yanı sıra, bu verilere uzak mesafelerden bilgisayar ağları üzerinden erişebilmek de mümkün hale gelmiştir. Artık çoğu veri toplama işlemi dijital ortamda yapılmakta ve dolayısıyla, güvenilir veriler kayıt altına alınabilmektedir²⁵. Örneğin süpermarket kasalarından, para çekme makinelerinden, kredi kartı cihazlarından, e-ticaret uygulamalarından her an milyonlarca veri, verilerin saklandığı merkezlere ulaşmaktadır. Bunların dışında, bir hastanedeki röntgen cihazından, bir güvenlik kamerasından, bir iris tanıma sisteminden, bir kumaş kalite ölçüm kamerasından, borsadaki işlemlerden yine birçok veri elde edilmekte ve analiz için beklenmektedir. Bu işlemlerin her birinde analizden farklı şeyler beklenmektedir. Bir süpermarket işletmecisi hangi tür ürünlerin bir arada satıldığını, bir borsa analisti hisse senedinin yarınki değerini, iris tanıma sistemi verinin kime ait olduğunu, kredi kart

sistemi kartı kullananın kartın sahibi olup olmadığını, bir güvenlik kamerası olağandışı bir durum olup olmadığını öğrenmek istemektedir. Bu sistemlerin tümünde yeni verinin işlenmesi için, geçmişteki verilerden faydalanılmaktadır. Çok büyük miktardaki verilerin elle işlenmesi, analizinin yapılması mümkün değildir. Bu problemlere çözüm bulmak amacıyla makine öğrenmesi metotları geliştirilmiş ve hala geliştirilmeye devam edilmektedir. Makine öğrenmesi metotları; geçmişteki verileri kullanarak, veriye en uygun modeli bulmaya çalışmakta ve yeni gelen verileri de bu modele göre analiz etmektedirler²⁵.

Makine öğrenmesi yöntemlerinde veri madenciliğinin aksine, veri setlerinin büyük olması gerekmemektedir ve bu yöntemler, küçük veri setleri için de başarılı sonuçlar vermektedir. Makine öğrenmesi yöntemleri, yeni örnekler geldiğinde, kendi kendine öğrenebilen modeller oluşturmaktadır. Veri madenciliği ise, makine öğrenmesi olarak adlandırılan yöntemler bütünündeki modelleri büyük veri setlerine uyarlayarak ortaya örüntü çıkarmayı amaçlamaktadır. Makine öğrenmesi modelleri, veri madenciliğinde kullanıldığında bunlardan öğrenme beklenmez³.

Makine öğrenmesi, insan ve hayvan öğrenmesi ile paralellik göstermektedir ve makine öğrenmesi teknikleri; hayvan ve insan öğrenmesinden yola çıkarak geliştirilmiş tekniklerdir²⁶. Burada, sistem önceden bilmediği ve yapamadığı işlemleri, belirli bir eğitimden sonra gerçekleştirebiliyorsa, sistemin bu işlemleri öğrendiği düşünülmektedir²⁷.

Öğrenme Simon tarafından, “*zaman içinde yeni bilgilerin keşfedilmesi yoluyla davranışların iyileştirilmesi süreci*” olarak tanımlanmaktadır²⁸. Genel anlamda veriden öğrenme ise, sonlu sayıda eğitime verisi kullanan tahmin fonksiyonunun öğrenilmesini sağlayan öğrenme makinesi veya algoritmanın oluşturulması olarak tanımlanır²⁹. Makine öğrenmesinde bu öğrenme işinin, bilgisayar tarafından gerçekleştirilmesi amaçlanmaktadır. Burada, öğrenmenin tanımında da verildiği gibi, bilgisayarın da insan gibi zaman içerisinde tecrübe kazanması istenmektedir. Bilgisayarın öğrenebilmesi ve tecrübe sahibi olabilmesi, bilgisayarın ilgili olay hakkında bilgiler ile donatılmasına bağlıdır.



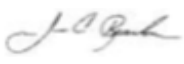

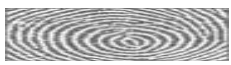

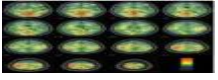



Makine öğrenmesi, “*bilgisayarın bir olay ile ilgili bilgileri ve tecrübeleri öğrenerek, gelecekte oluşacak benzer olaylar hakkında kararlar verebilmesi ve problemlere çözümler üretebilmesi*” olarak da tanımlanabilir²⁸. Daha başka bir tanım olarak;

“makine öğrenmesi; aynı görevin yinelenmesinde ya da aynı popülasyondan çekilen farklı bir görevde, makinenin ikinci sefer daha iyi çalışmasını sağlayan, sistemdeki herhangi bir değişimdir”³⁰.

İlk makine öğrenmesi modeli *perceptron*, F. Rosenblatt tarafından 1957’de önerilmiştir. Bu, öğrenme işlemlerinin matematiksel analizinin gerçek anlamda başladığı zamandır. Perceptron ile ilgili deneyler kapsamlı olarak bilinir hale gelince, başka tür öğrenen makineler önerilmiş ve daha sonra da gerçek dünya problemlerini çözmeye kullanılabilecek bilgisayar programlarından sinir ağları ve bazı araştırmacılar tarafından sınıflamada kullanmak için karar ağaçları geliştirilmiştir³¹.

Günümüzde, makine öğrenmesi uygulamaları birçok alanda sıklıkla kullanılmaktadır. Bu uygulamaların bir kısmı Tablo 2.1’de verilmiştir.

Tablo 2.1: Makine Öğrenmesi Yönteminin Uygulama Alanları

Girdi (pattern, nesne)	Örnek	Çıktı
El yazısı/Kitap yazısı		Resmi metne çevirme
İris		İrisin sahibini bulma
İmza		İmza gerçek olup olmadığını bulma
Yüz resmi		Resimdeki kişiyi tanıma
Parmak izi		Parmak izinin sahibini bulma
İnsan sesi		Sesin sahibini bulma
		Sesi metne çevirme
Tıbbi veri		Kişinin hasta olup olmadığını bulma
Metin/Mail		Metnin konusunu bulma
		Metnin yazarını bulma
		Mail’in spam olup olmadığını bulma
Güvenlik kamerası kayıtları		Olağan dışı bir durum olup olmadığını bulma
İşitme engellilerin kullandığı işaret dili kamera kayıtları		Görüntüdeki kişinin anlatmak istediğini bulma

Sonuç olarak makineler insanlığın işgücüne sağladıkları katkıyı, makine öğrenmesi metotları sayesinde insanlığın beyin gücüne de sağlamaya başlamışlardır. Her tür uygulama için çok miktarda verinin analiz edilerek gelecekle ilgili varsayımlar geliştirilmesine, kararlar verilmesine yardımcı olan makine öğrenmesi metotlarının önemi ve katkıları her geçen gün arttırmaktadır⁵.

2.1.2.1. Makine öğrenmesi türleri

Makine öğrenmesi yöntemlerinde değişik öğrenme türleri kullanılmaktadır. Öğrenmeyi gerçekleştirecek olan sistem ve kullanılan öğrenme algoritması bu öğrenme türlerine göre değişmektedir³². Eğitim verisinden öğrenmek için kullanılan öğrenme türleri aşağıdaki gibi sıralanabilir:

- Eğitici (Supervised) öğrenme
 - ✓ Sınıflama
 - ✓ Regresyon
- Eğitici (Unsupervised) öğrenme
- Yarı-eğitici (Semi-supervised) öğrenme
- Pekiştirerek (Reinforcement) öğrenme
- Aktif (Active) öğrenme

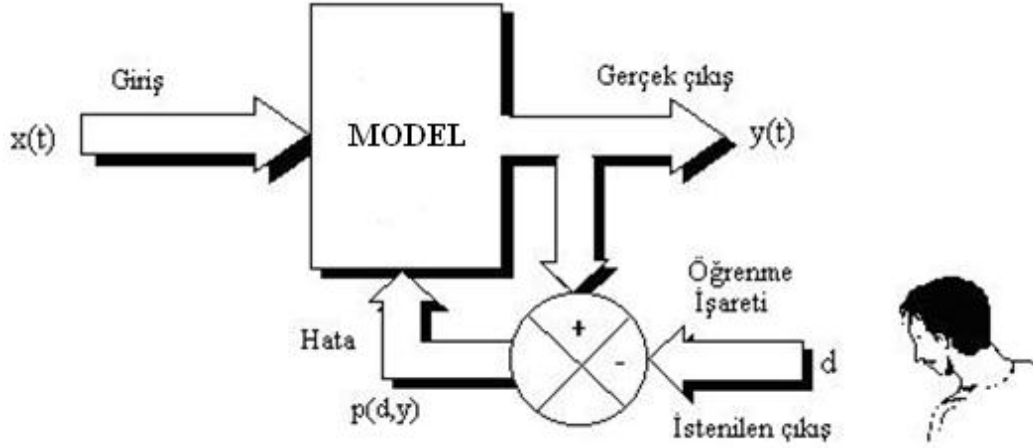
Bu öğrenme türlerinden eğitici ve eğitici öğrenme, en sık kullanılan öğrenme türleridir.

2.1.2.1.A. Eğitici öğrenme

Eğitici öğrenmede, sınıflar ve her sınıfın önceden sınıflandırılmış verileri bir eğitici tarafından tanımlanır ve makinenin bir eğitim seti ile öğrenmesi sağlanır³¹. Yani, eğitim için eğitici veriler (eğitim seti) kullanılmaktadır. Eğitim seti giriş bilgileri ve istenen (hedef) bilgiler olmak üzere iki ayrı vektör gibi düşünülebilir. Vektörlerin her bir karşılıklı elemanları bir eğitim çiftini oluşturmaktadır. Eğitim seti, makinenin eğitimine başlamadan önce belirlenmektedir³³.

Bu tür öğrenmede eğitici, sisteme öğrenilmesi istenen olay ile ilgili örnekleri eğitim seti olarak verir. Bir başka ifadeyle, her örnek için hem girdiler, hem de o girdiler karşılığında oluşturulması gereken çıktılar sisteme gösterilir. Sistemin görevi girdileri

eğiticinin belirlediği çıktılara haritalamaktır. Bu sayede olayın girdileri ile çıktıları arasındaki ilişkiler öğrenilmektedir³². Girilen değerle istenen değer arasındaki fark, hata değeri olarak önceden belirlenen değerden küçük oluncaya kadar eğitime devam edilir. Sistem belirli bir sıralamadaki girdiler için istenen istatistiksel doğruluğu elde ettiği zaman, eğitime işlemi tamamlanmış kabul edilir ve süreç sona erer³⁴. Şekil 2.3'te eğiticili öğrenme yapısı gösterilmiştir³³.



Şekil 2.3: Eğiticili Öğrenme Yönteminin Yapısı

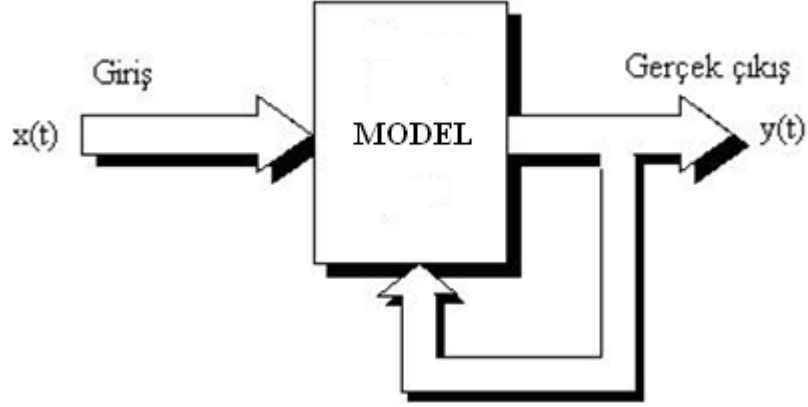
Eğiticili öğrenmede temel olarak, sınıflama ve regresyon yöntemleri bulunmaktadır. Sınıflama ve regresyon, önemli veri sınıflarını ortaya koyan veya gelecek veri eğilimlerini tahmin eden modelleri kurabilen iki veri analiz yöntemidir. Sınıflama bireylerin ait oldukları grupları belirlemede kullanılırken, regresyon belirli değerlerin tahmin edilmesinde kullanılır. Sınıflamada bir sınıfı diğerinden ayırmada kullanılacak olan bir karar sınırı öğrenilir. Regresyonda ise, örneklerden sürekli girdi-çıkı haritalaması öğrenilmektedir³⁵.

Karar ağaçları, regresyon yöntemleri, sinir ağları, destek vektör makineleri, k-en yakın komşuluğu gibi yöntemler eğiticili öğrenme yöntemleridir.

2.1.2.1.B. Eğitici olmayan öğrenme

Eğitici olmayan öğrenmede, sınıfları tanımlayan bir eğitici söz konusu değildir ve sistem kendi sınıfını bulmak zorundadır³¹. Bu yöntemlere kendi kendine öğrenen modeller de denilmektedir. Eğitici olmayan öğrenmede eğitim seti kullanılmamaktadır ve eğitim için sadece giriş bilgilerinin olması yeterlidir. Bu algoritmalar, birbirine benzer giriş bilgilerini gruplamakta veya giriş bilgisinin hangi gruba ait olduğunu göstermektedir.

Burada modelin performansını kendiliğinden izlemesi söz konusudur. Ancak, modelin kendini nasıl organize edeceği hakkında bir miktar bilgiye sahip olmalıdır³³. Şekil 2.4'te eğitici-siz öğrenme yapısı gösterilmiştir.

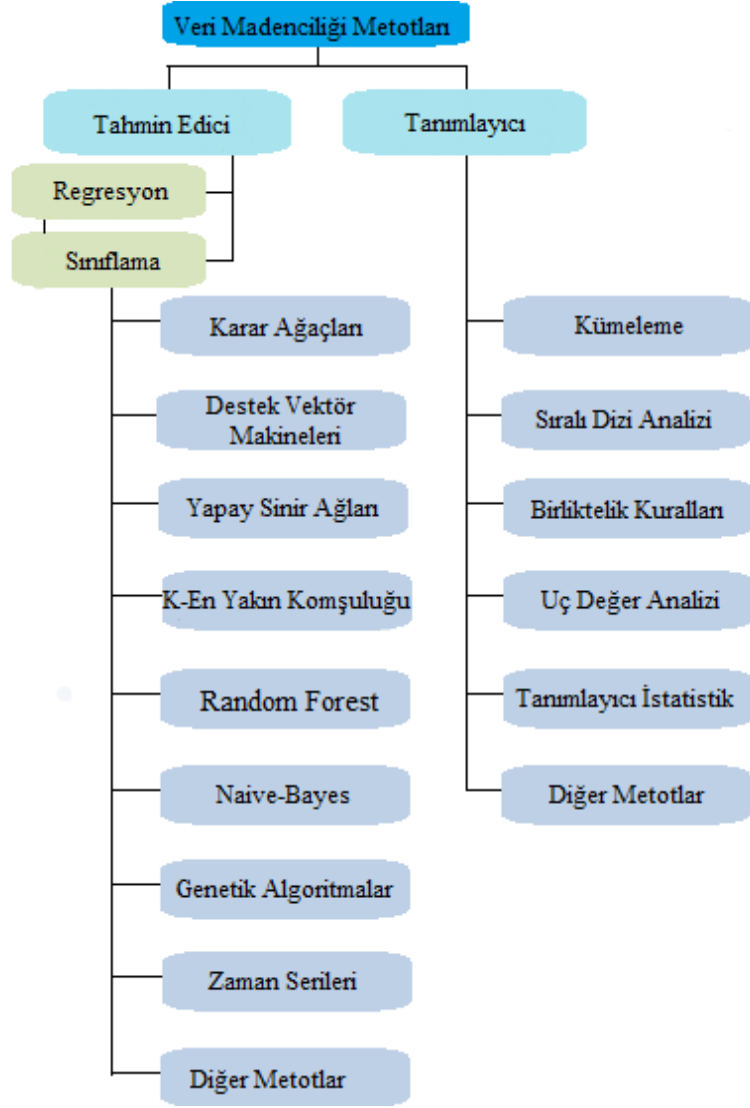


Şekil 2.4: Eğitici-siz Öğrenme Yönteminin Yapısı

Kümeleme analizi, korelasyon analizi, faktör analizi ve istatistiksel ölçümler eğitici-siz öğrenme yöntemlerindedir.

2.1.3. Veri madenciliği yöntemlerine genel bakış

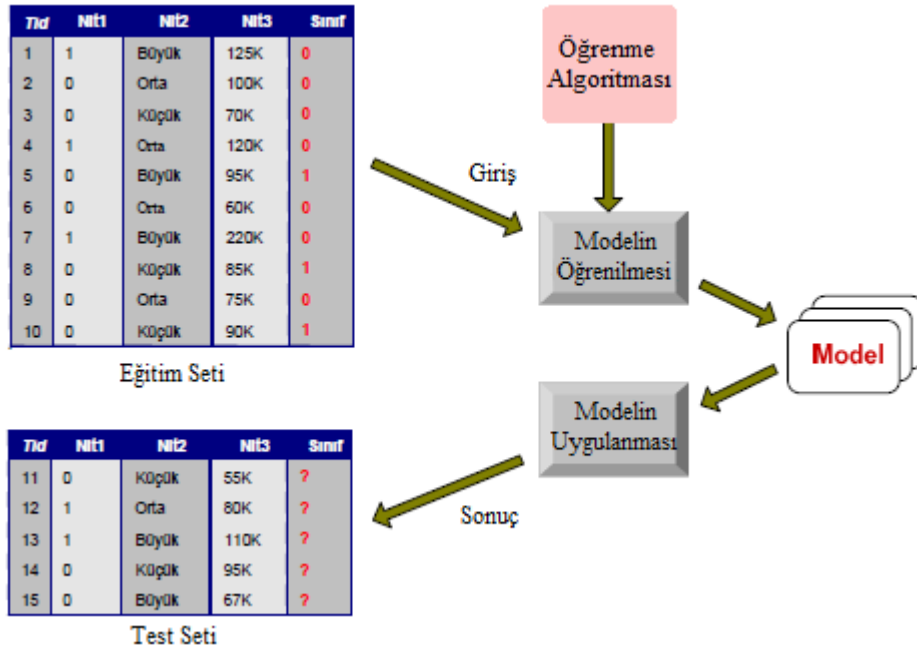
Veri madenciliği yönteminde modeller temel olarak, tahmin edici ve tanımlayıcı şeklinde iki ana başlık altında incelenebilir^{20,35}. Tahmin edici modellerde, sonuçları bilinen verilerden hareket edilerek bir model geliştirilmesi ve kurulan bu modelden yararlanılarak sonuçları bilinmeyen veri kümeleri için sonuç değerlerinin tahmin edilmesi amaçlanmaktadır. Tanımlayıcı modellerde ise, karar vermeye yardımcı olabilecek mevcut verilerdeki örüntülerin tanımlanması sağlanmaktadır²⁰. Şekil 2.5'te veri madenciliği metotları gösterilmektedir.



Şekil 2.5: Veri Madenciliği Metotları

Tahmin edici metotlardan biri olan sınıflama, veri madenciliği araştırmalarının önemli bir parçasıdır. Sınıflama temel olarak; dış gözlemlere dayanarak bir nesneyi farklı sınıflardan birine atayacak bir kural bulma, başka bir ifadeyle, bilinmeyen bir veri örneğinin sonlu sayıda sınıftan hangisine ait olduğunu tahmin edilmesi işlemidir³⁶. Amaç ise, yeni bir örneğin (verinin), belirli sınıflar içinde hangi sınıfa ait olduğunu belirleyecek bir sınıflayıcı oluşturmaktır. Başka bir deyişle sınıflamanın amacı, sınıf özellikleri ve çıkış özellikleri arasındaki bazı ilişkileri keşfederek, bu ilişkiler ile sınıflandırılmamış yeni bir veri örneğinin sınıfını tahmin etmektir. Burada önemli olan, her bir sınıfın özelliklerinin önceden net bir şekilde belirlenmiş (eğitilmiş) olmasıdır^{29,26,37,38}. Sonuç olarak, sınıflama metodu ile bir model oluşturulmakta ve yeni gelen bir örneğin hangi sınıfa ait olacağı belirlenebilmektedir³⁷.

Sınıflama işlemi genelde iki ana adımdan oluşmaktadır. İlk adımda, önceden belirlenmiş olan verilerin sınıflarını tanımlayan bir model oluşturulur. Bu model, çeşitli özelliklerle tanımlanmış veri gruplarının analiz edilmesiyle öğrenilir. Bu analiz edilen veri grubu eğitim setini oluşturur ve her bir eğitim verisinin sınıfı belirtildiği için bu çeşit öğrenme eğitici öğrenme olarak bilinir²⁹26. İkinci adımda model, sınıfı henüz bilinmeyen örneklerin sınıf etiketlerini belirlemek amacıyla kullanılır. Eğitici öğrenmede, eldeki veriler *eğitim seti (training set)* ve *test seti (test set)* şeklinde iki gruba ayrılmakta ve eğitim setiyle öğrenilen model, test setinde doğru sınıflama yapmayı amaçlamaktadır. Eğitim seti, sınıflama modelinin elde edileceği verileri içermektedir. Test seti ise, eğitim setiyle elde edilen modelin doğru sınıflama yapıp yapmadığını belirlemek için kullanılır. Yani, eğitim setiyle model inşa edilirken, test setiyle model doğruluğu kontrol edilir. Burada hedef; yeni örneklerin doğru şekilde daha önceden belirlenmiş sınıflara atanmasıdır³⁹. Şekil 2.6'da sınıflamanın görsel bir özeti verilmiştir.



Şekil 2.6: Sınıflamanın Görselleştirilmesi

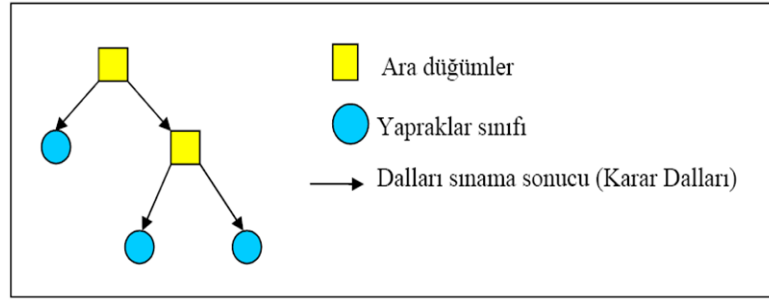
Her biri önceden belirlenmiş bir sınıfa ait olan veriler kümesi verildiğinde, yeni bir verinin hangi sınıfa ait olacağına karar verecek modeli bulma, sınıflandırma problemidir. Oluşturulan model eğitim verisinin analizine dayanmaktadır³⁷. Şekil 2.5'te de görüldüğü gibi, veri madenciliğinde sınıflama yapmak için çeşitli yöntemler

bulunmaktadır. Bu yöntemlerden özellikle son yıllarda uygulamasına rastlananlar aşağıdaki gibi sıralanabilir:

- Karar ağaçları
- Random forest
- Yapay sinir ağları
- Naïve-Bayes
- K-En yakın komşuluğu
- Genetik algoritmalar
- *Destek vektör makineleri*

Karar ağaçları: Karar ağaçları; kolay oluşturulması, yorumlanmalarının kolay olması, veri tabanı sistemleri ile kolayca entegre edilebilmeleri ve güvenilirliklerinin iyi olması nedenleri ile veri madenciliği sınıflama modelleri içerisinde en yaygın kullanıma sahip tekniktir^{20,40}35. Adından da anlaşılacağı gibi karar ağaçları, bir ağaç görünümünde olan, tahmin edici bir yöntemdir^{20,35}35.

Karar ağaçları karar düğümü, dallar ve yapraklardan oluşur³35. Karar düğümleri yapılacak olan testi, dallar testteki değerleri, yapraklar ise sınıfı belirtir. Burada, karar düğümleri ile belirtilen testin sonucu ağacın veri kaybetmeden dallara ayrılmasını sağlar. Her düğümde test ve dallara ayrılma işlemleri ardışık olarak gerçekleşir ve bu ayrılma işlemi üst seviyedeki ayrımlara bağlıdır. Ağacın her bir dalı sınıflama işlemini tamamlamaya adaydır. Eğer bir dalın ucunda sınıflama işlemi gerçekleşemiyorsa, o dalın sonucunda bir karar düğümü oluşur. Ancak dalın sonunda belirli bir sınıf oluşuyorsa, o dalın sonunda yaprak vardır³⁵20. Bu yaprak, veri üzerinde belirlenmek istenen sınıflardan biridir. Karar ağacı işlemi, kök düğümünden başlar ve yukarıdan aşağıya doğru yaprağa ulaşana dek ardışık düğümleri takip ederek gerçekleşir. Karar ağaçlarında kök ve her düğüm bir soruyla etiketlenir. Düğümlerden ayrılan dallar ise ilgili sorunun olası yanıtlarını belirtir. Her düğüm de söz konusu sorunun çözümüne yönelik bir tahmini temsil eder²35,37.



Şekil 2.7: Karar Ağacı Şeması

Karar ağaçları; bir karar ağacı, ağacı oluşturacak bir algoritma ve ağacı veriye uygulayacak ve söz konusu problemi çözecek bir algoritma şeklinde üç bölümden oluşan bir modeldir².

Karar ağaçlarının avantajları şu şekilde özetlenebilir^{1,3}. Modellenmesinin ucuz olması

- Yorumlanmasının kolay olması
- Veri tabanı sistemleri ile kolaylıkla birleştirilebilmesi
- Yüksek güvenilirlikte sonuçlar üretmesi
- Nitel, nicel, sürekli, kesikli tüm değişkenlere uygulanabilen algoritmalara sahip olması

Random Forest (RF): RF, çok sayıda karar ağacının bir araya gelmesiyle oluşan bir tekniktir. Sınıflama yapan her bir karar ağacı bireysel oy almakta ve işlem sonunda en yüksek oyu alan karar ağacının yaptığı sınıflama kullanılmaktadır. Her bir karar ağacı eğitildiği veri grubundan farklı bir veri grubuyla karşılaştığında aynı performansı gösteremeyeceği için, yöntem çok sayıda karar ağacını birleştirmekte ve bu sayede sınıflama performansını ve doğru sınıflama oranını artırmaktadır. RF yönteminde ağaç, bütün verinin oluşturduğu tek bir düğümlerle başlamakta, eğer örneklerin hepsi aynı sınıfa ait ise düğüm, yaprak olarak sonlanmakta ve sınıf etiketi verilmektedir. Eğer örnekler aynı sınıfa dahil değilse, örnekleri sınıflara en iyi bölecek olan özellik seçilmektedir⁴¹.

RF'nin sağladığı avantajlar aşağıdaki gibi sıralanabilir:

- Çok sayıda girdi değişkeni ile çalışabilmektedir.
- Sınıflama yaparken değişkenlerin önemlerini kestirmektedir. Özellikle gen çalışmalarında, önemli genin seçimi için oldukça faydalıdır.
- Sınıflama uygulamalarında yüksek doğruluk elde edilmektedir.
- Eksik ve kayıp verilerin analizinde kullanılabilir.

- Aşırı uyum sorunu birçok yöneme göre daha az görülmektedir.
- Gözlemler arası yakınlık hesaplayarak kümeleme analizinde kolaylık sağlamaktadır.
- Öğrenme süreci oldukça hızlı işlemektedir.

*Yapay sinir ağları (YSA):*YSA, beynin ve sinir sisteminin yapısından esinlenerek ortaya atılmış, sinir sisteminin çalışmasını taklit ederek veriden öğrenme sağlayan bir bilgi işleme sistemidir^{23,42}. Bu yöntem, canlılardaki sinir sisteminde gerçekleşen öğrenme, ilişkilendirme, sınıflandırma, genelleme, tahmin ve özellik belirleme gibi işlevlerini yerine getirmeyi hedefler. YSA dışarıdan aldığı dinamik verilere göre dinamik sonuçlar üretebilen, birbirine bağlı basit elemanlardan oluşan bir yapıya sahiptir⁴³42.

YSA; girdi ve çıktı değerleri arasında, küçük hesaplama birimlerinden elde edilen sonuçları birleştirerek özetleyen bir modelleme yöntemidir1. YSA'nın çalışma sistemi, nöronlara benzeştirilmiş işlem öğeleri arasındaki ilişkilerle yapılandırılmıştır ve YSA'da sinir sisteminin ana birimi olan nöronların çalışma prensibi kopyalanılarak işlem birimleri üretilmektedir2. Yapay sinir hücreleri ya da düğüm (node) olarak da bilinen bu işlem birimlerinin birbirine bağlanmasıyla yapay sinir ağları oluşturulmaktadır. Oluşturulan yapay sinir ağları öğrenme, hafızaya alma, sınıflama, genelleme, tahmin ve veriler arasındaki ilişkileri ortaya çıkarma gibi görevleri gerçekleştirmektedir.

Temel bir yapay sinir ağı hücresi, biyolojik sinir sistemi hücresine göre daha basit bir yapıya sahiptir. Sinir ağı ilişkilendirmeyi yapan iç kuralları üretir ve bu kuralları, bunların sonuçlarını örneklerle karşılaştırarak düzenler. YSA'da öğrenme, girdi seti üzerinden yapılır ve girdi setindeki değişiklikler kullanılarak, deneme yanılma yoluyla bir çıktı üretilir3. YSA'nın avantajları aşağıdaki gibidir⁴³.

- YSA doğrusal olmayan bir yapıya sahiptir ve YSA'nın temel işlem birimlerinin doğrusal olmayan yapısı, algoritmayı doğrusal olmayan problemlerin çözümünde en temel araçlardan biri haline getirmektedir.
- YSA paraleldir ve bu özellik yavaş birimin etkisini azaltmaktadır. Çünkü seri işlemlerde bir birimin hızı tüm sistemin çalışma hızını etkilerken, paralel sistemlerde yavaş birimin etkisi azdır. Dolayısıyla YSA'ların paralel işlemler kullanması toplam bilgi işleme hızını artırmaktadır.

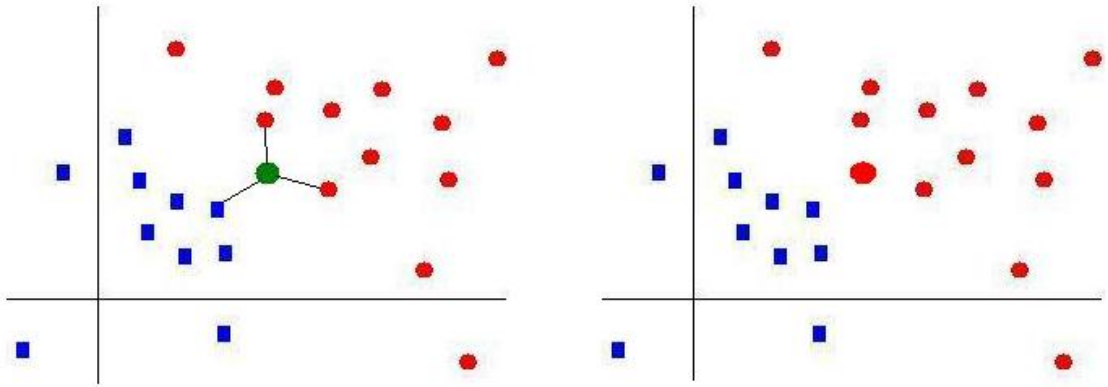
- YSA yerel bilgi işleme sistemine sahip bir tekniktir. Bu nedenle, her bir işlem birimi iş bölümü yaparak, problemin tümü yerine gerekli bölümleri ile ilgilendiği için çok karmaşık problemler kolaylıkla çözülebilmektedir.
- YSA, hataları tolere edebilen bir sistemdir. Yani, paralel bilgi işleme yaptığı için, sistemdeki herhangi bir birimin hatalı çıkması toplam bilgi işleme sisteminin çalışmasının bozulmasını önlemekte, sadece hatanın çıktığı birimin bulunduğu ağın performansını etkilemektedir.
- YSA, genelleme yeteneğine sahiptir ve öğrenmesini tamamladıktan sonra, eğitim sırasında karşılaşmadığı test örnekleri için de istenilen sonuçları üretebilmektedir.
- YSA uyarlanabilirlik özelliğine sahiptir. YSA'ların eğitiminde ağırlıklar kullanıcı tarafından ayarlanabildiği için, problemler değiştiğinde de ağırlıklar ayarlanabilmekte ve ağlar yeniden eğitilebilmektedir.
- Analiz ve tasarım kolaylığı da YSA'nın avantajları arasındadır.

Naïve-Bayes: Naïve-Bayes yöntemi, kolay uygulanabilir olduğu kadar üstün performansı ile da sınıflandırma çalışmalarında en çok kullanılan metotlardan biri haline gelmiştir⁴⁴. Bu yöntem, verileri sınıflamak için güçlü varsayımlar ile Bayes teoremine dayanılarak oluşturulmuş olasılıklı bir sınıflayıcıdır ve sınıfın kısmi özelliklerinin varlığının (ya da yokluğunun), başka bir özelliğin varlığı (ya da yokluğu) ile ilişkili olup olmadığı ile ilgilenir. Olasılıklara dayandığı için eğitici öğrenme şeklinde uygulanabilir. Birçok uygulamada Naïve-Bayes modeli için parametre kestirimi, en çok olabilirlik yöntemi ile yapılmaktadır. Dolayısıyla Bayes olasılıkları ve Bayes metodu kullanılmadan da Naïve-Bayes sınıflayıcı çalışabilmektedir. Bu yöntem, sıklıkla metin sınıflamada kullanılmaktadır³. Naïve Bayes yönteminin avantajları aşağıdaki gibidir:

- Küçük miktardaki çalışma verisi ile parametre kestirimi yapabilmektedir.
- Değişkenler bağımsız kabul edildiği için sadece değişken varyanslarının belirlenmesi yeterlidir, kovaryans matrisi girişine gerek duyulmamaktadır.

K-En yakın komşuluğu: K-En yakın komşuluğu, özellikle büyük veri tabanlarında kullanılan, en kolay sınıflandırma yöntemlerinden birisidir. Bu yöntemde sınıflama, n boyutlu uzayda sınıflanacak örneğe en yakın verilerin bulunmasına bağlıdır^{45,46,47}.

K-En yakın komşuluğu sınıflayıcıları, verilen test seti ile eğitim setinin benzerliklerinin kıyaslanmasıyla öğrenmeye dayanmaktadır³. Sınıfı belli olmayan örnekler eğitim setindeki verilere benzetilerek sınıflama yapılır. Sınıflanacak örneğe en yakın k komşu bulunur ve bu k komşu en fazla hangi sınıfa ait ise, yeni gelen örnek de o sınıfa atanır^{46,48}. k değeri kaç tane en yakın komşunun sınıflandırma için kullanılacağını belirtir. Bu değer, 1'den büyük ve genellikle tek sayı olarak seçilen pozitif bir tam sayıdır. Eğer k=1 ise nesne en yakın komşusunun bulunduğu sınıfa dahil olacaktır⁴⁹. Örneğin, k = 3 için yeni bir örnek sınıflandırılmak istenirse, bu durumda eski sınıflandırılmış verilerden en yakın 3 tanesi alınır ve bu verilerin çoğunluğu hangi sınıftaysa, yeni örnek de o sınıfa dahil edilir⁵⁰. Şekil 2.8'de, yeşil nokta ile gösterilen yeni verinin kırmızı noktalarla gösterilen sınıfa dahil edilmesi gerektiği görülmektedir. Burada, k=3 olarak seçilmiş ve yeni veriye en yakın 3 nokta bulunarak, bu noktaların çoğunluğu hangi sınıftaysa, yeni gelen veri de o sınıfa dahil edilmiştir.



Şekil 2.8: K-En Yakın Komşuluğu Yönteminde, k=3 için Yeni Verinin Sınıflandırılması

En yakın komşular bulunurken, çeşitli metrikler kullanılabilir. Sürekli değişkenlerde öklid uzaklığı aradaki mesafeyi bulmak için kullanılırken, metin sınıflandırma, örtüşme metriği gibi durumlarda Hamming uzaklık hesabı kullanılabilir⁴⁸.

K-En yakın komşuluğu yönteminde sınıflama yapmak için; k pozitif tam sayısına, sınıflandırılmış örnekler kümesine ve yakınlık ölçüsünü belirten metrik değerine ihtiyaç duyulmaktadır ve yöntem, bütün eğitim veri kümesindeki noktalara uzaklık hesabını gerektirdiği için yavaştır⁴⁶.

Genetik Algoritmalar (GA): Genetik algoritmalar, doğal seçim ilkelerine dayanan bir arama ve optimizasyon yöntemidir. Bu yöntem, problemlere tek bir çözüm üretmek yerine farklı çözümlerden oluşan bir çözüm kümesi üretir. Böylelikle, arama uzayında

aynı anda birçok nokta değerlendirilmekte ve sonuçta bütünsel çözüme ulaşma olasılığı yükselmektedir. Çözüm kümesindeki çözümler birbirinden tamamen bağımsızdır ve her biri çok boyutlu uzay üzerinde bir vektördür⁵¹.

Genetik algoritmalar problemlerin çözümü için evrimsel süreci bilgisayar ortamında taklit ederler. Diğer optimizasyon yöntemlerinde olduğu gibi çözüm için tek bir yapının geliştirilmesi yerine, böyle yapılardan meydana gelen bir küme oluştururlar. Problem için olası pek çok çözümü temsil eden bu küme genetik algoritma terminolojisinde nüfus adını alır. Nüfuslar vektör, kromozom veya birey adı verilen sayı dizilerinden oluşur. Birey içindeki her bir elemana gen adı verilir. Nüfustaki bireyler evrimsel süreç içinde genetik algoritma işlemcileri tarafından belirlenirler⁵².

Problemin bireyler içindeki gösterimi problemde değişiklik gösterir. Genetik algoritmaların problemin çözümündeki başarısına karar vermedeki en önemli faktör, problemin çözümünü temsil eden bireylerin gösterimidir. Nüfus içindeki her bireyin problem için çözüm olup olmayacağına karar veren bir uygunluk fonksiyonu vardır. Uygunluk fonksiyonundan dönen değere göre yüksek değere sahip olan bireylere, nüfustaki diğer bireyler ile çoğalmaları için fırsat verilir. Bu bireyler çaprazlama işlemi sonunda çocuk adı verilen yeni bireyler üretirler. Çocuk kendisini meydana getiren ebeveynlerin (anne, baba) özelliklerini taşır. Yeni bireyler üretilirken düşük uygunluk değerine sahip bireyler daha az seçileceğinden bu bireyler bir süre sonra nüfus dışında bırakılırlar. Yeni nüfus, bir önceki nüfusta yer alan uygunluğu yüksek bireylerin bir araya gelip çoğalmalarıyla oluşur. Aynı zamanda bu nüfus önceki nüfusun uygunluğu yüksek bireylerinin sahip olduğu özelliklerin büyük bir kısmını içerir. Böylelikle, pek çok nesil aracılığıyla iyi özellikler nüfus içersinde yayılırlar ve genetik işlemler aracılığıyla da diğer iyi özelliklerle birleşirler Uygunluk değeri yüksek olan ne kadar çok birey bir araya gelip, yeni bireyler oluşturursa arama uzayı içerisinde o kadar iyi bir çalışma alanı elde edilir. Genetik algoritmaların avantajları aşağıdaki gibidir⁵³.

- GA'nın en önemli avantajı paralel çalışmasıdır. Büyük problemler için bu önemli bir kazançtır.
- GA, çok geniş bir çözüm uzayını hızlı bir şekilde arayabilmektedir.
- GA, geniş bir problem aralığında kullanılabilir.

- GA, başlangıçta problem hakkında bilgi olmadan da problemleri çözebilmektedir.

Destek vektör makineleri (DVM): Bu yöntem hakkında detaylı bilgi bölüm 2.3’de verilmiştir.

2.2. İstatistiksel Öğrenme Teorisi (Vapnik-Chervonenkis Teorisi)

İstatistiksel öğrenme teorisi, 1960’lı yılların sonlarında Vladimir Vapnik ve Alexey Chervonenkis tarafından ortaya atılmış, ancak 1990’lı yıllara kadar sadece veri toplama ve verilerden fonksiyon tahmin etme probleminin teorik bir analizi olarak kullanılmıştır. 1990’ların ortalarında ise, bu teoriye dayalı yeni bir öğrenme algoritmasının (DVM) önerilmesiyle, istatistiksel öğrenme teorisi daha yaygın biçimde kullanılmaya başlanmıştır²⁸. DVM’lerin gerçek dünya problemlerini çözmedeki başarısı ile istatistiksel öğrenme teorisi, teorik analizler için kullanılmasının dışında, çok boyutlu fonksiyonların tahmini için pratik algoritmalar geliştiren bir araç haline gelmiştir⁵⁴. Vapnik, bir teoriye neden ihtiyaç duyulduğu ile ilgili olarak: “hiçbir şey, iyi bir teoriden daha pratik değildir” demiştir ve bu da teorinin önemini vurgulamaktadır⁵⁵.

İstatistiksel öğrenme teorisi esas olarak; bilgi kazancı, tahminler yapma, karar verme ya da bir veri setinden modeller kurma gibi çıkarımsal problemler için bir çerçeve hazırlamaktadır⁵⁵. Bu teori, istatistiksel bir bakış açısıyla öğrenme işlemini açıklamaya çalışan hesapsal öğrenme teorisinin bir biçimidir ve öğrenme ile ilgili temel kavram ve prensipleri, iyi tanımlanmış problem formülasyonlarını ve kendi içinde tutarlı matematiksel teoriyi birleştirmektedir²⁸.

İstatistiksel öğrenme teorisinin asıl amacı, dağılımdan bağımsız yöntemler ile sınıflandırma ve tahminler için hata sınırları üretmek olduğu kadar, küçük popülasyonlar arasındaki istatistiksel ilişkileri araştırmaktır^{3,56}. Bu teoride, özel veri kümeleri üzerindeki hata yerine, genelleme hatasının karakterize edilmesi amaçlanmıştır⁵⁷.

İstatistiksel öğrenme teorisinde, tahmin edilecek hem nitel hem de nicel hedefler için, tahmin modelleme probleminde yeni bir bakış açısı getirilmektedir ve bir modelin uygun olup olmadığının veya başka bir modelin daha iyi sonuç verip vermeyeceğinin kararı ile ilgili sorulara yanıt bulunabilmektedir²⁸. Klasik istatistik, doğru model formunun bilindiğini varsayıp, amacını modelin parametrelerini belirlemek olarak görürken;

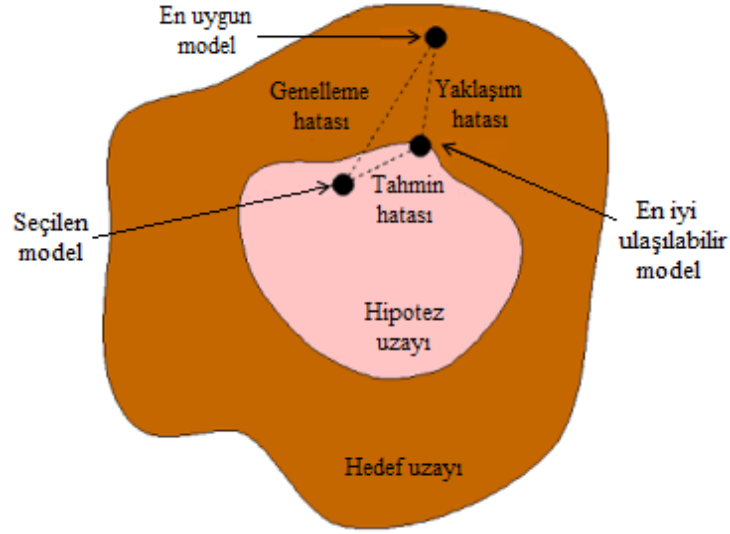
istatistiksel öğrenme teorisi, model formunun bilinmediğini kabul edilmekte ve doğru olabilecek modeller arasından en iyi modelin bulunmasını hedeflemektedir²⁸. Teorinin özünü oluşturan bu modele ilişkin iki temel varsayım söz konusudur. Bunlardan ilki; gelecek gözlemler (test verileri) ile geçmiş gözlemlerin (eğitim verileri) ilişkili olduğudur⁵⁵. Eğer ilişki yoksa tahmin imkansızdır. İkincisi ise, geçmiş ve gelecek gözlemlerin, model içerisindeki aynı dağılımdan bağımsız olarak örneklendirildiğidir. Buradaki bağımsızlık varsayımı, her yeni gözlemin maksimum bilgiyi sağlayacağı, aynı dağılım varsayımı ise gözlemlerin dayanak olayı (underlying phenomenon)(olasılık dağılımı) hakkında bilgi vereceği anlamına gelmektedir⁵⁸. Verilen bir eğitim verisi için mükemmel bir fonksiyon tanımlanabilse de, outlier (sapan değer)'lerin varlığı gelecek gözlemler üzerinde kötü bir tahmine yol açmaktadır⁵⁹.

Modellemenin amacı, hipotez uzayından hedef uzayındaki dayanak (underlying) fonksiyona en yakın (bazı hata ölçümleri bakımından) modeli seçmektir. Bunu yaparken ortaya çıkan hatalar, iki durumdan kaynaklanmaktadır¹⁶**Hata! Yer işareti anımlanmamış..**

Yaklaşım Hatası (Approximation Error): Hipotez uzayının, hedef uzayından daha küçük olmasının bir sonucudur ve dolayısıyla dayanak (underlying) fonksiyonu hipotez uzayının dışında kalabilir. Zayıf bir model uzayı seçimi, model yanlış uyumu (model mismatch) olarak ifade edilen, büyük bir yaklaşım hatasına neden olacaktır.

Tahmin Hatası (Estimation Error): Öğrenme prosedüründen dolayı, hipotez uzayından optimal olmayan modelin seçilmesinden kaynaklanan hatadır**Hata! Yer işareti anımlanmamış..**

Şekil 2.9'da hataların modellenmesine ait görsel bir ifade yer almaktadır. Burada yaklaşım hatası ve tahmin hatası gösterilmektedir. Bu hata formları birlikte genelleme hatasını oluşturmaktadır ve sonuçta riski minimize eden f fonksiyonu bulunmak istenmektedir.



Şekil 2.9: Hataların Modellenmesi

Daha iyi bir eğitim için daha çok eğitim verisinin kullanılması uygundur. Eğitim veri kümesi yeterince büyük olduğu zaman (sonsuzaya yakın), eğitim hatası çok küçük (sıfıra yakın) olabilir. Fakat eğer eğitim kümesi çok küçük ise, eğitim hatası oldukça büyük olacak ve öğrenme işleminin sonucunun güvenilir olması sağlanamayacaktır. Eğer eğitim veri kümesi outlier içeriyorsa, eğitim hatasını azaltmak için yine daha çok eğitim verisine ihtiyaç duyulur. Kecman (2001), örneklem boyutunun belirlenmesi için eğitim verisi sayısının VC (Vapnik- Chervonenkis) boyutuna oranını ölçü almıştır. Bu oran değeri 20 den küçük çıkan veri kümeleri küçük, 20 den büyük çıkanlar ise orta ölçekli veri kümeleri olarak ele alınmıştır²⁹. Buraya kadar bahsedilenler, sadece eğitim hatası düşünülerek planlanmıştır. Genelleme hatası (bilinmeyen test verisi üzerindeki hata) veya öğrenme metodunun genelleme kapasitesi hiç hesaba katılmamıştır. Eğer giriş veri kümesi yüksek boyutlu (değişken sayısı çok ise) veya verinin karakteristiğini ifade eden temel fonksiyon çok karmaşık ise, öğrenme işleminde daha çok veriye ihtiyaç duyulur. Ayrıca, boyut sıkıntısı girdi uzayının boyutunu arttıracaktır. Bu iki problemin ele alınmasıyla birlikte, yeni öğrenme makinelerinin geliştirilmesi ihtiyacı ortaya çıkmıştır ve DVM yöntemi geliştirilmiştir. DVM; yüksek boyutlu ve küçük sayıda eğitim verisinden öğrenebilen istatistiksel öğrenme teorisi çatısı altında yönlendirilmiş yeni nesil bir öğrenme metodudur²⁹.

Sonlu örnek külelerden bağımlılıkların tahmini için genel bir teori olarak ortaya atılan istatistiksel öğrenme teorisi; VC boyutu ve yapısal risk minimizasyonu prensibini göz önünde bulundurarak öğrenmeyi amaçlamaktadır.

2.2.1. Beklenen risk ve deneysel risk

Genel olarak öğrenme, olasılık temeline göre işleyen bir süreçtir. Veriden öğrenme işlemi üç temel bileşenden oluşur: üretici, danışman ve öğrenme makinesi⁶⁰. Üretici bileşeni, giriş vektörlerini (bilinmeyen bir dağılıma uygun x vektörleri) üretir. Danışman, her bir giriş vektörüne göre y eğitime cevabı değerini döndürür. x girişleri ve y cevapları öğrenme makinesinin eğitilmesi amacıyla kullanılır. Öğrenme makinesi bu giriş ve çıkış kümeleri arasındaki bağlantıyı $f(x, \alpha)$ fonksiyonlar kümesini kullanarak öğrenir²⁹. Bu fonksiyonlar kümesi istatistik literatüründe hipotez uzayı olarak adlandırılır. Veriden öğrenme problemi, danışmanın cevaplarını mümkün olan en iyi şekilde tahmin eden $f(x, \alpha)$ fonksiyonunun seçilmesidir.

En iyi tahmin modelinin seçilebilmesi için, x gözlemini doğru sınıflayıcı $Y(x) = y$ ve tahmin edilen sınıflayıcı $f(x, \alpha) = \hat{y}$ arasındaki tutarsızlığa ait bir ölçüm olmalıdır⁶¹. Yani, karar fonksiyonu $f(x, \alpha)$ 'nın kalitesini ölçmek için $L(y, f(x, \alpha))$ ile gösterilen kayıp fonksiyonu ölçümü kullanılmalıdır. Bu kayıp fonksiyonu, öğrenme makinesi tarafından üretilen yakınsama değeri ile danışmanın cevap değeri arasındaki farkı tanımlar²⁹. y 'yi tahmin etmek için $f(x, \alpha)$ 'nın kullanılmasından kaynaklanan kaybın ölçüsü, $L(y, f(x, \alpha))$ 'dir ve α , fonksiyonun parametre vektörüdür⁶². İstatistiksel araştırmalarda farklı uygulama alanları için birçok kayıp fonksiyonu kullanılmıştır. Beklenen risk veya kayıp fonksiyonunun beklenen değeri aşağıdaki formülle ifade edilebilir⁶⁰.

$$R(\alpha) = \int L(y, f(x, \alpha)) dP(x, y) \quad (2.1)$$

$$P(x, y) = P(x)P(y|x) \quad (2.2)$$

Burada $R(\alpha)$ beklenen riski, $dP(x, y)$ olasılık dağılımını göstermektedir. $L(y, f(x, \alpha))$, ise eğitim verisinden hesaplanan özel bir kayıp fonksiyondur ve bu fonksiyon 0 veya 1 değerini alabilir. İki sınıflı sınıflama problemi için kayıp fonksiyonu aşağıdaki gibi tanımlanır²⁹26:

$$L(y, f(x, \alpha)) = \begin{cases} 0, & y = f(x, \alpha) \\ 1, & y \neq f(x, \alpha) \end{cases} \quad (2.3)$$

Sınıflama probleminin amacı, eşitlik 2.1'deki beklenen riskin yani bilinmeyen bir olasılık dağılımından alınan eğitim verisi üzerindeki sınıflama hatası olasılığının minimum tutulmasıdır. Formül 2.3'e bakıldığında, danışmanın cevabı sistemin bulunduğu

fonksiyon çıkışı ile aynı olmadığı zaman bu risk değerinin artacağı söylenebilir²⁹. Yani, danışmanın verdiği y cevaplarıyla, $f(x, \alpha)$ fonksiyonunun verdiği cevapların farklı olması durumunda bu kayıp fonksiyonu için $R(\alpha)$, sınıflama hatasının olasılığını vermektedir⁶³.

Beklenen risk değerini minimum tutmaya çalışan tümevarımsal öğrenme makinesi metotları, yapısal risk minimizasyonu prensibindeki öğrenme metotları olarak bilinirler²⁹. Formül 2.1'de verilen beklenen risk değeri, bilinmeyen $P(x, y)$ dağılımını içerdiğinden doğrudan minimumlaştırılmaz. Bu durumda, beklenen risk fonksiyonunun minimumlaştırılması için, deneysel (ampirik) risk minimizasyonu (DRM) prensipleri kullanılır. $D(x_i, y_i)$, $i=1, \dots, \ell$ eğitim kümesi verilsin:

$$R_{amp} = \frac{1}{l} \sum_{i=1}^l L(y_i, f(x_i, \alpha)) \quad (2.4)$$

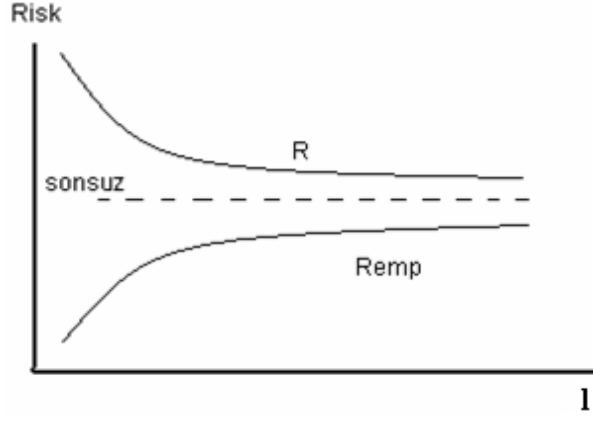
DRM prensibi, eğitim verisi sınırlı sayıda olduğunda (örneğin, örnek küçük olduğunda), $R(\alpha)$ beklenen risk değerinin küçük olmasını garanti edememektedir⁶². Deneysel risk, kayıp fonksiyonları kullanarak beklenen riske doğrusal olarak yakınsamalıdır:

$$\lim_{l \rightarrow \infty} R_{amp}(\alpha) = R(\alpha) \quad (2.5)$$

$$\lim_{l \rightarrow \infty} \min_{\alpha \in \Lambda} R_{amp}(\alpha) = \min_{\alpha \in \Lambda} R(\alpha) \quad (2.6)$$

Burada, l eğitime verilerinin sayısı sonsuza doğru artarken, deneysel risk değeri beklenen risk değerine yakınsamaktadır. Her ne kadar bu prensip deneysel riski minimum yapan fonksiyonu garanti altına almasa da, beklenen riski minimum yapan fonksiyona veya gerçek riske yakınsar⁶⁴.

Deneysel risk, eğitim verileri için yanlış sınıflama olasılığını temsil ederken, beklenen risk bilinmeyen dağılımdaki farklı örneklerin yanlış sınıflama olasılığını temsil eder. DRM genelleme prensibini temel alan metotlar, deneysel risk fonksiyonu yapısında sadece verilen eğitim kümesi ile ilgili bilgiler barındırdığı için iyi bir performansı garanti edemezler²⁹.



Şekil 2.10: Beklenen ve Deneysel Risk

Şekil 2.10’da görülen uniform yakınsama prensibine göre, l değeri artarken deneysel risk fonksiyonu beklenen risk fonksiyonuna yaklaşmaktadır. Fakat bahsedilen DRM’nin uygunluğu ve yakınsama prensibi yapısal prosedürlerin nasıl oluşturulacağı hakkında bir yol göstermez. Gerekli modeller ancak, yapısal prosedürlerin tanıtılması, Yapısal Risk Minimizasyonu (YRM) tümevarımsal prensibi ve VC teorisinin geliştirilmesi ile oluşturulabilmiştir. Bu kavramlar istatistiksel öğrenme teorisinin ve aynı zamanda DVM öğrenme metodunun da temelini oluşturmaktadır^{29,65}.

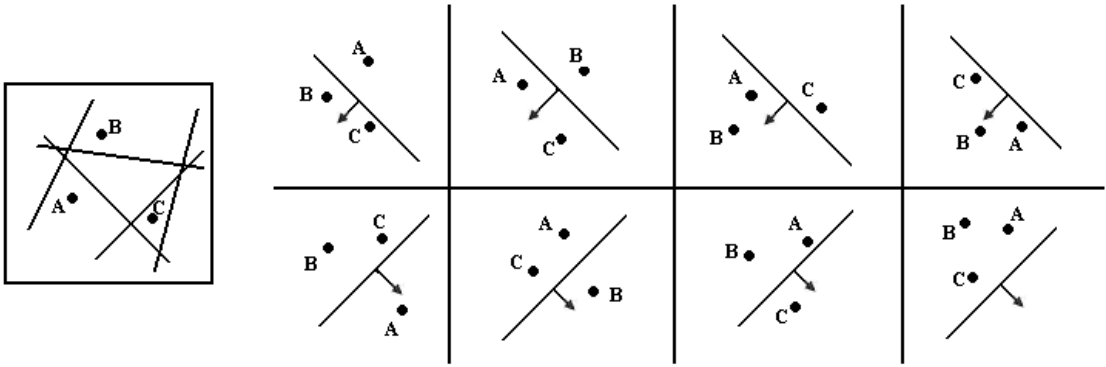
2.2.2. Vapnik-Chervonenkis (VC) boyutu

VC boyutu, $\{f(\alpha)\}$ sınıflama fonksiyonlarının bulunduğu hipotez uzayının kapasitesini ölçen skaler bir değerdir^{16,57,61}. Kapasite ise, hipotez uzayının karmaşıklığının bir ölçüsüdür ve tahmin gücü ile fonksiyon setinin esnekliğini ölçer. Fonksiyonlar kümesi için bir VC boyutu, doğrusal olarak parçalanabilen eğitim örneklerinin maksimum sayısı olarak tanımlanır. VC boyutu, istatistiksel öğrenme teorisi ile elde edilen tahmin setinin özelliğidir ve VC boyutunun temel düşüncesi, iki sınıfı doğru sınıflayabilme ve sonrasında genellemedir¹⁰.

VC boyutu bir model ailesinin karmaşıklığını ölçtüğü için, “eğer iki model veriyi aynı şekilde açıklıyorsa, o zaman daha basit olan model tercih edilmelidir” prensibinden (Occam’ın kılıcı prensibi) yola çıkılarak, veriyi aynı derecede açıklayan modellerden, daha küçük VC boyutuna sahip olan model seçilmelidir. Ayrıca VC boyutu genellenebilme (yeni veride benzer sonuçlar verebilme) kabiliyeti ile ilişkili olduğundan, daha küçük VC boyutuna sahip modelin genellenebilme performansının daha yüksek olacağı söylenebilir⁶⁶.

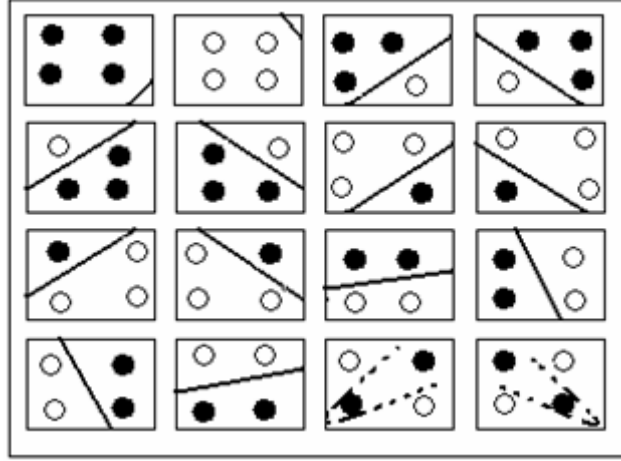
VC boyutunu hesaplamak için parçalama (shattering) kullanılmaktadır. İki sınıflı bir sınıflama problemi için, m tane gözlem, bir fonksiyon seti tarafından $\{-1, +1\}$ sınıflarına tüm mümkün 2^m yolla ayrılabilirdiğinde, örneklerin o fonksiyon seti ile parçalandığı söylenebilir^{61,65}. Başka bir ifadeyle, tüm mümkün etiketleme yolları ile hatasız bir şekilde parçalanabilen deneme noktalarının maksimum sayısı VC boyutunu verir^{3,29}. Şekil 2.2'de, iki boyutlu uzay (R^2) için basit bir VC boyut örneği gösterilmiştir. Burada; A, B ve C gözlemleri R^2 'de tanımlanmış üç gözlem olsun. Bu gözlemler, bir fonksiyon seti ile $\{-1, +1\}$ sınıflarına aşağıda verilen tüm mümkün kombinasyonlarla ayrılabilir:

Parçalama no	-1	1
1	A	BC
2	B	AC
3	C	AB
4	AB	C
5	AC	B
6	BC	A
7	ABC	\emptyset
8	\emptyset	ABC



Şekil 2.11: R^2 Uzayında Tanımlanmış A, B, C Noktaları için VC Boyutu Örneği

Şekil 2.11'de, A, B, C noktalarını ayıran 8 mümkün durum görülmektedir. Şekilde ok ile gösterilen kısım pozitif sınıfa ait örnekleri, diğer kısım ise negatif sınıfa ait örnekleri tanımlamaktadır. Ancak, dördüncü bir gözlem eklendiğinde bu şekilde bir ayırma mümkün olmayacaktır (Şekil 2.12). Dolayısıyla hatasız bir şekilde tüm mümkün yollarla sadece 3 nokta parçalara ayrılabilirdiğinden R^2 'de VC boyutu 3'tür³.

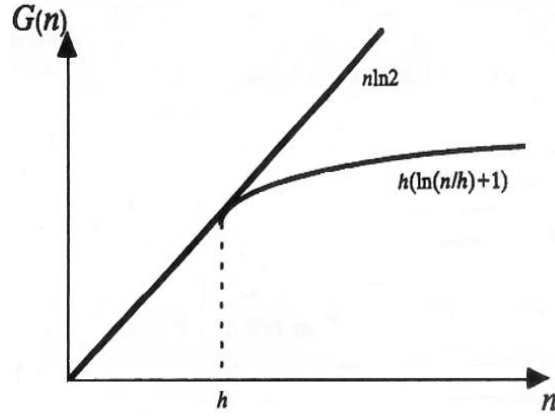


Şekil 2.12: R^2 Uzayında Dört Noktanın Ayrımı

Şekil 2.12 incelendiğinde, iki boyutlu bir uzayda, dört nokta için mümkün olan $2^4 = 16$ durumun tümünde doğrusal ayırım yapılamamaktadır ve sonuçta iki boyutlu uzayda VC boyutu 3'tür. İki boyutlu uzaydaki doğrusal olmayan sınıflandırıcılar için VC boyutu daha büyük olabilir²⁹.

Bu örnek genelleştirilse, eğer h tane nokta tüm mümkün yollarla bir fonksiyon seti tarafından doğrusal olarak sınıflara ayrılabilir, ancak $h+1$ nokta aynı fonksiyon seti tarafından ayrılamıyor ise o fonksiyon setinin VC boyutu h olur^{65,65^{67,68}}. Başka bir ifadeyle, bir fonksiyon kümesi tarafından hatasız bir şekilde tüm mümkün durumları etiketlenebilen örnek sayısı VC boyutudur ve n boyutlu bir uzayda $h = n + 1$ şeklinde ifade edilir, n tane ağırlık vektörü (w) kullanılır. Bu son ifade VC boyutunun ağırlık vektör parametrelerinin sayısını arttırdığını gösterir. Bir başka deyişle, öğrenmenin yüksek boyutlarda fazla parametreyle, düşük boyutlarda ise az parametreyle gerçekleşeceği, yani parametre sayısı arttıkça VC boyutunun artacağı düşünülmektedir. Ancak Levin ve Danker tarafından bir parametreye sahip bir öğrenme makinesinin sonsuz sayıda VC boyutuna sahip olduğu gösterilmiştir^{3,29}. **Hata! Yer işareti anımlanmamış..**

Hipotez uzayının karmaşıklığını veya H hipotez uzayının m sayıda gözlem üzerindeki açıklayıcı gücünü ölçen VC boyutu, bir yükselme fonksiyonu kullanılarak hesaplanabilir. Vapnik ve Chervonenkis (1968), n sayıda gözlem için logaritmik bir yükselme fonksiyonu önerdiler ve bu fonksiyonun ya doğrusal bir fonksiyon olduğunu ya da n örnek sayısının logaritmik bir fonksiyonu tarafından sınırlandırıldığını kanıtladılar⁶⁹.

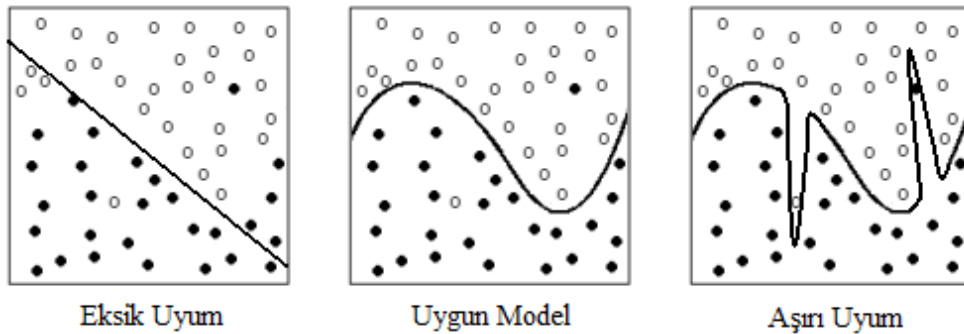


Şekil 2.13: Yükselme Fonksiyonu Grafiği

Şekil 2.13'deki yükselme fonksiyonu grafiğinde gösterilen, $n = h$ noktasında yükselme fonksiyonu eğrisinin düşüşe geçtiği noktaya VC boyutu denir ve h ile ifade edilir. Eğer VC boyutu sonlu (finite) bir değer ise, yükselme fonksiyonu yeterince büyük örnekler için doğrusal olarak artmaz ve logaritmik bir fonksiyon tarafından sınırlandırılır. VC boyutunun (h) sonluluğu; hızlı yakınsama oranı ve dağılımdan bağımsız olarak hesaplanan DRM'nin tutarlılığı için gerek ve yeter şartları sağlar. Diğer taraftan, eğer herhangi bir n için sınır doğrusal olarak kalıyorsa; $G(n) = n \ln 2$ olur⁶⁹.

2.2.2.1. Aşırı uyum (overfitting) ve eksik uyum (underfitting) kavramları

Hipotez uzayının karmaşıklığını tanımlayan VC boyutu, sınıflandırma yapılırken model karmaşıklığından dolayı meydana gelebilen modelin veriye aşırı uyumu ve eksik uyumu problemlerinin analizinde de yarar sağlamaktadır. Şekil 2.14'te, üç farklı karar sınırı kullanarak çözülmüş olan bir ikili sınıflandırma örneği görülmektedir.



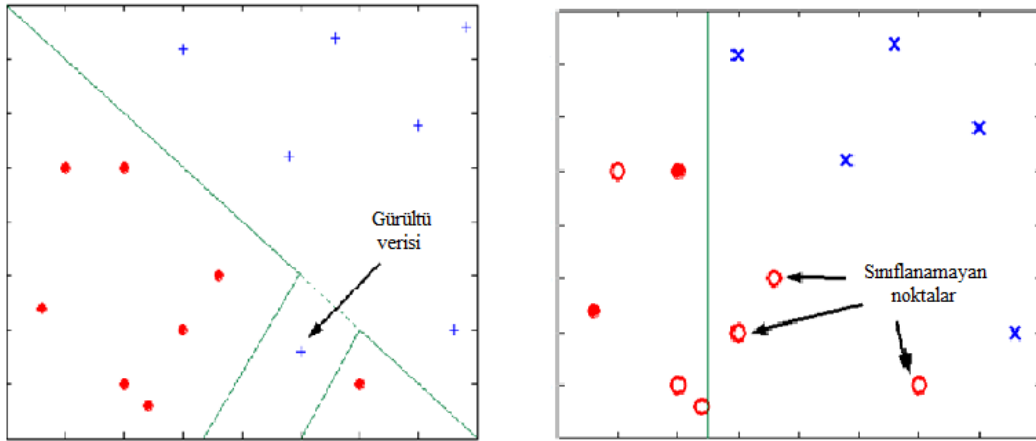
Şekil 2.14: Üç Farklı Model Kullanılarak Oluşturulmuş Bir İkili Sınıflandırma Örneği

Burada görülen üç model arasından yapılacak en iyi seçim, ortadaki gibi olacaktır. Eğer en soldaki model seçilirse, karşılaşılabilecek her sınıf çifti doğrusal olarak ayrılabilir kabul edilecektir. Fakat bu durum sunulan veriden kaynaklanmayacak, sınıflandırıcıya zorla

kabul ettirilen bir eğilim olacaktır. Çok sayıda yanlış sınıflandırılan noktanın göz ardı edildiği bu duruma “eksik uyum” denilmektedir. Eğitim verisine eksik uyum gösterme, düşük tahmin performansına sahip çok basit modeller üretilmesine sebep olacaktır⁷⁰. Eksik uyum; modelin veri yapısını öğrenmede yetersiz olduğunu yani tam öğrenemediğini göstermektedir⁷¹.

En sağdaki modelde sınıflandırma fonksiyonu, bütün eğitim noktalarını doğru biçimde ayırabilmektedir ancak burada birçok özelliği göz önüne alan karmaşık bir model yapısı söz konusudur. Schölkopf ve Smola (2002)’nin da belirttiği gibi, yeterli derecedeki bir fonksiyonla, “veriye her zaman mükemmel bir şekilde uyum sağlanabilse de, gerçek model, ölçümlerin hassaslığına bağlı olarak büyük dalgalanmaları da beraberinde getirecek ve model, değişimlerden çok fazla etkilenecektir”⁷². Bu değişimlerin, test örneklerinin düzenliliğinden kaynaklanıp kaynaklanmadığı durumu ise açık değildir. Örneğin model, karşıt sınıflar içinde gömülü biçimde bulunan özel bölgeleri de ayırabilecek hale gelmiş olabilir. Bu duruma “aşırı uyum” denir. Yani model, eğitim verisine aşırı uyum sağlamıştır. Aşırı uyum sorunundan, model karmaşıklığı kontrol altına alınarak kaçınılabılır.

Aşırı uyum, sapan değerlerin varlığından veya temsili örneklerin eksikliğinden kaynaklanabilmektedir. Şekil 2.15’te aşırı uyuma neden olan bu sorunlar gösterilmektedir.



Şekil 2.15: Aşırı Uyumun Meydana Gelme Sebepleri

Eğitim hatası, eğitim verilerindeki yanlış sınıflandırma sonucu oluşan hataların sayısıdır. Genelleştirme hatası ise, önceden görülmeyen veri modelinin tahmin edilmeyen hatasıdır. İyi bir model, düşük eğitim hatasına ve düşük genelleştirme

hatasına sahip olmalıdır. Çok fazla parametreye sahip, yani çok özellik kullanılarak oluşturulan bir model, eğitim setine iyi uyum sağlayacak, ancak eğitim setinin seçimine duyarlı olacaktır. Bu durum, *aşırı uyumdan kaynaklanan genelleme hatasına* neden olur. Bunun tersine, çok az parametreye sahip, yani az özellik kullanılarak oluşturulan bir model, eğitim setine iyi uyum sağlamayacak, ancak farklı eğitim setleri için tutarlı olacaktır. Bu durum ise, *eksik uyumdan kaynaklanan genelleme hatasına* neden olur⁷¹.

Eksik uyum ve aşırı uyum problemleri arasında bir denge her zaman mevcuttur. Bu denge noktası, düzenleyici bir parametre olan VC boyutu yardımıyla sağlanabilir²⁹.

2.2.3. Yapısal risk minimizasyonu

YRM, sonlu sayıda eğitim verisi ile öğrenmede kullanılacak model seçimi için yararlı olan tümevarımsal bir prensiptir^{64,73,74}. YRM, kapasite kontrolünün genel modelini tanımlar ve hipotez uzayının karmaşıklığı (VC boyutu) ile eğitim setinin uyum kalitesi (deneysel risk) arasındaki dengeyi (tradeoff) sağlar⁷⁵. YRM tümevarımsal prensibi ve VC boyutu, istatistiksel öğrenme teorisinde önemli bir yere sahiptir.

Olasılıkta uniform yakınsama, öğrenme makineleri için beklenen risk ile deneysel risk arasındaki sapma değerine bir sınır getirir⁷⁶. İstatistiksel öğrenme teorisinde bu sınırlar ancak, yakınsama fonksiyonları kümesini daha küçük bir hipotez uzayında kısıtlayarak ve aynı zamanda bu yakınsama fonksiyonlarının esneklik ve karmaşıklıklarını da kontrol ederek minimum tutulabilir.

YRM prensibi ve VC teorisindeki temel fikir; çok sayıda aday model arasından beklenen risk veya genelleme hatasını minimum yapacak doğru oranda karmaşıklığa sahip (deneysel riski ve yakınsama fonksiyonlarının karmaşıklığını aynı anda minimum yapan) modelin seçilmesidir. Böylece, seçilen en iyi öğrenme makinesi verilen eğitim kümesi için en uygun kapasiteye sahip olacaktır^{29,77}.

Klasik yöntemler risk minimizasyonu için DRM'yi kullanırken, DVM yöntemi YRM'yi kullanmaktadır ve yöntemin temel farklılığı da buradan kaynaklanmaktadır. Bu kurala göre genelleme hatası; deneysel hata, denemedeki örnek sayısı ve kapasiteden etkilenmektedir^{3,65}.

DRM sadece eğitim verisinin kendisindeki hatayı minimize etmeye dayanmaktadır. Eğer, eğitim verisi dağınık ve/veya temel dağılımı temsil etmiyorsa, sistem zayıf

eğitilecek ve bu yüzden sınırlı bir performansa sahip olacaktır. Eğitim setindeki hatayı minimize eden DRM'nin aksine, YRM verinin dağınıklığını göz önünde bulundurur ve riskin üst sınırın minimize eder **Hata! Yer işareti tanımlanmamış.**⁷⁸. İstatistiksel öğrenmenin amacı, DVM'nin genelleme yeteneğini arttırmaktır ve bu fark bunu sağlamaktadır.

Doğrusal olmayan modellerin VC boyutunu tahmin etmenin zorluğuna yapısal risk minimizasyonu yoluyla karşılık bulunabilir. YRM, bilinmeyen veride düşük hata oranı elde etmek için, verilen bir modelin gelecek hatası üzerinde analitik bir üst sınır sağlamaktadır. Bu üst sınır;

$$R(\alpha) \leq R_{amp}(\alpha) + \Phi(h, n, \delta) \quad (2.7)$$

şeklinde tanımlanır. YRM, 2.7 eşitsizliğinin sağ tarafını minimize etmeye çalışır. Burada, VC güvenilirliği veya kapasite olarak tanımlanan $\Phi(h, n, \delta)$ terimi,

$$\Phi(h, n, \delta) = \sqrt{\frac{h \ln\left(\frac{2n}{h} + 1\right) - \ln\left(\frac{\delta}{4}\right)}{n}} \quad (2.8)$$

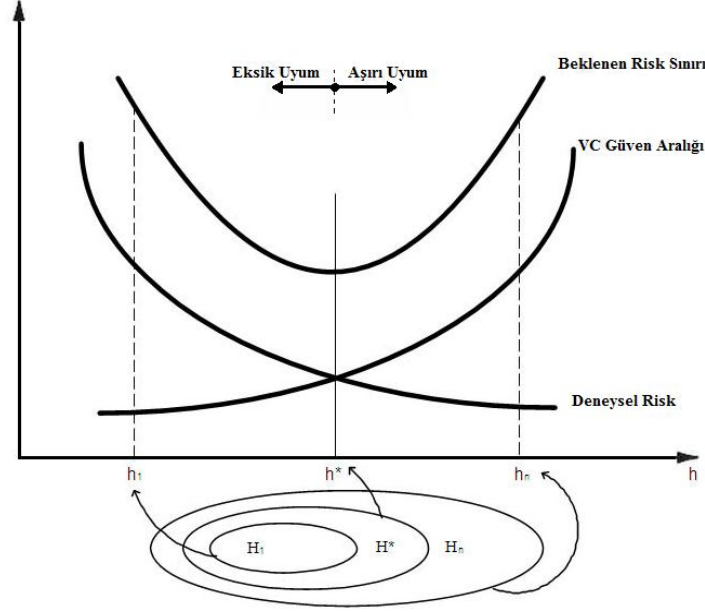
olarak ifade edilir ve son olarak beklenen risk için üst sınır,

$$R(\alpha) \leq R_{amp}(\alpha) + \sqrt{\frac{1}{n} \left(h \ln\left(\frac{2n}{h} + 1\right) + \ln\left(\frac{4}{\delta}\right) \right)} \quad (2.9)$$

halini alır. Formül 2.9'daki $R(\alpha)$ beklenen risk değeri, $R_{amp}(\alpha)$ deneysel risk değeri, n eğitim verilerinin sayısı, h VC boyutudur ve $n > h$ 'dir. δ ise, $0 < \delta < 1$ aralığında değer alabilen yeterli küçüklükte sabit bir ifadedir ve Formül 2.9'da verilen sınır $1 - \delta$ olasılıkla oluşturulmuştur. Burada, n/h oranı arttıkça, VC güvenilirliği azalır ve deneysel risk beklenen riske yaklaşır³.

Deneysel risk ve gerçek risk makine öğrenmesinin deneme süreci tarafından belirlenen özel bir fonksiyona bağlıyken, VC güvenirlilik terimi fonksiyon sınıflarının seçimine bağlıdır. Eğer seçilen fonksiyonlar kümesinin alt kümesi olan fonksiyon bulunursa, bu alt küme için risk sınırı minimize edilebilir⁷⁹. YRM'de, artan VC boyutuna (karmaşıklık) sahip hiyerarşik bir hipotez uzayı söz konusudur ve fonksiyonlar sınıfı, iç içe geçmiş alt kümeler bölünerek yapı (structure) oluşturulur³. Yani fonksiyonlar sınıfı için, $H_1 \subset H_2 \subset \dots \subset H^* \subset \dots \subset H_n$ şeklinde iç içe geçmiş alt kümeler tanımlanmıştır

ve bu alt kümelere ait VC boyutları, $h_1 < h_2 < \dots < h^* < \dots < h_n$ şeklindedir. VC boyutu kavramına göre kurulan bu tip modellerin hedefi, eğitime hatası ve model karmaşıklığını dengelemek amacıyla uygun sayıda bağımsız parametre içeren alt modellerden birisinin bulunmasıdır²⁹26.



Şekil 2.16: Beklenen Risk Fonksiyonunun Yapısı

Şekil 2.16'da, H_1 fonksiyonlar sınıfı en az karmaşıklığa, dolayısıyla en küçük VC boyutuna ve en yüksek deneysel risk değerine sahipken, H_n fonksiyonlar sınıfı en yüksek karmaşıklığa, dolayısıyla en büyük VC boyutuna ve en küçük deneysel risk değerine sahiptir. Beklenen risk (genelleme hatası) sınırını minimum yapan optimum fonksiyon en basit ve en karmaşık olan fonksiyonlar sınıfının arasında bir karmaşıklığa sahiptir ve deneysel risk ile VC güven aralığının birlikte dengede oldukları, yani fonksiyonlar sınıfının mümkün olduğu kadar basit tutulabildiği ve eğitim hatasının da minimuma ulaştığı yerde bu fonksiyon optimumdur⁶⁹.

2.3. Destek Vektör Makineleri

2.3.1. Destek vektör makinelerinin tanımı

DVM, 1960'lı yılların sonunda Vladimir Vapnik ve Alexey Chervonenkis tarafından geliştirilmiş, temel olarak istatistiksel öğrenme teorisine dayanan bir makine öğrenmesi yöntemidir⁸⁰. Ancak, ilk başarılı uygulamaları 1990'lı yıllarda yapılmıştır. Bu durumun nedeni, pratik uygulamalarda DVM'nin uygun bir yöntem olmadığı düşünülmesi ve

ileri bilgisayar teknolojisi gerektirmesidir⁸¹. Vladimir Vapnik, Bernhard Boser ve Isabella Guyon tarafından 1992 yılında yapılan yayının ardından DVM ilgi odağı haline gelmiş ve gerçek dünya problemlerini sınıflamada çok iyi sonuçlar vermesi nedeniyle, birçok kullanıcı tarafından tercih edilmeye başlanmıştır²⁹.

DVM metodu, son yıllarda özellikle veri madenciliğinde değişkenler arasındaki örüntülerin bilinmediği veri setlerindeki sınıflama problemleri için sıklıkla kullanılmaktadır. Bu metod, temelde iki sınıflı problemlerin çözümünde doğrusal bir sınıflayıcı olarak düşünülmüş, daha sonra doğrusal olarak ayrılamayan veya çok sınıflı sınıflama problemlerinin çözümüne de geliştirilerek, bu problemlerin çözümünde de yaygın olarak kullanılmaya başlanmıştır³. DVM'nin uygulama alanları günümüzde çeşitlilik kazanmakla birlikte ilk olarak optik karakter tanıma alanına uygulanmış, daha sonra Vapnik ve Chervonenkis, Boser, Guyon, Cortes, Schölkopf tarafından bugünkü gelişmiş seviyesine çıkarılmıştır⁷⁴. Son zamanlarda ise örüntü tanıma, yüz bulma ve tanıma, veri madenciliği, dil yapısını inceleyen mantıksal programlamalar, uçak alt basınç profillerinin modellenmesi, biyoloji ve diğer bioinformatik uygulamalarda, gen analizlerinde ve proteinlerin sınıflandırılmasında DVM yöntemi kullanılmaktadır¹⁰.

DVM, klasik çok katmanlı perceptron yapay sinir ağları yöntemleri ile yakın akrabalık gösteren bir yöntemdir. DVM'nin dayandığı ana fikir, verileri sınıflara en uygun biçimde ayırabilecek maksimum marjlinli bir hiperdüzlem bulmaktır. Diğer birçok doğrusal öğrenme makinesi de bu mantıkla çalışmasına karşın DVM, bu yöntemlerden farklı olarak, yapısal risk minimizasyonu ile test edilecek veriler için yanlış sınıflandırma olasılığını en aza indirecek bir çözüm sunabilir¹³.

DVM literatüründe, tahmin edici veya bağımsız (predictor) değişkene *doğal nitelik (attribute)*, optimum hiperdüzlemi belirlemek için kullanılan dönüştürülmüş doğal niteliğe *belirleyici nitelik (feature)* ve bir deneği (gözlemi) tanımlayan belirleyici nitelik setine ise, *vektör* denilmektedir⁸².

DVM'de, verilerin doğrusal olarak ayrılabilmediği veya doğrusal olarak ayrılamadığı şeklinde iki durum söz konusudur. Doğrusal olarak ayrılabilen veri yapılarında, iki sınıf birbirinden bir doğru ile ayrılabilirken, doğrusal olarak ayrılamayan verileri sınıflandırmak için, veri çeşitli yollarla farklı bir boyuta taşınır ve burada en iyi ayırıcı hiperdüzlem bulunarak veri sınıflandırılır. DVM'de bir sınıflandırıcı, bir fonksiyonlar kümesinin kapasitesini tanımlayan, minimize edilmiş VC boyutuyla yaratılır. Eğer VC

boyutu düşükse, sınıflandırma hatasının beklenen olasılığı da düşüktür. Parametrelerin ince ayarını gerekli bulmamasıyla DVM'ler, diğer istatistiksel örüntü sınıflandırma yöntemlerine kıyasla daha iyi genelleme kabiliyeti sergilerler²⁸.

DVM, basit fikirler üzerine kurulmuş olması ve pratik uygulamalarda yüksek performans göstermesi bakımından oldukça kullanışlıdır. DVM'de kullanılacak örnek sayısı önemli değildir. DVM eğitim esnasında görülmemiş verileri de sorunsuz olarak sınıflandırabilmektedir ve bu durum DVM'nin genelleştirebilme yeteneğini göstermektedir. Genelleştirebilme özelliği DVM'yi diğer tekniklere göre (YSA, karar ağacı vs..) iyi bir alternatif yapmaktadır¹⁰.

DVM'nin işleyişinde optimizasyon ve Lagrangian teorilerinden faydalanılmaktadır. Optimizasyon teorisi matematiğin, problemlerin çözümlerini karakterize etme üzerine yoğunlaşan ve onlar için verimli algoritmalar geliştiren bir dalıdır. Dolayısıyla makine öğrenmesi problemi optimizasyon teorisinin çerçevesi içinde analiz edilecek hale çevrilmiştir. Optimizasyon teorisi, sadece algoritmik teknikler sağlamakla kalmaz, çözüm olarak verilen bir fonksiyon için gerekli ve yeterli şartları da açıklar.

Langrangian teorisinin amacı ise, başlangıçta hiçbir eşitsizlik sabiti olmayan bir optimizasyon probleminin çözümünü karakterize etmektir⁵⁷. Bu teorisin ana konsepti Langrange çarpanları ve Langrangian fonksiyonudur. Bu metot 1797'de mekaniksel problemler için Langrange tarafından geliştirilmiştir.

$f(x)$ amaç fonksiyonlu ve $h_i(x) = 0, i = 1, \dots, m$ eşitlik sabitli verilen bir optimizasyon problemi için,

$$L(x, \alpha) = f(x) + \sum_{i=1}^m \alpha_i h_i(x) \quad (2.10)$$

Langrangian fonksiyonudur ve α_i katsayıları Langrange çarpanları olarak adlandırılır.

Bu teorileri de içeren DVM, istatistiksel öğrenme teorisinde iyi şekilde kurulmuş bir teoriye sahiptir ve sınıflandırma ile regresyon problemlerine çözüm için uygundur. Vapnik'in teorisi eğitim kümesindeki hata ile VC boyutuna göre ifade edilen hipotez uzayının karmaşıklığının her ikisini de küçükleyen çözümün bulunduğunu göstermektedir.

DVM algoritması öğrenme teorisinin ve pratiğinin kesiştiği bir uygulamadır. Gerçek dünya uygulamaları, teorik olarak çözülmesi zor ve karmaşık olan uygulamalardır.

DVM algoritması bu iki zorluğu da basitçe kaldırabilir ve karmaşık modellere de çözüm getirebilir.

2.3.2. Destek vektör makinelerinin tarihçesi

DVM'nin önemi, son yirmi yıllık bir süreçte kavranmış olsa da, DVM yaklaşık 50 yıla yakın bir gelişim sürecine sahip bir makine öğrenmesi yöntemidir. Ancak, DVM'nin temeli daha gerilere gidilerek, 1936 yılında R.A Fisher tarafından tanımlanan ilk örüntü tanıma algoritması Fisher'in doğrusal diskriminant yöntemine dayandırılabilir. Bu yöntemin geliştirilmesinden sonra, 1957'de DVM'nin asıl gelişim sürecinin başlangıcı olarak görülebilecek olan ilk doğrusal sınıflayıcı *perceptron* Frank Rosenblatt tarafından ortaya atılmıştır. Perceptronun ardından, 1963 yılında Vapnik ve Lerner genelleştirilmiş betimleme algoritmasını (generalized portrait algorithm) tanıtmışlardır. DVM'nin uyguladığı algoritma da, Vapnik ve Lerner tarafından ortaya atılan bu algoritmanın doğrusal olmayan genelleştirilmiş halidir. Aizerman, Braverman ve Rozonoer 1964 yılında, doğrusal olmayan DVM'nin temelini oluşturan çekirdek fonksiyonların geometrik yorumları ile ilgili çalışmalar yapmışlardır. Yine aynı yıl (1964) içinde Vapnik ve Chervonenkis, genelleştirilmiş betimleme algoritmasını daha da geliştirerek, en iyi hiper düzlem tekniği ile örüntü tanıma algoritmasını sunmuşlardır. En iyi hiperdüzlem fikrinin ortaya atılmasından sonra, Cover 1965 yılında büyük marjlinli hiperdüzlem yaklaşımını öne sürmüştü ve bu yaklaşım Mangasarian (1965) tarafından örüntü tanımlamada kullanılmıştır. Daha sonra, Duda ve Hart (1973) ve Anlauf ve Biehl (1989) de büyük marjlinli hiperdüzlem yaklaşımı ile ilgili çalışmalar yapmışlardır. 1968 yılında Smith tarafından, DVM'nin sınıflama başarısını etkileyen önemli etkenlerden olan gevşek değişkenlerle ilgili çalışmalar yapılmıştır. DVM'nin temelini oluşturan istatistiksel öğrenme teorisinin tanınması, 1974 yılında Vapnik ve Chervonenkis ile başlamış ve 1979 yılında bu teorisin Vapnik tarafından geliştirilmesiyle DVM ilgi çekici bir sınıflama yöntemi haline almıştır. 1979 yılında Vapnik ve Chervonenkis'in Rusça olan çalışmaları ilk olarak Almancaya çevrilmiştir. Daha sonra 1982 yılında Vapnik, 1979'da yaptığı çalışmasını "*Estimation of Dependences Based on Empirical Data*" ismiyle İngilizceye çevirerek kitap haline getirmiş ve istatistiksel öğrenme teorisi ile DVM yöntemi bu sayede daha geniş bir çevreye ulaşmıştır. Bu gelişmelerden sonra, Poggio ve Girosi (1990) ve Wahba (1990) çekirdek fonksiyonlarla ilgili çalışmalar yapmışlardır. Bennett ve Mangasarian (1992), 1968'de Smith'in gevşek değişkenler ile ilgili yaptığı çalışmayı geliştirmişlerdir.

Bunlarla birlikte, Boser, Guyon ve Vapnik tarafından 1992 yılında yapılan bir çalışmayla, DVM bugünkü halini almıştır. DVM'lerle ilgili ilk detaylı çalışma ise, 1998 yılında Scholkopf ve Burges tarafından yapılmıştır. Daha sonra, 2005 yılında Wang ve Kecman, 2006 yılında Tan, Steinbach ve Kumar tarafından çeşitli çalışmalar yapılmıştır.

2.3.3. Destek vektör makinelerinin uygulama alanları

Sınıflama için sıklıkla kullanılan DVM, birçok alanda kabul görmüş bir yöntemdir. DVM'nin kullanım alanları arasında; görüntü sınıflandırma, metin sınıflandırma (Joachims, 1997), nesne tanıma (Blanz ve ark., 1996), el yazısı tanıma (Cortes ve Vapnik, 1995; Scholkopf, Burges ve Vapnik, 1995; Scholkopf, Burges ve Vapnik, 1996; Burges ve Scholkopf, 1997), ses tanıma (Schmidt, 1996) ve yüz tanıma (Osuna, Freund ve Girosi, 1997) gibi çeşitli örüntü tanıma uygulamaları yer almaktadır⁶⁸. Sağlık alanında da kullanımı yaygınlaşan DVM, bu alanda temel olarak; kanser morfolojisinde, tedavi başarısının ve ilgili genin belirlenmesinde, çeşitli hastalıkların teşhisinde kullanılmaktadır.

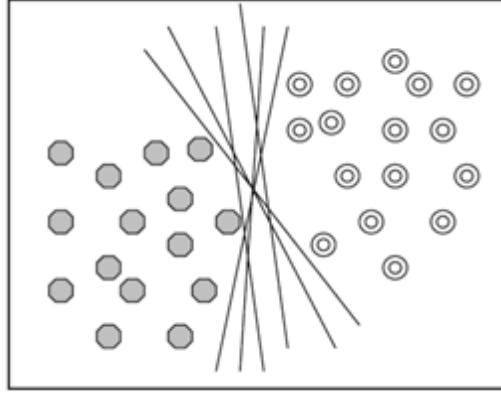
DVM'nin kullanım alanlarından optik karakter tanıma (OKT) alanına el yazısı tanıma, metin sınıflama; biyometri alanına yüz tanıma, parmak izi tanıma ve ses tanıma; tanı koyma alanına tıbbi tanımlar (x-ray ve EKG analizleri) ve makine tanımları (defolu mal bulma); askeriyedeki uygulama alanlarına ise, otomatik hedef tanıma, saldırı bulma ve resim segmentasyonu ve analizi uygulamaları örnek olarak verilebilir.

2.3.4. Destek vektör sınıflaması

2.3.4.1. Doğrusal destek vektör makineleri ile sınıflama

2.3.4.1.A. Verilerin tamamının doğrusal olarak ayrılabilirdiği durum (sert marjin)

Destek vektör makinelerinin en basit ve ilk olarak tanıtılan uygulaması 2 sınıflı doğrusal olarak ayrılabilen sınıflama problemlerinin çözümüdür³. DVM'lerin bu şekli sadece doğrusal olarak ayrılabilen belirleyici nitelik uzayı için geçerlidir ve bu nedenle çoğu gerçek dünya probleminde kullanılamamaktadır^{28,57}. Doğrusal olarak ayrılabilen sınıflama problemleri, DVM'nin temel yapı taşını oluşturmasının yanı sıra, bu çeşit öğrenen makineyi karakterize eden anahtar özellikleri sergilemektedir. Tanımı çok daha gelişmiş sistemleri anlamak için önemlidir⁸³.



Şekil 2.17: Doğrusal Olarak Ayrılabilen İki Sınıflı Sınıflama Problemi

Şekil 2.17’de gösterilen iki sınıflı veriler doğrusaldır ve bu verileri birbirinden direkt olarak ayırabilen birçok hiperdüzlem (doğru) çizilebilmektedir^{68,80}. Ancak DVM’nin amacı, bilinmeyen veri seti ile karşılaştığında sınıflama hatasını en küçük yapacak hiperdüzlemi seçmektir. Bu hiperdüzlem, iki örnek grubuna eş uzaklıkta olacaktır. Bunun için maksimum marjınlı hiperdüzlem tekniği önerilmiştir⁶⁰.

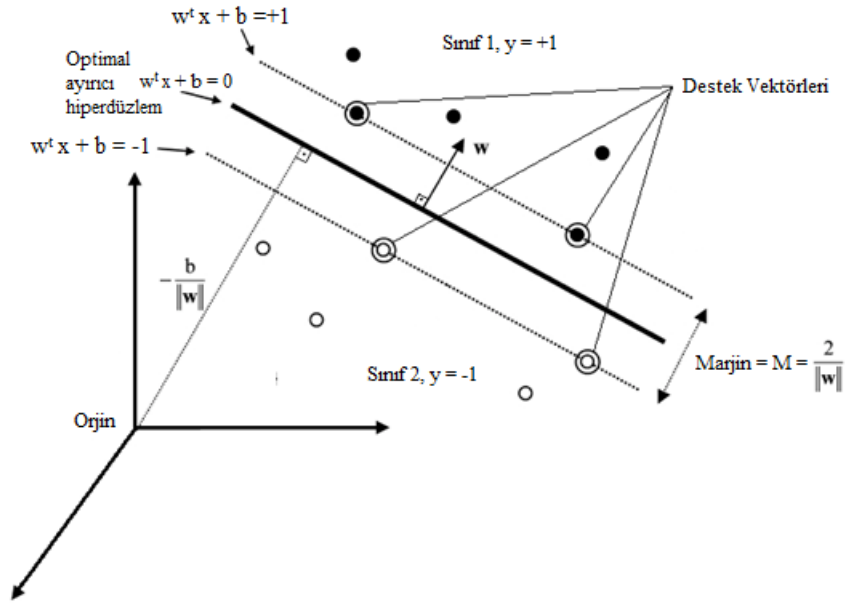
Şekil 2.17’deki hiperdüzlemler;

$$\langle w \cdot x \rangle + b = 0 \quad (2.11)$$

şeklinde formüle edilir.

Formülde w hiperdüzlemin normalini (ya da ağırlık vektörü) ve b yanlılığı (bias) gösteren bir değerdir. x ise $\langle w \cdot x \rangle + b = 0$ hiperdüzlemi üzerinde herhangi bir noktadır. Burada, $\langle w \cdot x \rangle$ bir iç çarpımı göstermektedir ve $\langle w \cdot x \rangle = w^t x$ şeklinde ifade edilebilir.

Doğrusal olarak ayrılabilen iki sınıflı bir sınıflandırma probleminde DVM’nin eğitimi için k sayıda örnekten oluşan eğitim seti; $\{x_i, y_i\} \ i=1,2,\dots,k$ için, sınıflar $y_i \in \{-1, +1\}$ ve girdi vektörü (doğal nitelikler) $x_i \in R^d$ olacak şekilde bir veri seti göz önüne alınsın. Özellik vektörlerinin ayrılabilir olduğu ve doğrusal bir karar sınırı tarafından ayrılabilirdiği varsayılmaktadır. Tanımlanan veri kümesi için, iki veri sınıfını ayırabilen, hiperdüzlemler kümesi bulunmaktadır. Doğrusal ayrılabilir veriler için, veri kümesini verilen etiketlere göre bir hiperdüzlemlerle ayırıp, aynı sınıfa ait bütün veri noktalarını hiperdüzlemin aynı tarafında bırakmak mümkündür^{28,59}.



Şekil 2.18: Verilerin Tamamının Ayrılabilirliği İçin Doğrusal Ayrıcı Hiperdüzlem

$\frac{|b|}{\|w\|}$: hiperdüzlemden orjine olan dik uzaklık,

$\|w\|$: w 'nin Öklid normu olarak ifade edilir.

Şekil 2.18'de görülen, sınırı maksimuma çıkararak en uygun ayrımı yapan hiperdüzlem "optimum ayırıcı hiperdüzlem" ve sınır genişliğini belirleyen noktalar ise "destek vektörleri (dv)" olarak adlandırılır. Optimum hiperdüzlemin belirlenebilmesi için bu düzleme paralel ve sınırlarını oluşturacak iki hiperdüzlemin belirlenmesi gerekir⁸⁴. Şekil 2.18'de kesikli çizgilerle gösterilen ve ayırıcı hiperdüzleme paralel olarak çizilmiş eşit uzaklıkta iki hiperdüzlem bulunmaktadır. Bu iki hiperdüzlem arasındaki uzaklığa "marjin" adı verilmektedir. Bu hiperdüzlemlerin fonksiyon gösterimleri aşağıdaki gibidir:

$$w^t x^+ + b = +1, \quad y_i = +1 \text{ için} \quad (2.12)$$

$$w^t x^- + b = -1, \quad y_i = -1 \text{ için} \quad (2.13)$$

Bu iki eşitlik birbirinden çıkarılırsa, $w^t (x^+ - x^-) = 2$ eşitliği elde edilir. Bu eşitliğin her iki tarafı $\|w\|$ 'ye bölünürse; $\frac{w^t (x^+ - x^-)}{\|w\|} = \frac{2}{\|w\|}$ elde edilir. Burada $\|w\| = \sqrt{\sum_{i=1}^n w_i^2}$, w 'nin Öklid normudur ve dik uzaklıkların hesabı için kullanılmaktadır. $\frac{w^t}{\|w\|}$ ise, birim vektördür (uzunluğu 1 olan vektör). Böylece marjin uzunluğu,

$$M = \|x^+ - x^-\| = \frac{2}{\|w\|} \quad (2.14)$$

olarak bulunur.

Veya marjin değerine şöyle de ulaşılabilir. Ayırıcı hiperdüzlem ve bir x örnek verisi arasındaki mesafe $|w^t \cdot x + b| / \|w\|$ 'dir⁶⁵. Burada da $\|w\|$, w 'nin Öklid normudur ve dik uzaklıkların hesabı için kullanılmaktadır. Bu durumda, Formül (2.12) ve Formül (2.13)'de verilen, sırasıyla sınıf 1 ve sınıf 2'yi ayıran eşitliklerden, bu hiperdüzlemlerin ayırıcı hiperdüzleme olan uzaklıkları $|w^t \cdot x + b| / \|w\|$ eşitliğinden, $1/\|w\|$ olarak bulunur. Dolayısıyla, her iki hiperdüzlemin ayırıcı hiperdüzleme olan uzaklıklarının toplamı ile elde edilen marjin değeri aşağıdaki gibidir:

$$\text{Marjin} = M = \frac{1}{\|w\|} + \frac{1}{\|w\|} = \frac{2}{\|w\|} \quad (2.15)$$

Ayırıcı hiperdüzlem, veri örneklerinin ayrılmasını tanımlayan aşağıdaki koşulları yerine getirir⁶⁵:

$$w^t x_i + b \geq 1, \quad y_i = +1 \text{ için} \quad (2.16)$$

$$w^t x_i + b \leq -1, \quad y_i = -1 \text{ için} \quad (2.17)$$

Formül (2.16) ve (2.17) tek formül olarak ifade edilecek olursa:

$$y_i(w^t x_i + b) \geq 1, \quad i = 1, 2, \dots, n \quad (2.18)$$

Verilen eğitim verisi için tüm ayırıcı hiperdüzlemler bu formda gösterilebilir⁶⁵.

(2.18) eşitsizliğini sağlayan hiperdüzlemin iki tarafındaki en yakın örneklere olan dik uzaklıkları toplamı marjindir ve optimum ayırıcı hiperdüzlem, marjini maksimum yapan hiperdüzlemdir⁸⁰. Böylece, optimum ayırıcı hiperdüzlemin bulunması problemi, Formül (2.14)'te verilen $M = \frac{2}{\|w\|}$ marjini maksimum yapan w değerinin bulunması işlemine dönüşmüş olur⁶⁵. $\frac{2}{\|w\|}$ değerini maksimum yapmak için, $\|w\|$ değerinin, dolayısıyla $\|w\|^2$ değerinin minimize edilmesi gerekmektedir. Burada $\|w\|^2 = w^t w$ 'dir. Bu durumda, en iyi ayırıcı düzlemi bulmak için, aşağıdaki denklemlerin çözümü gerekir:

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad (2.19)$$

$$y_i(w^t x_i + b) \geq 1, \quad \forall_i \quad (2.20)$$

Burada Formül (2.19) çözülecek problem (nesne fonksiyonu) ve Formül (2.20) problemin çözümü sırasında kullanılan koşul yani eşitsizlik kısıtıdır. (2.19)'deki ifade ikinci dereceden eşitsizlik kısıtlı bir doğrusal olmayan optimizasyon problemidir^{3,28,80,85}. Nesne fonksiyonu karesel bir fonksiyon ve w , b parametrelerini içeren eşitsizlik kısıtı doğrusal olduğu için Lagrange çarpanları yöntemi ile çözüm yapılır³. Lagrange çarpanları yöntemi iki nedenle önem kazanmaktadır: Birincisi problemin Lagrangian formülasyonu yapıldığında, kısıtlarla yer değiştirecek Lagrange çarpanlarının hesaplanması daha kolaydır. İkinci neden ise, problemin yeniden formülasyonunda, gerçek eğitimde ve test algoritmalarında, eğitim verisinin sadece vektörler arasındaki nokta çarpım şeklinde görülecek olmasıdır. Bu, prosedürün doğrusal olmayan duruma genellenmesine izin verecek olan önemli bir özelliktir^{28,68}. Burada, Lagrange teoreminin kuralı hatırlatılacak olursa: Lagrange oluşturmak amacıyla $c_i \geq 0$ formundaki kısıtlar için, kısıt denklemleri pozitif Lagrange çarpanları ile çarpılır ve amaç fonksiyonundan çıkarılır⁶⁸. Pozitif Lagrange çarpanları olan α_i 'ler kullanılarak dönüştürülen yeni optimizasyon problemi aşağıdaki gibidir:

$$\begin{aligned} L_P &= \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i (y_i (w^t x_i + b) - 1) \\ &= \frac{1}{2} w^t w - \sum_{i=1}^n \alpha_i y_i (w^t x_i + b) + \sum_{i=1}^n \alpha_i \end{aligned} \quad (2.21)$$

Bu Lagrangian formülü birincil (primal) değişkenler w ve b bakımından minimize edilmeli, ikincil (dual) değişkenler bakımından maksimize edilmelidir³.

Uygulamada, Formül (2.21)'de görülen primal problemi çözmek yerine, yaygın olarak dual (Wolfe dual) karesel optimizasyon problemi çözülmektedir ve (2.19) ile aynı sonucu vermektedir²⁸.

Formül (2.21)'de ifade edilen formülasyonun çözülmesi oldukça karmaşıktır. Çözümün bulunması için formüldeki w ve b parametrelerinin sadece α_i parametresiyle ifade edilmesini sağlayacak olan Karush-Kuhn-Tucker (K.K.T.) koşulları olarak bilinen yöntem kullanılır ve bu durumda, Formül (2.21) sadece α_i Lagrange çarpanlarına göre maksimumlaştırılması istenen bir dual probleme dönüştürülür^{65,80}. K.K.T. koşulları ile çözüm sağlamak için öncelikle formül (2.21)'in w ve b 'ye göre türevleri alınır:

$$\frac{\partial}{\partial b} L_P = 0 \Rightarrow b = \sum \alpha_i y_i = 0 \quad \alpha_i \geq 0 \quad (2.22)$$

$$\frac{\partial}{\partial w} L_P = 0 \Rightarrow w = \sum \alpha_i y_i x_i \quad \alpha_i \geq 0 \quad (2.23)$$

w ve b parametrelerini bulmak için ulaşılan bu formüller, bilinmeyen Lagrange çarpanlarını (α_i) içerdiği için halen çözüm üretmemektedir. Çözüm için (x_i, y_i) noktasında Formül (2.20)'yi eşitlik haline dönüştürecek 0'dan farklı α_i 'leri sağlamalıdır3:

$$\alpha_i [y_i < x_i, w > +b) - 1] = 0 \quad i = 1, \dots, m \quad (2.24)$$

α_i 'lerin 0 olmadığı yerlerde eşitliği sağlayan (x_i, y_i) noktaları destek vektörleridir ve ayırıcı vektöre paralel olan marjin doğrusu üzerinde yer alırlar.

(2.22) ve (2.23) ile ifade edilen koşullar Formül (2.21)'de yerine yazılırsa, aşağıdaki dual problem elde edilir:

$$\begin{aligned} L_D &= \frac{1}{2} w^t w - w^t \sum_{i=1}^n \alpha_i y_i x_i + b \sum_{i=1}^n \alpha_i y_i + \sum_{i=1}^n \alpha_i \\ &= \frac{1}{2} w^t w - w^t w + 0 + \sum_{i=1}^n \alpha_i \\ &= -\frac{1}{2} w^t w + \sum_{i=1}^n \alpha_i \\ L_D &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i^t x_j, \quad \alpha_i \geq 0 \quad \forall_i \end{aligned} \quad (2.25)$$

Artık b ve w parametreleri için çözüm üretebilecek Formül (2.22) ve Formül (2.23) kullanılarak, sınıfları ayıracak karar fonksiyonunu belirlenebilir:

$$f(x) = (\sum \alpha_i y_i x_i \cdot x) + b \quad (2.26)$$

Karar fonksiyonu bulunduktan sonra, yeni gelen bir örneğin hangi sınıfa ait olacağına aşağıdaki eşitsizlikler yardımıyla karar verilir:

$$f(x) > 0 \Rightarrow \text{sınıf 1} \quad (2.27)$$

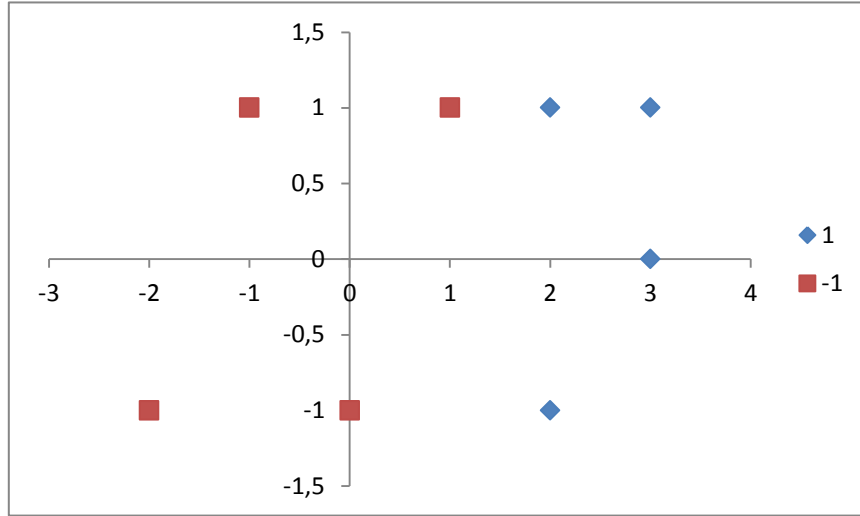
$$f(x) < 0 \Rightarrow \text{sınıf 2} \quad (2.28)$$

Örnek 2.1: Doğrusal olarak ayırma problemi için, aşağıda elle çözülmüş basit bir örnek verilmiştir.

Zekâ geriliğini belirlemek için önemli olan iki özellik x_1 ve x_2 olsun. Pozitif ve negatif gruplarda yer alan 4'er kişi için bu özellik değerlerinin aşağıdaki gibi olduğu varsayalım:

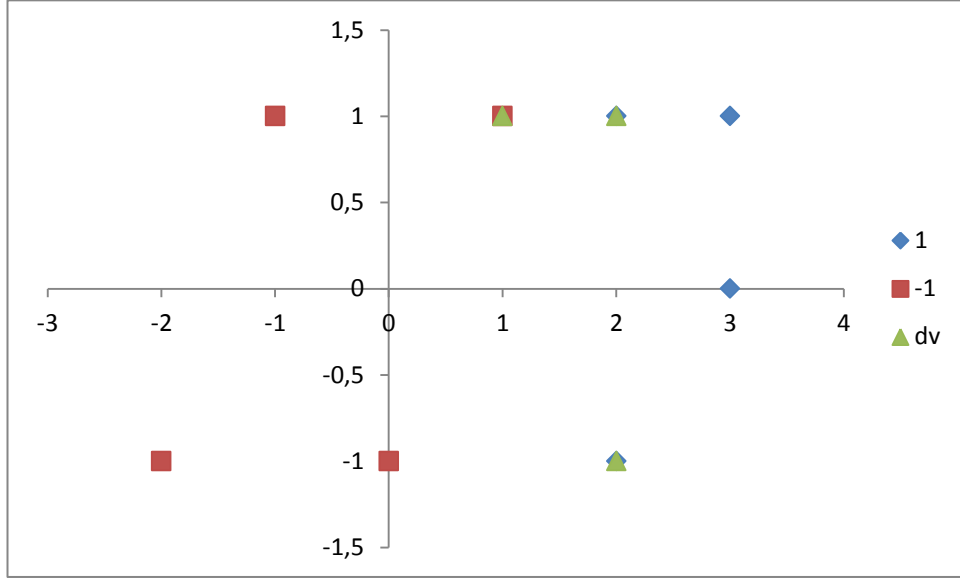
Pozitif Sınıf (\mathbb{R}^2 'de)		Negatif Sınıf (\mathbb{R}^2 'de)	
X_1	X_2	X_1	X_2
2	1	-1	1
2	-1	1	0
3	1	1	1
3	0	-2	-1

Burada, pozitif sınıf zeka geriliği olan kişileri, negatif sınıf ise zeka geriliği olmayan kişileri ifade etmektedir. Şekil 2.19'da pozitif ve negatif sınıflara ait bu noktaların iki boyutlu uzaydaki görünümü verilmiştir. Şekilde, mavi noktalar pozitif örnekleri, kırmızı noktalar ise negatif örnekleri göstermektedir.



Şekil 2.19: \mathbb{R}^2 'de Noktaların Görünümü

Bu örnekte veriler doğrusal olarak ayrılabilirdiği için doğrusal DVM kullanılacaktır. Şekil 2.20 incelendiğinde, üç adet destek vektörün (s_1, s_2, s_3) olduğu açıkça görülecektir. Destek vektörlerin koordinatları tanımlandığında; $\{s_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, s_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, s_3 = \begin{pmatrix} 2 \\ -1 \end{pmatrix}\}$ elde edilir.



Şekil 2.20: Yeşil Üçgenlerle Gösterilen Destek Vektörler

Daha sonra, bu destek vektörlerine yanlılığı (b) ifade eden l değeri eklenecek ve karışıklığı önlemek için destek vektörlerin bu hali \tilde{s}_1 \tilde{s}_2 ve \tilde{s}_3 olarak gösterilecektir. Böylece, $\tilde{s}_1 = (1 \ 1 \ 1)$, $\tilde{s}_2 = (2 \ 1 \ 1)$ ve $\tilde{s}_3 = (2 \ -1 \ 1)$ olacaktır. Aşağıdaki eşitlikler yardımıyla α_i değerleri bulunursa:

$$\begin{aligned} \alpha_1 \langle \tilde{s}_1, \tilde{s}_1 \rangle + \alpha_2 \langle \tilde{s}_2, \tilde{s}_1 \rangle + \alpha_3 \langle \tilde{s}_3, \tilde{s}_1 \rangle &= -1 \\ \alpha_1 \langle \tilde{s}_1, \tilde{s}_2 \rangle + \alpha_2 \langle \tilde{s}_2, \tilde{s}_2 \rangle + \alpha_3 \langle \tilde{s}_3, \tilde{s}_2 \rangle &= +1 \\ \alpha_1 \langle \tilde{s}_1, \tilde{s}_3 \rangle + \alpha_2 \langle \tilde{s}_2, \tilde{s}_3 \rangle + \alpha_3 \langle \tilde{s}_3, \tilde{s}_3 \rangle &= +1 \end{aligned}$$

ve

$$\begin{aligned} 3\alpha_1 + 4\alpha_2 + 2\alpha_3 &= -1 \\ 4\alpha_1 + 6\alpha_2 + 4\alpha_3 &= +1 \\ 2\alpha_1 + 4\alpha_2 + 6\alpha_3 &= +1 \end{aligned}$$

Yukarıdaki eşitlikler çözüldüğünde, $\alpha_1 = -8$, $\alpha_2 = 6,5$ ve $\alpha_3 = -1,5$ olarak bulunur. α değerleri bulunduktan sonra, pozitif ve negatif değerleri ayıracak hiperdüzlemi bulmak için;

$$\begin{aligned} \tilde{w} &= \sum_i \alpha_i \tilde{s}_i = -8 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + 6.5 \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} - 1.5 \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 2 \\ 0 \\ -3 \end{pmatrix} \end{aligned}$$

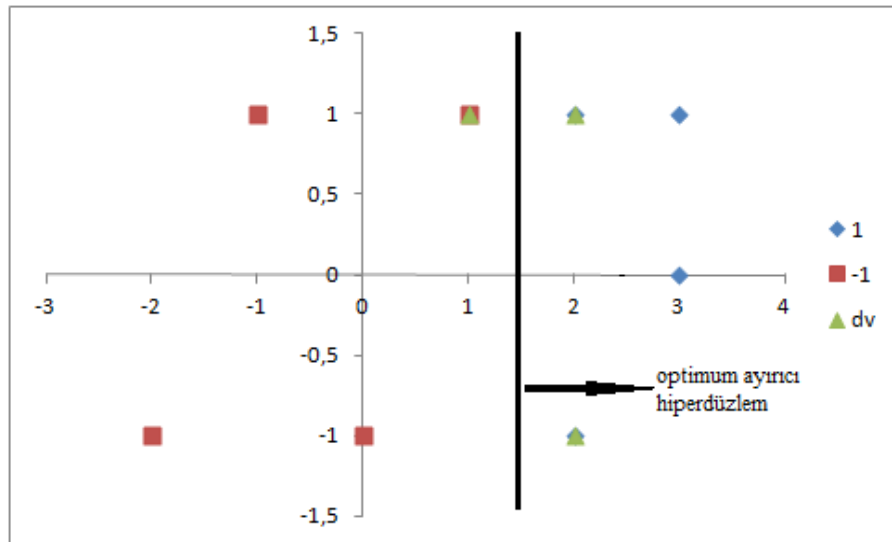
ağırlık vektörü bulunur.

Son olarak, ayırıcı hiperdüzlem eşitliğini $y = w^T x_i + b = 0$ şeklinde yazılır ve elde edilen vektörlere bir yanlılık değeri eklendiği hatırlanırsa, bu durumda y eşitliğinde $w = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$ ve $b = -3$ olur⁸⁶.

Buradan, hiperdüzlemin koordinatları belirlenirse:

$$y = w^T x_i + b = 0 \Rightarrow [2 \ 0] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + (-3) = 0 \Rightarrow 2x_1 + 0x_2 = 3$$

olur ve koordinatlar, $x_1 = 1.5$ ve $x_2 = 0$ olarak bulunur. Bu durumda, $x_1 = 1.5$ ve $x_2 = 0$ noktalarından geçen hiperdüzlem, optimum ayırıcı hiperdüzlem olacaktır.



Şekil 2.21: Bulunan $\alpha_1 = -8$, $\alpha_2 = 6.5$ ve $\alpha_3 = -1.5$ Değerleri İçin Ayırıcı Hiperdüzlem

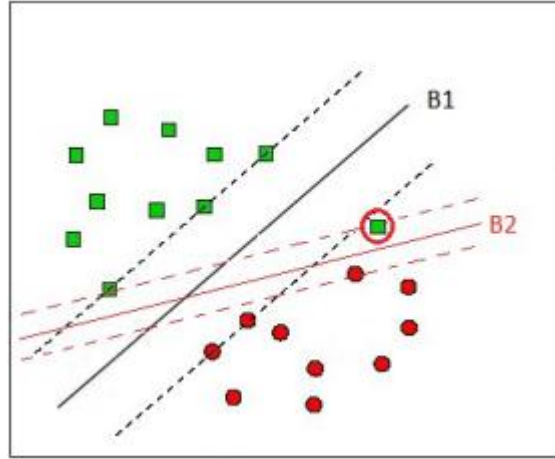
Yeni gelen bir kişinin x_1 ve x_2 özelliklerine ait özellik vektörü, $x = (4, 2)$ şeklinde olsun ve bu kişinin dahil olduğu sınıf belirlenmek istensin. Bu durumda, karar fonksiyonu hesaplanırsa;

$$\begin{aligned} f(x) &= \left(\sum \alpha_i y_i x_i \cdot x \right) + b = w^T x_i + b \\ &= w^T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + b \\ &= (2 \ 0) \begin{pmatrix} 4 \\ 2 \end{pmatrix} + (-3) \\ &= 5 \end{aligned}$$

şeklinde bulunur. $f(x)=5>1$ olduğundan, bu kişi pozitif sınıfa yani zeka geriliği olan kişiler sınıfına atanır.

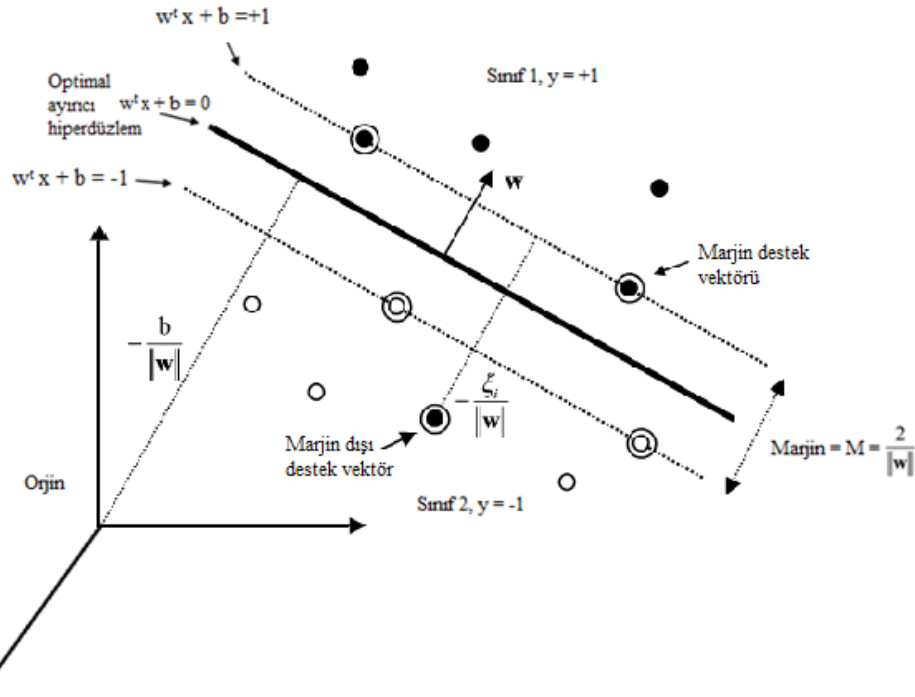
2.3.4.1.B. Verilerin tamamının doğrusal olarak ayrılmadığı durum (yumuşak (soft) marjin)

Bölüm 3.3.1.1.'de gösterilen işlemler ancak eğitim örneklerinin tamamen ayrılabilir olması durumunda çalışmaktadır. Ancak, pratikte genellikle veriler tam olarak ayrılamayabilmektedir.



Şekil 2.22: Verilerin Tamamının Doğrusal Olarak Ayrılmadığı Durum

Şekil 2.22 incelendiğinde, B_2 hiperdüzlemi noktaların tamamını doğru sınıflandırırken, B_1 hiperdüzlemi ise aynı şekilde noktaların tümünü doğru sınıflandıramamıştır. Sadece şekle bakarak tüm noktaları doğru sınıflandırdığı için B_2 hiperdüzleminin daha iyi bir ayırıcı düzlem olduğu düşünülebilir. Ancak bu doğru değildir, çünkü B_1 daha geniş bir marjine sahiptir ve bu tek gözlem outlier verisi de olabilir³. Dolayısıyla bu gibi durumlarda sert marjin yaklaşımı çalışmaz ve küçük deneme hatalarını tolere edebilecek başka yöntemlere gerek duyulur. Yumuşak (soft) marjin yaklaşımı bu tip problemlerin çözümü için öne sürülmüştür.



Şekil 2.23: Verilerin Tamamının Doğrusal Olarak Ayrılamadığı Durum İçin Optimum Ayırıcı Hiperdüzlem

Yumuşak marjin yaklaşımı, Cortes ve Vapnik tarafından geliştirilmiştir⁵⁷. Bu yaklaşımda DVM'ler, marjini maksimize eden ve aynı zamanda yanlış sınıflandırma hatalarının sayısı ile orantılı bir niceliği minimize eden bir hiperdüzlem bulmaya çalışır²⁸. Burada, marjinin maksimum hale getirilmesi ve yanlış sınıflandırma hatalarının minimum hale getirilmesi arasındaki denge, pozitif değerler alan ve C ile gösterilen bir düzenleme parametresi ($0 < C < \infty$) ile kontrol edilebilir⁶⁰. Bu parametre, Lagrange çarpanlarının alabilecekleri maksimum değeri göstermektedir. Bu şekilde Lagrange çarpanlarının $0 \leq \alpha_i \leq C$ aralığında kalması sağlanmaktadır⁸⁰. Bu parametreye aynı zamanda hata maliyeti de denmektedir ve yüksek C değeri yüksek hata beklentisini göstermektedir⁶⁸. Küçük bir C değeri ise, eğitim hatalarının sayısını arttırmaktadır⁸⁷. Ayrıca, yaklaşımda negatif olmayan gevşek (slack) değişken adı ile bir değişken ($\xi_i \geq 0$, $i=1, \dots, n$) tanımlanmıştır³. Bu gevşek değişken (ξ_i), uygun olan sınıfın sınırdan sapmasıdır ve minimize edilmiş bir hata derecesine izin verir^{28,65}. Eğitim setinde hatalı sınıflandırma oranı gevşek değişkenin değerini verir. Sert marjinde elde edilen kısıtlara gevşek değişken eklenerek aşağıdaki forma ulaşılır:

$$w^t x_i + b \geq +1 - \xi_i, \quad y_i = +1 \text{ için} \quad (2.29)$$

$$w^t x_i + b \leq -1 + \xi_i, \quad y_i = -1 \text{ için} \quad (2.30)$$

$$\xi_i \geq 0, \quad \forall_i \quad (2.31)$$

Gevşek değişkenler, ayırım kısıtlarını gevşetmek için tanıtılmaktadır ve $\xi_i = 0$ olması durumunda x_i örneği doğru sınıflandırılmış, $0 < \xi_i < 1$ olması durumunda x_i örneği doğru sınıflandırılmış marjini belirleyen hiperdüzlemler arasında yer alıyor, $\xi_i \geq 1$ ise x_i örneği yanlış sınıflandırılmış demektir⁷⁷. Burada, eğitim setini en az hatayla sınıflamak için kullanılan ξ_i 'leri minimize etmek, aşağıdaki ifadeyi minimize etmeyi gerektirecektir⁶⁰:

$$\Phi(\xi) = \sum_{i=1}^n \xi_i^\sigma \quad (2.32)$$

Burada σ , birtakım küçük pozitif sabitlerdir. Varsayımı güvenli kılmak için genellikle $\sigma=1$ değeri kullanılmaktadır. Çünkü karşılık gelen dual ξ içermez ve bu nedenle daha basit bir optimizasyon problemi sunar^{28,65}. Formül (2.32) minimize edilerek, eğitim verisinin küçük bir alt kümesi bulunabilir:

$$(x_{i1}, y_{i1}), \dots, (x_{ik}, y_{ik})$$

Eğer, bu veri seti eğitim verisinin dışındaysa, eğitim setinin kalan kısmı hatasız olarak ayrılabilir. Eğitim setinin kalan kısmını ayırmak için, optimal ayırıcı hiperdüzlem bulunabilir⁶⁰.

Sert marjin yaklaşımında kullanılan nesne fonksiyonu $\|w\|^2/2$, doğrusal olarak ayrılamayan veriler için, düzenleme parametresi ve gevşek değişken kullanılarak aşağıdaki gibi ifade edilir:

$$\min_{wb} \frac{1}{2} \|w\|^2 + C \sum_i \xi_i \quad (2.33)$$

Burada, $\sum_i \xi_i$ deneme hatalarının sayısına ilişkin bir üst sınır verirken, C marjin maksimizasyonu ile deneme hatası minimizasyonu arasındaki ödünleşimi (trade-off) belirler ve Lagrange çarpanları için bir üst sınır oluşturur³:

$$0 \leq \alpha_i \leq C \quad (2.34)$$

Formül (2.33)'deki ilk terim, öğrenen makinelerin ele alınan sınıfının VC boyutunu; bu yolla da öğrenen makinelerin öğrenme kapasitesini veya güven aralığını kontrol etmek, makinenin aşırı uyumdan kaçınmasını sağlamak ve belirleyici nitelik uzayındaki bir ayırıcı hiperdüzlemin marjini maksimum yapabilmek amacıyla minimize edilmektedir. İkinci terim ise, yanlış sınıflandırılmış öğelerin sayısını kontrol etmek, yani ampirik risk terimini kontrol etmek için minimize edilmektedir⁸⁸.

Yumuşak marjin yaklaşımı için birincil (primal) Lagrangian formülü:

$$L_P = \frac{1}{2} \|w\|^2 + C \sum_i \xi_i - \sum \alpha_i \{y_i(x_i w + b) - 1 + \xi_i\} - \sum \mu_i \xi_i \quad (2.35)$$

Burada μ_i , ξ_i değerlerinin pozitif olmasını sağlamak için kullanılan Lagrange parametreleridir⁸⁰. α_i 'ler yine pozitif Lagrange çarpanlarıdır. Sert marjin için yapılan işlemler aynı şekilde burada da uygulanır. Çözümü zor olan birincil Lagrangian'ı, dual forma dönüştürmek için K.K.T. koşulları uygulanırsa:

$$\frac{\partial}{\partial b} L_P = 0 \Rightarrow \sum \alpha_i y_i = 0 \quad (2.36)$$

$$\frac{\partial}{\partial w} L_P = 0 \Rightarrow w = \sum \alpha_i y_i x_i \quad (2.37)$$

$$\frac{\partial}{\partial \xi_i} L_P = 0 \Rightarrow C = \alpha_i + \mu_i \quad (2.38)$$

$$\xi_i \geq 0 \quad (2.39)$$

$$\alpha_i \geq 0 \quad (2.40)$$

$$\mu_i \geq 0 \quad (2.41)$$

ifadeleri elde edilir. Bu ifadeler Formül (2.35)'de yerine yazılırsa aşağıdaki dual fonksiyona ulaşılır:

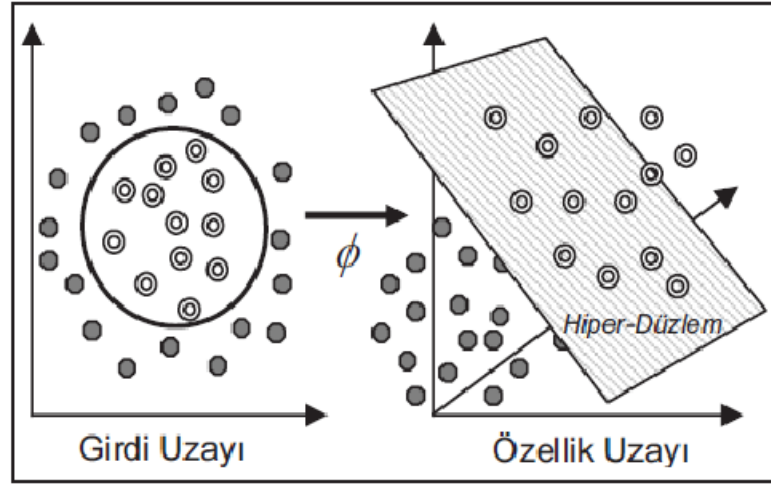
$$L_D = \sum \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i x_j, \quad 0 \leq \alpha_i \leq C, \quad \forall_i \quad (2.42)$$

Bu problemin çözümünde $0 \leq \alpha_i \leq C$ aralığında yer alan Lagrange çarpanlarına karşılık gelen x_i değerleri DV'lerdir. Elde edilen dual fonksiyon, sert marjin yaklaşımında elde edilen fonksiyonlar ile aynı olmasına karşın, farklılık Lagrange çarpanlarında yer almaktadır. Sert marjin yaklaşımında α_i 'lerin pozitif olma koşulu varken, yumuşak marjin yaklaşımında α_i 'ler C üst sınırına sahiptir, bu sınırı aşamazlar³.

2.3.4.2. Doğrusal olmayan destek vektör makineleri ile sınıflama

Bölüm 2.3.4.1.A ve 2.3.4.1.B'de bahsedilmiş olan DVM, örneklerin tamamının doğrusal olarak ayrılabilirdiği ya da tamamı ayrılamasa da çeşitli parametrelerle ayrımın yapılabileceği durumlar için problemlere çözüm aramaktaydı⁸⁰. Ancak, pratikte

verilerin doğrusal olarak ayrılabilirdiği durumlarla pek karşılaşmamaktadır ve bu durumda, pratik uygulamaların çoğu, doğrusal DVM'lerle çözülememektedir.

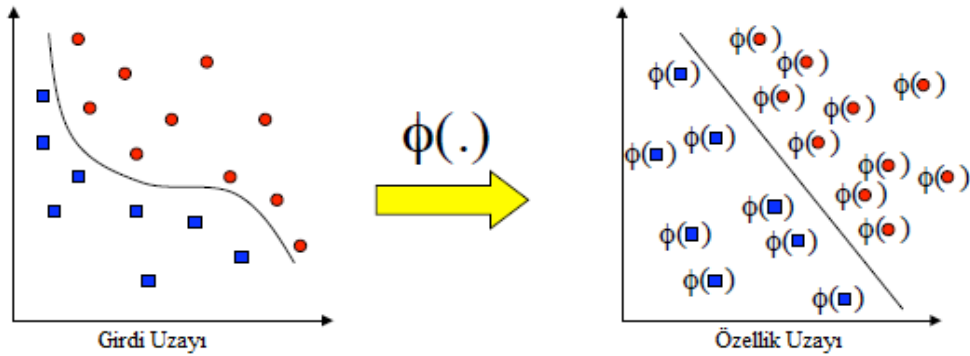


Şekil 2.24: Doğrusal Olarak Ayrılamayan Verilerin Farklı Boyutlardaki Uzaylara Aktarılması

Şekil 2.24'te iki sınıf iç içe geçmiş durumdadır. DVM'ler böyle doğrusal olmayan problemlerle karşılaştığında, orijinal verilerden sınıflandırma özelliklerini çıkarmak için, doğrusal olmayan haritalama (mapping) yaparak, verileri n boyutlu orijinal girdi uzayından daha yüksek boyuta sahip belirleyici nitelik (feature) uzayına taşır²⁸.

$$x \in R^n \rightarrow \Phi(x) \in R^f \quad (2.43)$$

DVM daha sonra, belirleyici nitelik uzayında maksimum marjini bulmak için doğrusal sınıflandırma kuralını öğrenir. Sınıflandırma kuralı, belirleyici nitelik uzayında doğrusal olması gerçeğine karşın, orijinal girdi uzayına izdüşüm yapıldığında doğrusal değildir²⁸.



Şekil 2.25: Doğrusal Olarak Ayrılamayan Verilerin, Doğrusal Olmayan Haritalama Fonksiyonları ile Daha Yüksek Boyutlu Belirleyici Nitelik Uzayına Taşınması

Doğrusal olmayan DVM, verilerin taşındığı bu yeni boyutta doğrusal DVM gibi çalışarak verileri ayıracak optimum hiperdüzlemi arar. Doğrusal olmayan bir problemin,

belirleyici nitelik uzayında doğrusal ayrılabilirliği gibi davranabilmesi için; haritalama işleminde, bir dezavantaj olarak, boyut sayısının belirgin şekilde artırılması gerekmektedir²⁸.

Dönüştürme işlemi için kullanılacak olan fonksiyon $\Phi(x)$ olarak belirlensin. Bu durumda doğrusal DVM'den tek farkı x yerine $\Phi(x)$ kullanılması olacaktır³. Buradan hareketle dönüştürülmüş uzayda kullanılacak karar fonksiyonu:

$$w^t \Phi(x) + b = 0 \quad (2.44)$$

şeklinde olacaktır. Ayırıcı hiperdüzleme paralel olan ve üzerinde destek vektörlerinin bulunduğu doğruların ayırdığı veriler aşağıdaki eşitsizlikler yardımıyla sınıflanır:

$$w^t \Phi(x) + b \geq 1 \quad (2.45)$$

$$w^t \Phi(x) + b \leq -1 \quad (2.46)$$

Bulunması gereken nesne fonksiyonu ve buna ilişkin Formül (2.43) ile Formül (2.44)'ün birleşiminden oluşan kısıt aşağıdaki gibidir:

$$\min_{wb} \frac{1}{2} \|w\|^2 \quad (2.47)$$

$$y_i(w^t \Phi(x) + b) - 1 \geq 0, \quad \forall i \text{ için} \quad (2.48)$$

Ancak, genelde $\Phi(x)$ fonksiyonu elde edilebilir değildir, hesaplanamaz hatta mevcut değildir²⁸. Uygulanan haritalama fonksiyonu biliniyorsa bile, kurulan optimizasyon probleminin yüksek boyutlu belirleyici nitelik uzayında çözümü karmaşık ve zor hesaplamalar gerektirecektir³.

w ve b parametrelerini hesaplamak için aşağıdaki eşitlikler çözülmelidir:

$$w = \sum \alpha_i y_i \Phi(x_i) \quad (2.49)$$

$$f(x) = (\sum \alpha_i y_i \Phi(x_i) \cdot \Phi(x)) + b = 0 \quad (2.50)$$

Bu denklemler dönüştürülmüş uzaydaki iki vektörün iç çarpımını içermektedir. Boyut sorunundan (curse of dimensionality) dolayı bu iç çarpımların hesaplanması zordur. Bu sorunu önlemek amacıyla çekirdek düzenlemesi olarak adlandırdığımız *kernel trick* yöntemi önerilmiştir. Doğrusal olmayan haritalama Φ ; destek vektörleri $\Phi(x_i)$ ile

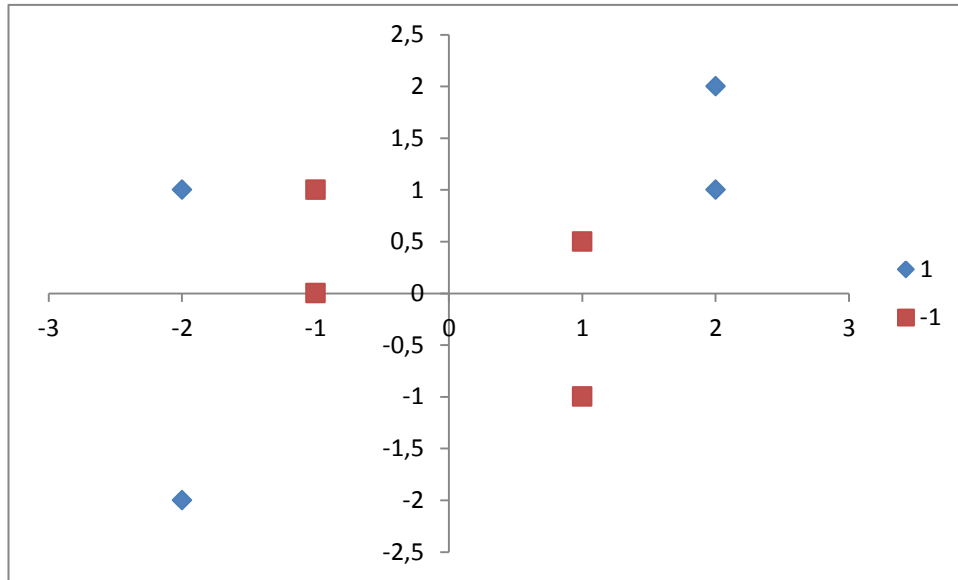
belirleyici nitelik uzayındaki örüntü vektörü $\Phi(x)$ arasındaki iç çarpımı hesaplamak için Mercer koşullarına uyan çekirdek fonksiyonlarını ($K(x_i, x)$) kullanır⁵⁴.

Örnek 2.2: Grupların doğrusal olarak ayıramadığı ve $\Phi(x)$ fonksiyonunun bilindiği bir durum için, aşağıda elle çözülmüş bir örnek verilmiştir.

Zeka geriliğini belirlemek için önemli olan iki özellik x_1 ve x_2 olsun. Dört kişi için bu özellik değerlerinin aşağıdaki gibi olduğu varsayalım:

Pozitif Sınıf (R^2 'de)		Negatif Sınıf (R^2 'de)	
X_1	X_2	X_1	X_2
-2	1	-1	0
2	1	1	0.5
2	2	1	-1
-2	-2	-1	1

Burada, pozitif sınıf zeka geriliği olan kişileri, negatif sınıf ise zeka geriliği olmayan kişileri ifade etmektedir. Şekil 2.26'da pozitif ve negatif sınıflara ait bu noktaların iki boyutlu uzaydaki görünümü verilmiştir. Şekilde, mavi noktalar pozitif örnekleri, kırmızı noktalar ise negatif örnekleri göstermektedir.



Şekil 2.26: R^2 'de Doğrusal Olarak Ayıramayan Noktaların Görünümü

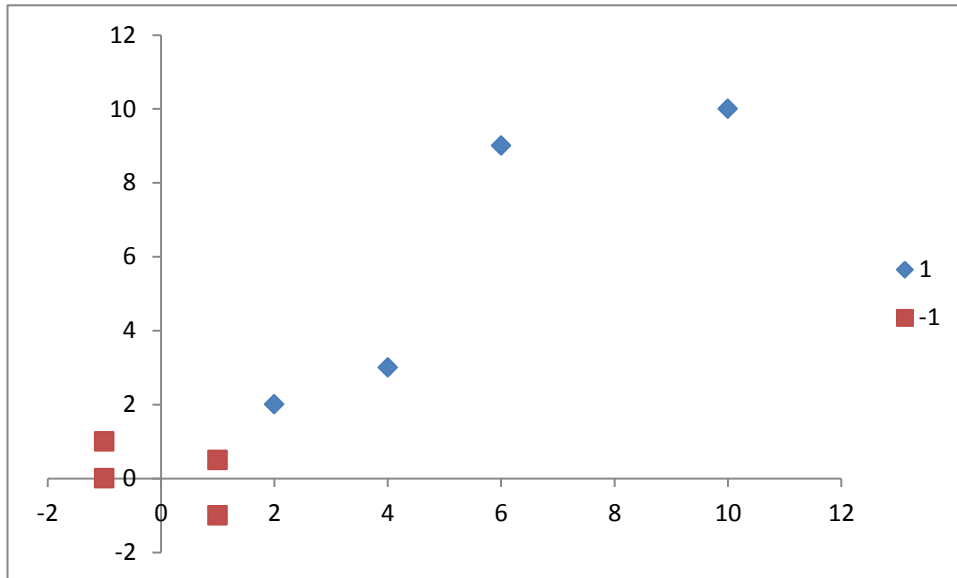
Amaç, iki sınıfı en doğru şekilde ayırabilen en iyi hiperdüzlemi seçmektir. Ancak orijinal veride (girdi verisi) bu en iyi hiperdüzlemi bulmak mümkün olmayacaktır (Şekil 2.26). Bu nedenle doğrusal olmayan DVM kullanılarak veriler, haritalama fonksiyonu

(Φ) yardımıyla girdi uzayından, belirleyici nitelik uzayına taşınmaktadır. Bu örnek için haritalama fonksiyonunu aşağıdaki gibi tanımlansın:

$$\Phi_1 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{cases} \begin{pmatrix} 4 - x_2 + |x_1 - x_2| \\ 4 - x_1 + |x_1 - x_2| \end{pmatrix}, & \sqrt{x_1^2 + x_2^2} > 2 \text{ ise,} \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, & \text{diğer durumlarda.} \end{cases}$$

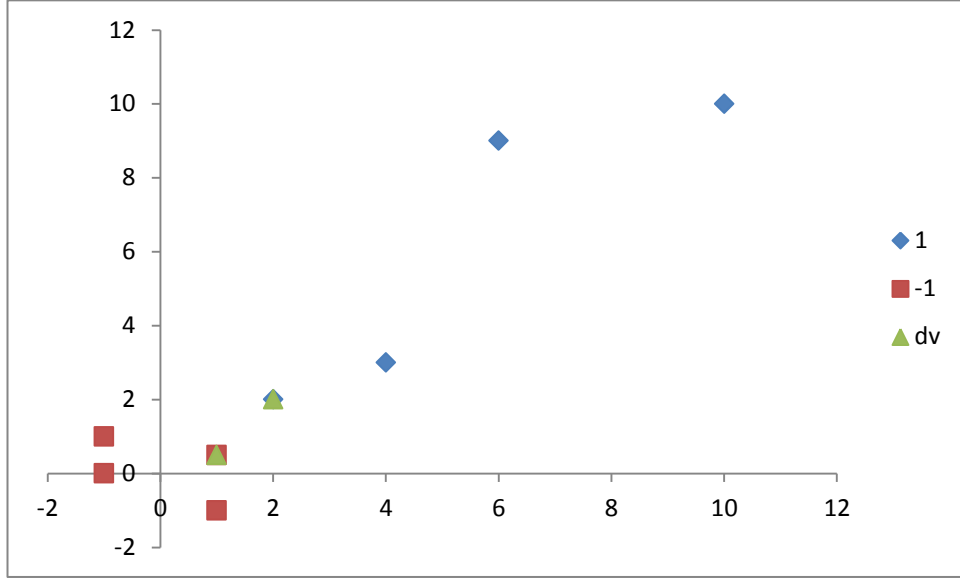
Böylece veri, belirleyici nitelik uzayında aşağıdaki gibi yeniden düzenlenir:

Pozitif Sınıf (\mathbb{R}^2 'de)		Negatif Sınıf (\mathbb{R}^2 'de)	
X_1	X_2	X_1	X_2
6	9	-1	0
4	3	1	0.5
2	2	1	-1
10	10	-1	1



Şekil 2.27: Belirleyici Nitelik Uzayına Taşınan Verilerin Görünüşü

Şekil 2.27 incelediğinde, verilerin belirleyici nitelik uzayındaki koordinatları görülebilir. Bu koordinatlar içerisinde, aşağıda tanımlanan s_1 ve s_2 destek vektörler olarak seçildiğinde bu vektörlerin koordinatları ise; $\{s_1 = \begin{pmatrix} 1 \\ 0.5 \end{pmatrix}, s_2 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}\}$ olur.



Şekil 2.28: Yeşil Üçgenlerle Gösterilen Destek Vektörler

Şekil 2.28'de destek vektörler yeşil üçgenlerle gösterilmiştir. Destek vektörlerine yanlılığı (b) ifade eden 1 değeri eklenerek α_i değerleri bulunacaktır. "b" eklenmiş destek vektörleri \tilde{s}_1 ve \tilde{s}_2 ile gösterirse $\tilde{s}_1 = (1 \ 0.5 \ 1)$ ve $\tilde{s}_2 = (2 \ 2 \ 1)$ olacaktır. α_i 'lere ait eşitlikler yazılırken haritalama fonksiyonu da göz önüne alınacaktır.

$$\begin{aligned}\alpha_1 \Phi_1(s_1) \cdot \Phi_1(s_1) + \alpha_2 \Phi_1(s_2) \cdot \Phi_1(s_1) &= -1 \\ \alpha_1 \Phi_1(s_1) \cdot \Phi_1(s_2) + \alpha_2 \Phi_1(s_2) \cdot \Phi_1(s_2) &= +1\end{aligned}$$

Yukarıda tanımlanan iki eşitlik $\Phi_1(s_i) \cdot \Phi_1(s_j) = \tilde{s}_i \cdot \tilde{s}_j$ $i, j = 1$ ve 2 şeklinde yazıldığında α_i değerleri aşağıdaki denklemler yardımıyla bulunur. Burada eşitlikler, doğrusal DVM'deki eşitliklere dönüşmektedir. Ancak eşitliklerdeki \tilde{s}_i 'ler haritalanmış değerlerdir. Aşağıdaki eşitlikler yardımıyla α_i değerlerini bulunursa:

$$\begin{aligned}\alpha_1 \tilde{s}_1 \cdot \tilde{s}_1 + \alpha_2 \tilde{s}_2 \cdot \tilde{s}_1 &= -1 \\ \alpha_1 \tilde{s}_1 \cdot \tilde{s}_2 + \alpha_2 \tilde{s}_2 \cdot \tilde{s}_2 &= +1\end{aligned}$$

ve

$$\begin{aligned}1.25\alpha_1 + 3\alpha_2 &= -1 \\ 3\alpha_1 + 8\alpha_2 &= +1\end{aligned}$$

Yukarıdaki eşitlikler çözüldüğünde, $\alpha_1 = -11$ ve $\alpha_2 = 4.25$ olarak bulunur. α değerleri bulunduktan sonra, pozitif ve negatif değerleri ayıracak hiperdüzlemi bulmak için,

$$\tilde{w} = \sum_i \alpha_i \tilde{s}_i$$

$$= -11 \begin{pmatrix} 1 \\ 0.5 \\ 1 \end{pmatrix} + 4.25 \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} -2.5 \\ 3 \\ -6.75 \end{pmatrix}$$

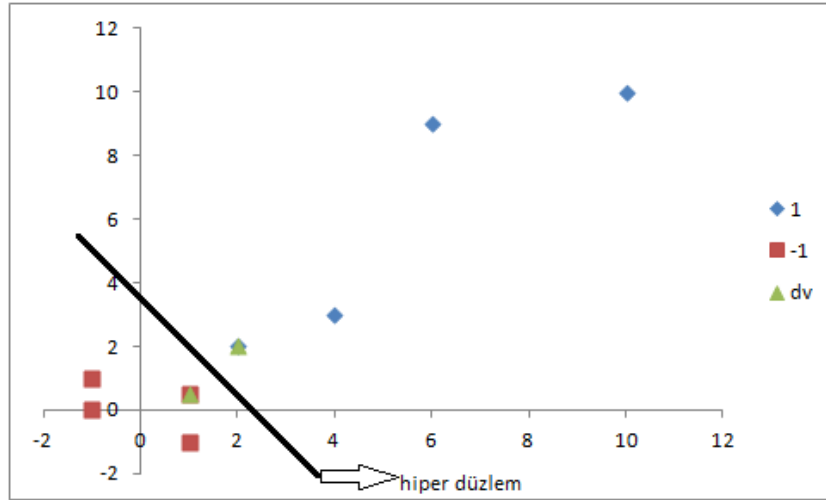
ağırlık vektörü bulunur.

Son olarak, ayırıcı hiperdüzleme ait denklem, $y = w^t x_i + b = 0$ şeklinde tanımlanır ve elde edilen vektörlere bir yanlılık değeri eklendiği hatırlanırsa y eşitliğinde $w = \begin{pmatrix} -2.5 \\ 3 \end{pmatrix}$ ve $b = -6.75$ olarak elde edilir⁸⁶.

buradan hiperdüzlemin koordinatları belirlenirse:

$$y = w^t \cdot x_i + b = 0 \Rightarrow [-2.5 \ 3] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + (-6.75) = 0 \Rightarrow 2.5x_1 + 3x_2 = 6.75$$

olur ve koordinatlar, $x_1=2.7$ ve $x_2=2.25$ olarak bulunur. Bu durumda, $x_1=2.7$ ve $x_2=2.25$ noktalarından geçen hiperdüzlem, optimum hiperdüzlem olacaktır.



Şekil 2.29: Bulunan $\alpha_1 = -11$, $\alpha_2 = 4.25$ Değerleri için Ayırıcı Hiperdüzlem

Yeni gelen bir kişinin x_1 ve x_2 özelliklerine ait özellik vektörü, $x = (2,3)$ şeklinde olsun ve bu kişinin dahil olduğu sınıf belirlenmek istensin. Bu durumda, bu özellik vektörü yukarıda verilen haritalama fonksiyonuna göre yeniden düzenlenirse, $\Phi_1 \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ olur. Buradan karar fonksiyonu hesaplanırsa;

$$f(x) = \left(\sum \alpha_i \Phi(s_i) \cdot \Phi(x) \right) + b$$

$$\begin{aligned}
&= -11\Phi\left(\begin{matrix} 1 \\ 0.5 \end{matrix}\right)\Phi\left(\begin{matrix} 2 \\ 3 \end{matrix}\right) + 4.25\Phi\left(\begin{matrix} 2 \\ 2 \end{matrix}\right)\Phi\left(\begin{matrix} 2 \\ 3 \end{matrix}\right) \\
&= -11\begin{pmatrix} 1 \\ 0.5 \\ 1 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + 4.25\begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \\
&= -3.75
\end{aligned}$$

şeklinde bulunur. $f(x) = -3.75 < 1$ olduğundan, bu kişi negatif sınıfa yani zeka geriliği olmayan kişiler sınıfına atanır.

2.3.4.3. Çekirdek fonksiyonlar (kernel functions)

DVM’de genellikle, doğrusal olmayan iki sınıf arasındaki en iyi sınırı bulabilmek için, veri, giriş uzayından daha yüksek boyutlu bir belirleyici nitelik uzayına haritalama yoluyla taşınır⁶⁰. Yani girdi verisi, sınıfların bir düzlemle ayrılabilmediği daha yüksek boyutlu bir uzaya Φ haritalama fonksiyonu yardımıyla haritalanır⁸. Ancak, bu haritalama fonksiyonunu bilmek veya elde etmek genellikle zordur. Bu nedenle, haritalama işlemini yapabilmek için çekirdek fonksiyonlarından faydalanılır. Çekirdek fonksiyonları, bilinen bir haritalama fonksiyonu olmadan yüksek boyutlu uzayda nokta çarpımları hesaplamaya olanak sağlar ve belirleyici nitelik uzayındaki iç çarpımı gerçekleştirdiğinden Φ dönüşümünün analitik olarak bilinmesine gerek yoktur^{89,90}. Yani, vektörleri daha yüksek boyutlu bir belirleyici nitelik uzayına açık olarak haritalayıp iç çarpımı orada hesaplamak yerine, değeri doğrudan iki vektörün iç çarpımını veren bir çekirdek fonksiyonu kullanmak daha uygun olacaktır⁹¹.

Haritalama fonksiyonlarının bilinmediği durumlarda girdi uzayındaki verilerden elde edilen bir çekirdek fonksiyonu oluşturularak işlemler gerçekleştirilir. Bu şekilde yapılan çekirdek düzenlemesinin iki önemli faydası bulunmaktadır³:

- Direk girdi uzayındaki veriler kullanılacağı için Φ haritalama fonksiyonu hakkında kesin bilgiye gerek yoktur.
- Çekirdek fonksiyonu kullanarak iç çarpım hesaplamak, dönüştürülmüş özellik seti $\Phi(x)$ kullanarak hesaplamaya kıyasla daha kolaydır ve maliyeti düşüktür.

Doğrusal olarak ayrılamayan verilerde kullanılan DVM metodunun performansı, sınıflandırma yapılırken seçilmesi gereken çekirdek fonksiyonu ile direk alakalıdır. Bu

fonksiyon yardımıyla DVM, sınıflandırılacak doğrusal olmayan verileri daha yüksek boyutlu belirleyici nitelik uzayına taşır⁹².

Çekirdek fonksiyonlarının dayandığı temel fikir, girdi uzayındaki bir takım bileşenlerin belirli bir kurala göre dönüştürülmesidir. Bu dönüşüm, aslında girdi uzayında da gerçekleştirilebilecek bazı işlemlerin, çok boyutlu bir uzayda gerçekleştirilmesi üzerine kurulmuştur ve böylece çekirdek fonksiyonlarını kullanan metotlar girdi uzayında karmaşık olan birçok uygulama üzerinde de çalışabilmektedirler²⁹.

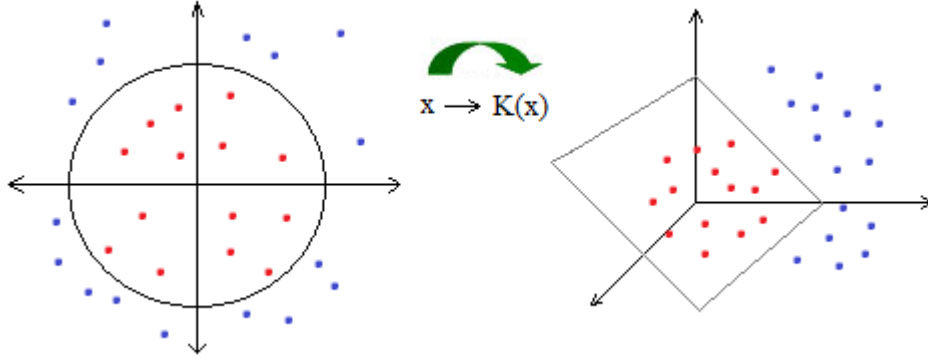
Verilerin doğrusal olarak ayrılamadığı durum için kullanılan karar fonksiyonu,

$$f(x) = \left(\sum \alpha_i y_i \Phi(x_i)^T \cdot \Phi(x_j) \right) + b \quad (2.51)$$

olarak tanımlanmıştı. Burada, $\Phi(x_i)^T \Phi(x_j)$ şeklinde gösterilen bir iç çarpım söz konusudur. Çekirdek fonksiyonu, beklenen belirleyici nitelik uzayında iki özellik vektörünün iç çarpımı ile ilgili olan bir fonksiyon olarak ifade edilmektedir⁹³. Tüm özellik vektörleri arasındaki iç çarpımlar kullanılarak oluşturulan matris, Gram matrisi (çekirdek matrisi) olarak adlandırılmaktadır. DVM’de, tüm veri girdileri çekirdek fonksiyonundan geçer ve çekirdek matrisinde sonlanır. Yani, çekirdek fonksiyonları, bütün veriyi özetleyen çekirdek matrisini oluştururlar⁹⁴. DVM yönteminde, matematiksel olarak $K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j)$ şeklinde ifade edilen bir çekirdek fonksiyonu yardımıyla doğrusal olmayan dönüşümler yapılabilen ve bu şekilde verilerin yüksek boyutta doğrusal olarak ayırılmasına imkan sağlanmaktadır⁸⁴. Bir çekirdek fonksiyonu olan $K(x_i, x_j)$ ifadesi aslında bir transfer fonksiyonudur⁹². Bu durumda, doğrusal olarak ayrılamayan veriler için kullanılacak karar fonksiyonu,

$$f(x) = \sum \alpha_i y_i K(x_i, x_j) \quad (2.52)$$

şeklinde olacaktır. Ancak burada doğrusal hiperdüzlem söz konusu olmadığından b terimi ihmal edilir. Çünkü b terimi çekirdek fonksiyonu içinde kapalı biçimde yer almaktadır⁹⁵. Şekil 2.29’de çekirdek fonksiyonlarının yaptığı dönüşüm gösterilmektedir.



Şekil 2.30: Verilerin Çekirdek Fonksiyonu ile Belirleyici Nitelik Uzayına Taşınması

Çekirdek fonksiyonları, girdi verileri arasındaki benzerliğin bir ölçüsü olarak düşünülebilir. Ancak, bütün benzerlik ölçüleri çekirdek fonksiyonu olarak kullanılamaz⁹³. Bir fonksiyonun çekirdek fonksiyonu olarak kabul edilebilmesi için, Mercer koşullarını yerine getirmesi gerekir^{29,89,94}. Bu koşullar, yüksek boyutta çalışılırken çekirdek fonksiyonların her zaman iki girdi vektörünün iç çarpımı şeklinde ifade edilmesini sağlamaktadır³.

Mercer teoremi, herhangi bir simetrik pozitif yarı-tanımlı (positive semidefinite) (özdeğerleri pozitif olan matris) matrisin, uzaydaki bir iç çarpım matrisi olarak kabul edilebileceğini belirtmektedir²⁸. Yani gerçel değerli $K(x_i, x_j)$ fonksiyonunun, $K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j)$ biçiminde yazılabildiğini sağlayan bir Φ eşlemesi varsa, pozitif yarı-tanımlı ve simetrik $K(x_i, x_j)$ bir çekirdek fonksiyonudur⁹⁵.

$K(x_i, x_j)$ fonksiyonunun bir çekirdek fonksiyonu olabilmesi için aşağıdaki koşulları gerçekleştirmesi gerekmektedir⁹⁵.

- Sürekli bir fonksiyon olmalıdır.
- Simetrik olmalıdır. Yani, $K(x_i, x_j) = K(x_j, x_i)$ koşulunu sağlamalıdır.
- Herhangi x_1, x_2, \dots, x_n değerleri için pozitif yarı-tanımlı olmalıdır:

$$K = \begin{bmatrix} K(x_1, x_1) & K(x_1, x_2) & \dots & K(x_1, x_n) \\ \dots & \dots & \dots & \dots \\ K(x_n, x_1) & K(x_n, x_2) & \dots & K(x_n, x_n) \end{bmatrix}$$

Doğrusal olarak ayırlamama durumu için DVM’de çeşitli çekirdek fonksiyonları kullanılmaktadır. Vapnik, deneylerle çekirdek fonksiyonları arasındaki seçimin ampirik performansta büyük bir farklılık göstermediğini vurgulamıştır²⁸. Daha önemli ve genellikle daha zor olan problem, çekirdek fonksiyonları kullanılırken bazı

matematiksel parametrelerin doğru tahmin edilmesidir⁹². Bu problem bir çapraz geçerlilik (birini dışarıda bırakma) prosedürü kullanarak çözülebilir fakat pratikte bu, DVM'nin kuadratik optimizasyon problemini hesaplamayı gerektirdiğinden çok zordur⁹¹.

Eğitim sürecinden önce, DVM'de kullanılan çekirdek fonksiyonunun parametrelerinin uygun bir şekilde seçilmesi gerekmektedir. Kullanıcı tarafından keyfi olarak seçilen bu parametrelerin seçilme sürecinin bir algoritmaya dayandırılması, sınıflandırmanın doğruluğu açısından önemlidir⁹². Literatürde, parçacık sürü optimizasyonu algoritması, genetik algoritmalar, ROC eğrisi, çapraz doğrulama (cross validation), yapı arama (pattern search) ve ızgara arama (grid search) gibi yöntemler, bu parametrelerin seçimi için uygun yöntemler olarak kullanılmaktadır. Bu yöntemlerden “yapı arama” ve “ızgara arama”, bölüm 2.3.4.5’de daha detaylı ele alınacaktır.

Çekirdek fonksiyonlarının özellikleri

Doğrusal olarak ayrılamayan veriler söz konusu olduğunda, DVM'de kullanılan çekirdek fonksiyonlarının özellikleri şu şekilde sıralanabilir⁷²:

- Eğer K_1 ve K_2 çekirdek fonksiyonları ve $\alpha_1, \alpha_2 \geq 0$ ise, o zaman $\alpha_1 K_1 + \alpha_2 K_2$ de bir çekirdek fonksiyondur.
- Eğer K_1 ve K_2 fonksiyonları birbirinden bağımsız çekirdek fonksiyonları ise, o zaman $(K_1 K_2)(x, \bar{x}) \equiv K_1(x, \bar{x}) K_2(x, \bar{x})$ şeklinde tanımlanabilen $K_1 K_2$ de bir çekirdek fonksiyondur.
- Eğer K_1 ve K_2 çekirdek fonksiyonları sırasıyla X_1 ve X_2 üzerinde tanımlı ise, o zaman $(K_1 * K_2)((x_1, x_2)(\bar{x}_1, \bar{x}_2)) = K_1(x_1, \bar{x}_1) K_2(x_2, \bar{x}_2)$ şeklinde tanımlanabilen kartezyen çarpım çekirdek fonksiyonu da $(x_1 * x_2) * (x_1 * x_2)$ üzerinde tanımlıdır.
- Eğer K_1 ve K_2 çekirdek fonksiyonları sırasıyla X_1 ve X_2 üzerinde tanımlı ise, o zaman $(K_1 + K_2)((x_1, x_2)(\bar{x}_1, \bar{x}_2)) = K_1(x_1, \bar{x}_1) + K_2(x_2, \bar{x}_2)$ şeklinde tanımlanabilen kartezyen toplam çekirdek fonksiyonu da $(x_1 * x_2) * (x_1 * x_2)$ üzerinde tanımlıdır.

2.3.4.4. DVM'de kullanılan çekirdek fonksiyonları

Farklı çekirdek fonksiyonlarının seçimiyle, farklı DVM'ler üretilir ve bu da farklı sınıflama performanslarına neden olabilir. Literatürde DVM'de kullanılan birçok çekirdek fonksiyon bulunmaktadır (Tablo 2.3). Ancak bu fonksiyonlardan doğrusal

fonksiyon, polinomiyal fonksiyon, sigmoid fonksiyon ve radyal tabanlı fonksiyon (RTF) en çok kullanılanlarıdır ve çoğu tahmin problemi polinomiyal ve radyal tabanlı fonksiyondan biriyle çözülebilmektedir²⁸. Polinomiyal çekirdek fonksiyon, RTF'ye göre daha fazla parametre içerir. Bu yüzden, RTF'nin hesaplanması daha kolaydır. Ancak, düşük derecelerde polinomiyal çekirdek fonksiyon ile iyi sonuçlar alınabiliyorsa polinomiyal çekirdek, sınıflandırılmayan kısımlar oluyorsa (C hata parametresini içerdiğinden dolayı) RTF kullanılmalıdır. Doğrusal çekirdek fonksiyon ise, özellik sayısının çok fazla olduğu durumlarda daha iyi sonuçlar vermektedir¹⁰.

Tablo 2.2: Yaygın Olarak Kullanılan Çekirdek Fonksiyonlar

Çekirdek fonksiyonların isimleri	Çekirdek fonksiyonların matematiksel ifadeleri
Doğrusal fonksiyon	$K(x_i, x_j) = x_i^T x_j$
Polinomiyal fonksiyon	$K(x_i, x_j) = (1 + x_i^T x_j)^d$
Sigmoid fonksiyon	$K(x_i, x_j) = \tanh(kx_i^T x_j - \delta)$
Radyal tabanlı fonksiyon	$K(x_i, x_j) = \exp\left(-\frac{\ x_i - x_j\ ^2}{2\gamma^2}\right)$
Üstel yarıçap temelli fonksiyon (üstel YTF)	$K(x_i, x_j) = \exp\left(-\frac{\ x_i - x_j\ }{2\sigma^2}\right)$
Ters İkinci Dereceden	$K(x_i, x_j) = (\ x_i - x_j\ ^2 - c^2)^{-1/2}$
B-spline	$K(x_i, x_j) = B_{2n+1}(x_i - x_j)$
Dalga Yapılı	$K(x_i, x_j) = \frac{\theta}{\ x_i - x_j\ } \sin\left(\frac{\ x_i - x_j\ }{\theta}\right)$

Doğrusal Fonksiyon

Girdi uzayında veriler doğrusal olarak ayrılabilirdiği durumda, veriyi yüksek boyuta taşımadan doğrusal çekirdek fonksiyonu yardımıyla sınıflama işlemi yapılır. Bu fonksiyonda herhangi bir boyut değeri ya da parametre söz konusu değildir.

$$K(x_i, x_j) = x_i^T x_j \quad (2.53)$$

Polinomiyal Fonksiyon

Polinomiyal çekirdek fonksiyon, d dereceli girdi vektörlerinin iç çarpımından oluşmaktadır ve aşağıdaki gibi ifade edilmektedir:

$$K(x_i, x_j) = (1 + x_i^T x_j)^d \quad (2.54)$$

Sigmoid Fonksiyon

Sigmoid fonksiyon k ve δ gibi iki parametre içermektedir. Bu fonksiyon sadece belli δ değerleri için tanımlanmıştır²⁹. Kaynaklar belirli parametreler için sigmoid fonksiyonun radyal tabanlı fonksiyon şeklinde çalıştığını göstermektedir⁹⁶.

$$K(x_i, x_j) = \tanh(kx_i^T x_j - \delta) \quad (2.55)$$

Radyal Tabanlı Fonksiyon

Radyal tabanlı fonksiyon çekirdek fonksiyonlar arasında kullanımı en yaygın çekirdek fonksiyondur ve çeşitli çalışmalarla üstünlüğü kanıtlanmıştır^{92,97}. Burada, σ çekirdek genişliğini veya yarıçap kontrolünü sağlayan parametredir⁹⁷.

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), \quad \sigma > 0 \quad (2.56)$$

Bu tez çalışmasının uygulama bölümünde, doğrusal olarak ayrılamayan verilerin sınıflandırılması için DVM’de radyal tabanlı çekirdek fonksiyon kullanılmıştır.

2.3.4.5. Hiperparametrelerin seçimi

DVM’nin başarısı, hatalı sınıflama maliyetinin ölçüsü olan C parametresinin, verilerin tamamının doğrusal ayrılamadığı durumlarda kullanılan ζ_i parametresinin ve çekirdek fonksiyonlardaki diğer parametrelerin seçimi ile doğrudan ilgilidir. Bu parametrelerin tamamı hiperparametre olarak adlandırılmaktadır ve bilinmeyen bu parametrelerin en iyi şekilde seçilmesi, DVM’nin veriyi doğru şekilde sınıflamasını sağlayacak olan en önemli etkidir. Eğitim hatası ve VC boyutu arasındaki ödünleşimi belirleyen C ve ζ_i parametreleri, uygulayıcılar tarafından seçilir⁶⁴.

Verilerin doğrusal olarak ayrılamadığı durumda, marjini maksimum yapan en iyi ayırıcı hiperdüzlemi bulmak için; düzenleme parametresi veya hata maliyeti olarak adlandırılan, C ile gösterilen bir parametre kullanılmaktadır. Bu parametre $0 < C < \infty$ aralığında yer alan ve keyfi olarak seçilen bir değerdir. Eğer bu değer çok büyük seçilirse, ayrılamayan noktalar için yüksek bir hata söz konusu olacağından çok fazla destek vektöre ihtiyaç duyulacak ve bu da aşırı uyuma (overfitting) neden olacaktır.

Eğer C değeri çok küçük seçilirse, bu durum da eksik uyuma (underfitting) neden olacaktır⁷⁷.

C parametresi belirlenirken ilk olarak, bir maliyet matrisi hazırlanır. Bu maliyet matrisi;

Tablo 2.3: Bilinmeyen C Parametresi için Maliyet Matrisi

	Tahmin Edilen Sınıf		
	C(i j)	Sınıf = evet	Sınıf = hayır
Gerçek Sınıf	Sınıf = evet	C(evet evet)	C(hayır evet)
	Sınıf = hayır	C(evet hayır)	C(hayır hayır)

şeklinde. Burada C(i|j), j sınıfına ait örneklerin, i sınıfına yanlış sınıflandırılmasının maliyetidir. Maliyet değeri hesaplanırken model doğruluğu (accuracy) da göz önüne alınmalıdır. Tablo 2.4’te maliyet matrisi ve verilen bir model matrisi için, hesaplama formüllerinin gösteriminde kullanılacak olan, harfle ifade edilen değerler yer almaktadır.

Tablo 2.4: Maliyet Matrisi ve Verilen Bir Model Matrisinin Harf ile İfade Edilmesi

Maliyet	Tahmin Edilen Sınıf		
		Sınıf=evet	Sınıf=hayır
Gerçek Sınıf	Sınıf=evet	p	q
	Sınıf=hayır	r	t

Model Matrisi	Tahmin Edilen Sınıf		
		Sınıf=evet	Sınıf=hayır
Gerçek Sınıf	Sınıf=evet	A	B
	Sınıf=hayır	C	D

Tablo 2.4 yardımıyla, doğruluk ve maliyet hesabı aşağıdaki gibi yapılır:

$$N=A+B+C+D \text{ olmak üzere}$$

$$\text{Doğruluk} = (A + D)/N \quad (2.57)$$

$$\text{Maliyet} = pA + qB + rC + tD \quad (2.58)$$

Eğer maliyet değerleri,

- $M(\text{evet}|\text{hayır}) = M(\text{hayır}|\text{evet})$
- $M(\text{evet}|\text{evet}) = M(\text{hayır}|\text{hayır})$

şeklinde yani Tablo 2.4'teki maliyet matrisinde $p=t$ ve $q=r$ ise, doğruluk maliyet ile orantılıdır. Bu durum için, Formül (2.57) ve (2.58) birlikte çözülmüşse;

$$\begin{aligned}
 \text{Maliyet} &= p(A + D) + q(B + C) \\
 &= p(A + D) + q(N - A - D) \\
 &= p(A + D) + qN - q(A + D) \\
 &= qN - (q - p)(A + D) \\
 &= N[q - (q - p) \times \text{Doğruluk}] \quad (2.59)
 \end{aligned}$$

olarak bulunur.

Örnek 2.3: İki ayrı model için (M_1 ve M_2), doğruluk ve maliyet hesabı aşağıdaki gibidir.

Tablo 2.5'te belli değerler ile hazırlanmış maliyet matrisi görülmektedir.

Tablo 2.5: Maliyet Matrisine Ait Değerler

Maliyet Matrisi	Tahmin Edilen Sınıf		
Gerçek Sınıf	C(ij)	+	-
	+	-1	100
	-	1	0

Tablo 2.6 ve 2.7'de, M_1 ve M_2 modellerine ait sınıflama matrisleri görülmektedir. Bu modeller için, Tablo 2.4'teki maliyet matrisi kullanılarak, sınıflama maliyetleri ve doğruluk oranları hesaplanırsa;

Tablo 2.6: M_1 Modeline Ait Matris

Model M_1	Tahmin Edilen Sınıf		
Gerçek Sınıf		+	-
	+	150	40
	-	60	250

$$N = 150 + 40 + 60 + 250 = 500$$

- Doğruluk (M_1) = $(150 + 250)/500$
= 0.80
- Maliyet (M_1) = $(-1)150 + (100)40 + (1)60 + (0)250$
= $-150 + 4000 + 60 + 0$
= 3910

Tablo 2.7: M_2 Modeline Ait Matris

Model M_2	Tahmin Edilen Sınıf		
Gerçek Sınıf		+	-
	+	250	45
	-	5	200

$$N = 250 + 45 + 5 + 200 = 500$$

- Doğruluk (M_2) = $(250 + 200)/500$
= 0.90
- Maliyet (M_2) = $(-1)250 + (100)45 + (1)5 + (0)200$
= $-250 + 4500 + 5 + 0$
= 4255

Örnek 2.3'teki sonuçlara bakıldığında, M_1 modelinde hem doğruluk hem de maliyet, M_2 modeline göre daha düşüktür. Aslında istenen, doğruluğun yüksek, maliyetin düşük olduğu modeli seçmektir. Bu durumda, doğruluk ve maliyetin dengede olduğu model seçilmelidir.

Verilerin tamamının doğrusal olarak ayrılamadığı durumda, doğru olarak seçilmesi gereken parametre ise negatif olmayan, gevşek değişken adı verilen ξ_i değişkenidir ($\xi_i \geq 0$, $i=1, \dots, n$). Bu gevşek değişken, uygun olan sınıfın sınırdan sapsmasıdır ve minimize edilmiş bir hata derecesine izin verir. Bu parametrenin doğru seçilebilmesi için, sapan değer bilgisine ihtiyaç duyulmaktadır. Çünkü ξ parametresi, sapan değerlerin varyansı ile orantılıdır ve doğru ξ parametresi için sapan değerlerin varyansının tahmini gerekmektedir. Bu varyans değeri, çok esnek bir tahmin edici kullanılarak, eğitim

verisinden tahmin edilebilir^{98,99,100}. Gevşek değişkenin tahmini değerinin elde edilmesi için aşağıdaki adımlar da kullanılabilir:

1. Karşıt sınıflardan rastgele çiftler seçilir.
2. Bu çiftler arasındaki uzaklık değeri hesaplanır.
3. Uzaklık değerleri büyükten küçüğe doğru sıralanır.
4. %90'lık yüzdeler değeri seçilir.

DVM' de belirlenmesi gereken C , ξ_i ve kullanılan çekirdek fonksiyonlarındaki bilinmeyen parametrelerin en uygun değerinin seçimi için geliştirilmiş çeşitli yöntemler mevcuttur. Bunlardan en yaygın kullanılan ve paket programlarda adı geçen yöntemlerden ikisi ızgara arama (grid search) ve yapı arama (pattern search) yöntemleridir.

Izgara arama yöntemi, her bir parametre değerini, belli bir arama aralığı üzerinden geometrik adımları kullanarak dener. Izgara arama, hesaplama yönünden zaman alıcı bir yöntemdir. Çünkü modelin her bir parametre için, ızgara içinde birçok noktada değerlendirilmesi gerekmektedir. Örneğin, 10 arama aralığına sahip bir ızgara arama kullanıldıysa ve iki parametreye sahip RTF çekirdek fonksiyonu söz konusu ise, bu durumda model $10*10=100$ ızgara noktası ile değerlendirilecektir. Eğer her model için çapraz doğrulama kullanılırsa, gerçek DVM hesaplamalarının sayısı, çapraz doğrulama katıyla (genellikle 4 ile 10 arasında) çarpılarak elde edilecektir. Izgara arama yönteminin bu hesapsal karmaşıklığından dolayı, parametrelerin az olduğu durumlarda kullanılması uygundur.

Yapı arama ("compass search", "line search") yöntemi ise, arama aralığının ortasından başlar ve her bir parametre için her talimatta deneme adımlarını oluşturur. Eğer model uygunluğu artarsa, arama merkezi bu yeni nokta olur ve süreç tekrar edilir. Eğer bir gelişme gözlenmezse, adım büyüklüğü azaltılır ve arama tekrarlanır. Arama adım büyüklüğü belirlenen tolerans değerine kadar düştüğünde, yapı arama sonlanır. Bu yöntem genellikle ızgara arama yöntemine göre daha az değerlendirme gerektirmektedir. Ancak, yapı arama yönteminin zayıflığı, bu yöntemin parametreler için global optimal nokta yerine yerel (lokal) optimal noktayı bulabilmesidir⁸².

Örnek 2.4: Polinomial çekirdek fonksiyonun kullanıldığı bir sınıflama örneği aşağıda verilmiştir¹⁰¹.

Sadece bir özelliğin ölçüldüğü ve 2 grubun bulunduğu bir veri seti ele alınsın. Bu veri setinde toplam beş bireye ait veriler; $x_1=1, x_2=2, x_3=4, x_4=5, x_5=6$ olarak tanımlansın ve bu bireylerden $x_1=1, x_2=2$ ve $x_5=6$ grup 1'e, $x_3=4$ ve $x_4=5$ bireyleri ise grup 2'ye ait olsun. Ait oldukları sınıfların kodları 1. grup için $y_i=1$ ve 2. grup için $y_i=-1$ ile gösterilsin.

Bu örnekte derecesi 2 olan polinamiyal fonksiyon kullanılmış ve $C=100$ olarak belirlenmiştir. Bu durumda kullanılacak çekirdek fonksiyon, $K(x_i, x_j) = (x_i^T x_j + 1)^2$ şeklinde olacaktır.

İlk olarak; $0 \leq \alpha_i \leq 100$ ($0 \leq \alpha_i \leq C$) ve $\sum_{i=1}^5 \alpha_i y_i = 0$ kısıtları altında, aşağıdaki formül yardımıyla, α_i $i=1, \dots, 5$ değerleri belirlenmelidir.

$$\sum_{i=1}^5 \alpha_i - \frac{1}{2} \sum_{i=1}^5 \sum_{j=1}^5 \alpha_i \alpha_j y_i y_j (x_i^T x_j + 1)^2$$

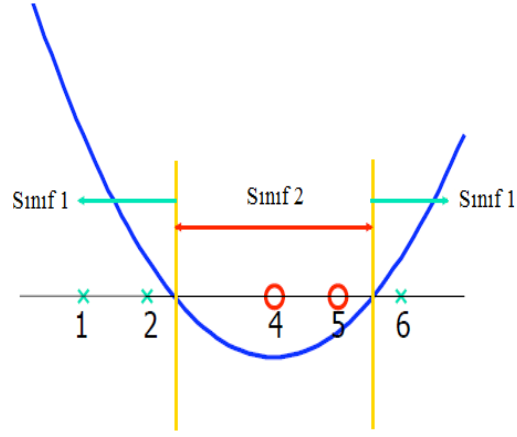
Buradan, karesel programlamayla α değerleri; $\alpha_1=0, \alpha_2=2.5, \alpha_3=0, \alpha_4=7.333$ ve $\alpha_5=4.833$ olarak bulunur. Bu durumda destek vektörleri; $x_2=2, x_4=5, x_5=6$ ve karar fonksiyonu aşağıdaki gibi olur:

$$\begin{aligned} f(z) &= 2.5(1)(2z + 1)^2 + 7.333(-1)(5z + 1)^2 + 4.833(1)(6z + 1)^2 + b \\ &= 0.6667z^2 + 5.333z + b \end{aligned}$$

b değeri ise; $f(2)=1, f(5)=-1$ veya $f(6)=1$ işlemlerinden biri çözülerek bulunabilir.

$$f(2)=1 \Rightarrow 0.6667(2)^2 + 5.333(2) + b = 1 \Rightarrow b = 9 \text{ olarak elde edilir.}$$

$f(z)$ karar fonksiyonunda z ve z^2 değerleri (yeni gelecek bir bireyin ölçülen x değeri ve bu değer karesi) yerine konulduğunda, $f(z)$ değerleri elde edilir. Çeşitli z değerleri için aşağıdaki grafik çizilebilir. Burada, destek vektörleri olan x_2 ve x_5 'in aynı, x_4 'ün ise farklı sınıfa ait olduğu görülmektedir.



Şekil 2.31: Örnek 3 için Karar Fonksiyonu

Şekil 2.31’de, Örnek 2.3 için bulunan karar fonksiyonunun sınıflandırması gösterilmiştir. Burada; 1., 2. ve 6. noktaların sınıf 1’e, 4. ve 5. noktaların ise sınıf 2’ye ait olduğu görülmektedir.

Örnek 2.5: İki sınıflı, her bir sınıfta 2 birey olan ve her bir bireyden 2 özelliğin ölçüldüğü XOR problemi aşağıdaki gibi çözülebilir:

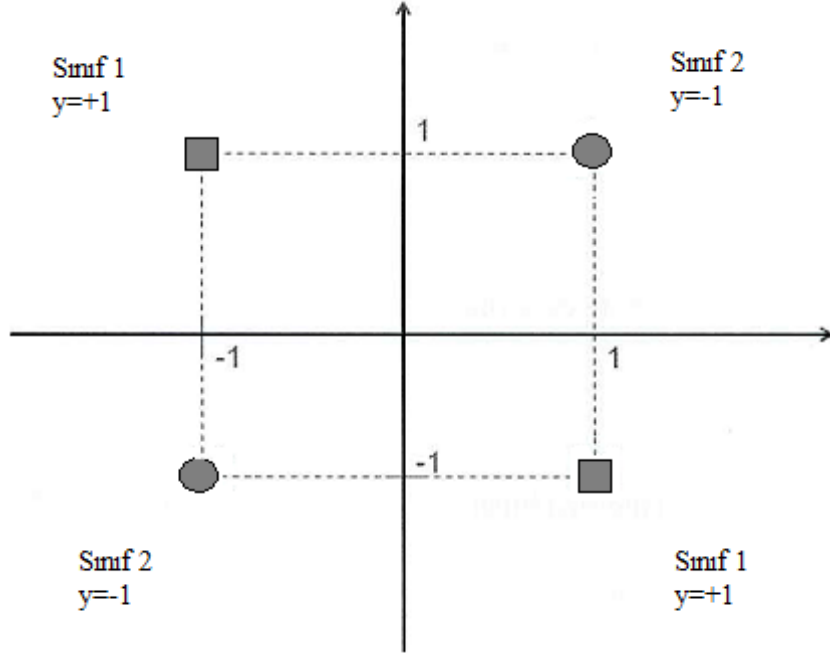
$$\text{sınıf 1} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} (\mathbb{R}^2 \text{de}),$$

$$\text{sınıf 2} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} -1 \\ -1 \end{pmatrix} (\mathbb{R}^2 \text{de})$$

olarak tanımlansın. Bu durumda ait oldukları sınıfların kodları ise,

$$y = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}$$

olsun. Şekil 2.32’de bu noktaların analitik düzlemdeki gösterimi görülmektedir.



Şekil 2.32: Verilen XOR Verilerinin Analitik Düzlemde Gösterimi

Önceki bölümlerde tanımlanan ve karar fonksiyonunun elde edilmesinde kullanılan aşağıdaki dual formun ikinci kısmındaki $y_i y_j K(x_i, x_j)$ ifadesi, çekirdek fonksiyonu olma özelliğini sağlayan Hessian matrisinin hesaplanması için kullanılır.

$$L_D = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

ve

$$y_i y_j K(x_i, x_j) = H_{ij}$$

Bu durumda, dual form aşağıdaki şeklini alacaktır:

$$L(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha^T H \alpha$$

İkinci dereceden polinom çekirdek fonksiyonu,

$$K(x_i, x_j) = (1 + x_i^T x_j)^2$$

olduğu bilinmektedir. Bu durumda Hessian matrisi aşağıdaki gibi hesaplanır:

$$H_{11} = y_1 y_1 (1 + x_1^T x_1)^2 = (1)(1) \left(1 + [1 \ -1] \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right)^2 = 9$$

$$H_{12} = y_1 y_2 (1 + x_1^T x_2)^2 = (1)(1) \left(1 + [1 \ -1] \begin{bmatrix} -1 \\ 1 \end{bmatrix}\right)^2 = 1$$

.

.

.

Bu şekilde bütün H_{ij} değerleri hesaplanırsa, aşağıdaki Hessian matrisi elde edilir:

$$\begin{bmatrix} 9 & 1 & -1 & -1 \\ 1 & 9 & -1 & -1 \\ -1 & -1 & 9 & 1 \\ -1 & -1 & 1 & 9 \end{bmatrix}$$

Buradan α değerlerini elde etmek için $L(\alpha)$ 'nın α 'ya göre türevi alınarak sıfıra eşitlenir.

$$\frac{\partial L}{\partial \alpha} = I - H\alpha = 0$$

Bu durumda,

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 9 & 1 & -1 & -1 \\ 1 & 9 & -1 & -1 \\ -1 & -1 & 9 & 1 \\ -1 & -1 & 1 & 9 \end{bmatrix} \alpha = 0$$

olur ve bu denklem sisteminin çözülmesiyle $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0.125$ şeklinde elde edilir. Bu durumda tüm örnekler destek vektörler olarak kabul edilir. Elde edilen bu sonuçlar aşağıda belirtilen kısıtları sağlamaktadır:

$$\sum_{i=1}^n \alpha_i y_i = 0 \quad \Rightarrow \quad \sum_{i=1}^4 \alpha_i y_i = \alpha_1 + \alpha_2 - \alpha_3 - \alpha_4 = 0$$

ve

$$\alpha_i \geq 0$$

$K(x_i, x_j) = (1 + x_i^T x_j)^2$ ikinci derece çekirdek fonksiyonu için, $\Phi(x)$ eşlemesi son bağıntıdan dolayı aşağıdaki gibidir:

$$\Phi(x_i) = [1 \quad x_{i1}^2 \quad \sqrt{2}x_{i1}x_{i2} \quad x_{i2}^2 \quad \sqrt{2}x_{i1} \quad \sqrt{2}x_{i2}]^T$$

Bu durumda ağırlık vektörü olan w aşağıdaki formül yardımıyla hesaplanır:

$$w = \sum_{i=1}^4 \alpha_i y_i \Phi(x_i)$$

$$\begin{aligned} w &= \alpha_1 y_1 \Phi(x_1) + \alpha_2 y_2 \Phi(x_2) + \alpha_3 y_3 \Phi(x_3) + \alpha_4 y_4 \Phi(x_4) \\ &= 0.125(1)\Phi(x_1) + 0.125(1)\Phi(x_2) + 0.125(-1)\Phi(x_3) + 0.125(-1)\Phi(x_4) \\ &= \frac{1}{8} [\Phi(x_1) + \Phi(x_2) - \Phi(x_3) - \Phi(x_4)] \end{aligned}$$

$$= \frac{1}{8} \begin{bmatrix} 1 \\ 1 \\ -\sqrt{2} \\ 1 \\ \sqrt{2} \\ -\sqrt{2} \end{bmatrix} + \frac{1}{8} \begin{bmatrix} 1 \\ 1 \\ -\sqrt{2} \\ 1 \\ \sqrt{2} \\ -\sqrt{2} \end{bmatrix} + \frac{1}{8} \begin{bmatrix} 1 \\ 1 \\ -\sqrt{2} \\ 1 \\ \sqrt{2} \\ -\sqrt{2} \end{bmatrix} + \frac{1}{8} \begin{bmatrix} 1 \\ 1 \\ -\sqrt{2} \\ 1 \\ \sqrt{2} \\ -\sqrt{2} \end{bmatrix}$$

$$w = \begin{bmatrix} 0 \\ 0 \\ -\sqrt{2}/2 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Burada, w 'nin ilk elemanı b 'yi verir. O halde, $b = 0$ 'dır.

Optimal hiperdüzlem; $w^T \Phi(x) + b = 0$ olarak verilmişti ve bu problem için $b=0$ olduğundan, $w^T \Phi(x) = 0$ 'dır. O halde karar fonksiyonu aşağıdaki gibi olacaktır:

$$f(x) = \begin{bmatrix} 0 & 0 & -\frac{\sqrt{2}}{2} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \\ \sqrt{2}x_1 \\ \sqrt{2}x_2 \end{bmatrix}$$

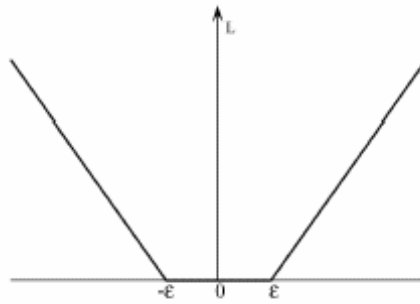
$$f(x) = -x_1x_2$$

2.3.5. Destek vektör regresyonu

Şimdiye kadar olan kısımda, destek vektör makineleri ile sınıflama yönteminin temellerinden ve işleyişinden bahsedilmiştir. Bu bölümde ise, destek vektör regresyonundan (DVR) kısaca bahsedilecektir.

Destek vektör makineleri ile sınıflamada DVM, belirleyici nitelik uzayındaki eğitim verilerine göre optimum ayırma düzlemini öğrenmeye çalışır. Regresyonda ise, mevcut eğitim verilerine göre giriş ve çıkış uzayları arasındaki haritalama fonksiyonunu öğrenmeye çalışır. DVR'nin en önemli özelliklerinden birisi, gözlenen eğitim hatasını minimize etmek yerine, genelleştirilmiş hata sınırını minimize etmeye çalışmasıdır. Bu genelleme hata sınırı, eğitim hatasının ve hipotez uzayının karmaşıklığının kontrol eden düzeltme teriminin bir kombinasyonudur¹⁰².

Destek vektör regresyonunda, optimal bir ayırıcı hiperdüzlem ve destek vektörleri arasındaki marj yerine bir “yaklaşım hatası” kullanılmaktadır. Destek vektör sınıflamasında olduğu gibi, regresyon problemi için de sınır üzerindeki optimum genelleme araştırılmaktadır ve sonlu bir gürültülü örnek kümesi üzerinde bilinmeyen sürekli fonksiyonun tahmin edilmesi amaçlanmaktadır²⁸. Sonuç olarak DVR’de temel fikir, marjini maksimumlaştıran hiperdüzlem ile hatayı minimize etmektir. Ancak burada hatanın bir kısmı tolere edilebilmektedir. Şekil 2.33’te DVR’ye ait hatanın nasıl hesaplandığı görülmektedir¹⁰³. Regresyonda veriyi modelleyen $f(x) = wx + b$ fonksiyonu ve gerçek değerli y_i hedefleri arasındaki sapma değeri olan ε ’a her iki yönde de izin vermek için $y_i - wx_i - b \leq \varepsilon$ ve $wx_i + b - y_i \leq \varepsilon$ şekline getirilir. Bu, $f(x)$ hipotez fonksiyonu etrafında $\pm \varepsilon$ boyutlu bir tüp olarak gösterilebilir ve bu tüpün dışında bulunan bir nokta eğitim hatasıdır. İki eğitim hatası tipi için ζ_i ve ζ_i^* gevşek değişkenleri tanımlandığında, bu gevşek değişkenler tüpün iç tarafında kayıp fonksiyonuna göre sıfırdır ve tüpün dışındaki noktalar için de doğrusal olarak artar^{29,104}.



Şekil 2. 33: Lineer ε Kayıp Fonksiyonu

Regresyonda, rasgele bir değişken olan sistem çıktısı, gerçek değerler kümesinden değerler alır ve böylece kararlı bir fonksiyon toplamı olarak yorumlanabilir.

$$y = g(x) + \varepsilon \quad (2.60)$$

Sonuç olarak; temel prensipte DVM sınıflaması ile benzer özellikler gösteren DVM regresyonunda, tahminler aşağıdaki formül yardımıyla yapılmaktadır:

$$f(x) = \sum \theta_i \Phi(x, x_i) + b \quad (2.60)$$

2.3.6. Destek vektör makinelerinin avantajları

DVM yönteminin avantajları aşağıdaki gibi sıralanabilir:

1. DVM’de yüksek doğruluk elde edilmektedir.
2. Parametreler uygun olarak seçildiğinde, DVM iyi bir genelleme yeteneğine sahip olmaktadır. Bu ifade; uygun genelleme düzeyi seçildiğinde, eğitim setinde bir takım bias’lar olsa bile, DVM’nin robust olabileceği anlamına gelmektedir.
3. Çok sayıda doğal nitelikle (bağımsız değişken) çalışabilmektedir.
4. Karmaşık (komplike) karar sınırlarını modelleyebilme yeteneğine sahiptir.
5. Hem doğrusal olarak ayrılabilen hem de doğrusal olarak ayrılamayan verilere uygulanabilmektedir.
6. Çekirdek fonksiyonlar dolaylı olarak doğrusal olmayan dönüşümler içerdiğinden, veriyi doğrusal olarak ayrılmasını sağlayan bu dönüşümle ilgili herhangi bir varsayım gerektirmemektedir ve bu dönüşüm robust teorik temellere dayanmaktadır.
7. Aşırı uyum sorunu diğer yöntemlerden daha azdır.
8. Optimallik problemi konveks olduğunda, DVM tek bir çözüm sunmaktadır. Bu durum, lokal minimumla ilgili çoklu çözümlere sahip olan ve bu nedenle farklı örnekler üzerinde robustluğu kimi zaman sağlayamayan yapay sinir ağlarıyla kıyaslandığında, DVM için bir avantajdır.

2.3.7. Destek vektör makinelerinin dezavantajları

DVM yönteminin dezavantajları aşağıdaki gibi sıralanabilir:

1. DVM’nin en önemli ayrıntısı olan çekirdek fonksiyonların ve parametrelerin uygun olarak seçilmesi gerekmektedir.
2. Eğitim ve test sırasında, hız ve boyut sorunu meydana gelebilmektedir.
3. DVM’de algoritma oldukça karmaşıktır ve karesel programlama kullanıldığından güçlü bilgisayar teknolojisine gerek duyulmaktadır.

3. GEREÇ ve YÖNTEM

Bu bölümde, öncelikle tezde kullanılan veri tanıtılıp, veri üzerinde yapılan işlemler ayrıntılı olarak açıklanacaktır. Yöntem bölümünde ise veri analizinde kullanılan DTREG programı tanıtılacaktır.

3.1. Veriler

Tez çalışmasında kullanılan veri seti, Zonguldak Karaelmas Üniversitesi Tıp Fakültesi psikiyatri polikliniğine, 1-31 Ocak 2011 tarihleri arasında gece yeme sendromu şikâyetiyle ayaktan başvuran ve çalışmaya katılmayı kabul eden 433 hastanın bilgilerini içermektedir. Verilerin tezde kullanımı için izin alınmıştır. Çalışmaya alınan kişilerden, yaş, cinsiyet, eğitim yılı, kardeş sayısı, medeni durum, çocuk sayısı gibi demografik özellikler, bel çevresi, kalça çevresi, beden kitle indeksi, fiziksel hastalık varlığı, psikolojik hastalık varlığı gibi çeşitli morfolojik ve klinik özellikler sorgulanmıştır. Ayrıca kliniğe başvuran kişilere gece yeme alışkanlığı var veya yok şeklinde klinik tanı konmuştur. Başvuran 433 hastanın 97'si klinik görüşmelerle gece yeme sendromu tanısı almış, geriye kalan 336 kişi ise gece yeme sendromu tanısı almamıştır. Bu sonuç gold standart olarak dikkate alınmıştır. Klinik tanı aşağıdaki kriterlere göre konulmuştur.

Gece yeme bozukluğu tanı kriterleri:

1. Günlük yeme paterninin akşam ve gece önemli ölçüde artış göstermesi. Aşağıdakilerden biri veya ikisiyle karakterizedir:
 - a. Günlük yenen yemeğin en az %25'i akşam yemeğinden sonra olması
 - b. Haftada en az iki gece yeme epizodunun olması
2. Akşam ve noktürnal yeme epizodlarına ilişkin farkındalık ve hatırlamanın söz konusu olması
3. Klinik aşağıdakilerden en az üçüyle karakterizedir:
 - a. Sabah yeme isteğinin olmaması ve/veya haftada 4 veya daha fazla sabah kahvaltının atlanması
 - b. Akşam yemeğinden uyku başlayıncaya kadar olan sürede ve/veya gece boyunca güçlü yeme arzusunun varlığı

- c. Haftada 4 veya daha fazla gece uyku başlangıcında ve/veya uykunun sürdürülmesinde insomnianın olması
 - d. Uykuyu başlatabilmek veya tekrar uykuya dalabilmek için kişinin yemesinin gerektiği inancının olması
 - e. Duygu durumun çoğunlukla depresif olması ve/veya duygu durumun akşamları kötüleşmesi
4. Bozukluk önemli derecede sıkıntı ve/veya fonksiyonellikteki bozulmayla ilişkilidir.
 5. Bozulmuş yeme paterni en az 3 ay sürmelidir.
 6. Bozukluk madde kötüye kullanımı veya bağımlılığına, tıbbi durumlara, medikasyona veya diğer psikiyatrik bozukluklara sekonder gelişmiş olmamalıdır.

Hastalara ayrıca, gece yeme anketi (GYA), beden şekli anketi (BSQ), semptom tarama listesi (SCL-90) ölçekleri de uygulanmıştır.

Sorgulama veya ölçme yoluyla elde edilen değişkenlerin özellikleri aşağıdaki gibidir:

- *GYStanı*: GYStanı değişkeni, hastalarla klinik görüşmeler yapılarak elde edilmiş 2 kategorili, bağımlı sonuç değişkenidir. Yapılan görüşmeler neticesinde, gece yeme sendromu tanısı alan ve almayanlar, 0 (tanı almadı) ve 1 (tanı aldı) şeklinde kodlanmıştır. Bu tanı gold standart kabul edilmektedir ve bu çalışmada bağımlı değişken olarak değerlendirilmiştir.
- *Yaş*: Hastalara doğrudan yaşları sorularak elde edilmiş sürekli bir değişkendir.
- *Cinsiyet*: Kadın ve erkek şeklinde iki kategorisi olan kategorik bir değişkendir.
- *Eğitim yılı*: Kişinin kaç yıllık bir eğitim aldığını gösteren sürekli bir değişkendir.
- *Kardeş sayısı*: Kişinin toplam kaç kardeşe sahip olduğunu gösteren, sürekli bir değişkendir.
- *Medeni durum*: Kişinin medeni halini belirten ve evli, bekar ve dul şeklinde üç kategorisi olan kategorik bir değişkendir.

- *Çocuk sayısı grubu:* Sahip olunan çocuk sayısı grubu olup 5 kategorili kategorik bir değişkendir.
- *Kilo vermek:* Kişinin daha önce kilo vermek için uğraşp uğraşmadığını sorgulayan, evet (1) ve hayır (2) şeklinde 2 kategorisi olan kategorik bir değişkendir.
- *Bel çevresi:* Kişinin bel çevresi ölçüsünü (cm) gösteren, sürekli bir değişkendir.
- *Kalça çevresi:* Kişinin kalça çevresi ölçüsünü (cm) gösteren, sürekli bir değişkendir.
- *Beden kitle indeksi:* Kişinin boy uzunluğu ile ağırlığı arasındaki oranı ($BKI = \text{ağırlık}/(\text{boy})^2$) belirten sürekli bir değişkendir.
- *Sigara:* Kişilerin sigara kullanıp kullanmadığını gösteren, kullanıyor (1) ve kullanmıyor (0) şeklinde iki kategorisi olan kategorik bir değişkendir.
- *Fiziksel hastalık:* Kişilerin fiziksel bir rahatsızlığı olup olmadığını belirten, var (1) ve yok (0) şeklinde iki kategorisi olan kategorik bir değişkendir.
- *Psikiyatrik tanı:* Kişilerin psikiyatrik bir tanısı olup olmadığını belirten, majör depresyon (1), anksiyete bozukluğu (2), Bipolar Afektif bozukluk (3), psikotik bozukluk (4) ve diğer (5) olmak üzere beş kategorisi olan kategorik bir değişkendir.
- *Antidepresan:* Kişilerin antidepresan kullanıp kullanmadığını gösteren, alıyor(1) ve almıyor (0) şeklinde iki kategorisi olan kategorik bir değişkendir.
- *Gece yeme anketi:* Gece Yeme Anketi 14 sorudan oluşan bir tarama anketidir ve anket sabah iştah ve gıda alımı, akşam ve gece yemeleri, aşermeler, uyanmalar, gece yemeleri sırasında farkındalık ve duygudurum ile ilgili sorular içermektedir¹⁰⁵. Her madde 1 ila 4 arasında puanlanmaktadır ve likert tipi bir ölçektir. Bu anketin sonucunda, elde edilen toplam puanlar GYA_puan olarak adlandırmıştır ve bu değişken sürekli bir değişkendir.
- *Vücut şekli anketi (BSQ):* BSQ vücut şekli ve kilo konusundaki endişeleri değerlendiren bir ankettir¹⁰⁶. Kişinin kendi doldurduğu 34 soruluk bir test olup, 6 puan üzerinden değerlendirilmektedir (1: asla, 2:sıklıkla, 3:bazen, 4: sıklıkla, 5:çok sık, 6:daima). Orijinal formunda bulimia hastaları için ortalaması 136.9, bulimia olmayanlar için 71.9 olarak saptanmıştır. Bu

anketin sonucunda, elde edilen toplam puanlar BSQ_puan olarak adlandırmıştır ve bu değişken sürekli bir değişkendir.

- *Semptom tarama listesi (SCL-90)*: SCL-90, bireylerdeki psikolojik belirtilerin ne düzeyde bulunduğunu ve bu belirtilerin hangi alanlarla ilgili olduğunu ölçen bir araçtır. Bu amaçla araç, psikolojik (ruhsal) belirti ve yakınmaları içeren 90 maddeyle 10 ayrı belirti boyutunda (somatizasyon, obsessif kompulsif belirtiler, kişiler arası duyarlılık, depresyon, anksiyete, öfke düşmanlık, fobik anksiyete, paranoid düşünceler, psikotizm, ek skalalar) değerlendirme yapma üzere yapılandırılmıştır. SCL-90, alt boyutları ayrılmaksızın bir bütün halinde ve her durumun (maddenin) “son bir ay içinde kişiyi ne ölçüde huzursuz ve tedirgin ettiğini” değerlendirmelerini isteyen bir yönergeyle uygulanan 5’li (0 Hiç, 1 Çok az, 2 Orta derecede, 3 Oldukça fazla, 4 Aşırı düzeyde) likert tipi bir ölçektir. Ölçekte boş bırakılanlar hariç tüm maddelere verilen puanların ortalaması “Genel Belirti Düzeyini (Global Symptom Index) GSI” verir. Bireyin ölçekten aldığı puanın yüksekliği, bireyin daha ileri düzeyde psikolojik belirtilere sahip olduğu anlamına gelmektedir. SCL-90 ölçeği uygulanarak elde edilmiş puan ortalamaları, ort_scl olarak adlandırılmıştır ve bu değişken sürekli bir değişkendir.

Tez çalışmasının uygulama bölümünde, kişilere ait yukarıda tanımlanan sosyo-demografik özellikler, bazı fiziksel ölçümler ve psikolojik durumları gösteren ölçek puanları kullanılarak, gece yeme alışkanlığı tanısı konulmaya çalışılmıştır. Burada amaç, klinik olarak konulan tanı ile bu özelliklere göre konulan tanının uyumunu ölçmek ve söz konusu özelliklerin tanı başarılarını belirlemektir. Tanı başarısının yüksek olması durumunda klinik tanıya gerek duyulmadan geliştirilen bir model yardımıyla başarılı bir tanı konulabilecektir. Bu amacı gerçekleştirmek için 17 doğal nitelik yardımıyla doğrusal olmayan DVM yöntemi kullanılmıştır.

Veri analizinin ilk aşamasında, amaca uygun olarak çeşitli tek değişkenli klasik istatistik yöntemler kullanılmış ve elde edilen sonuçlar bulgular bölümünün girişinde özetlenmiştir.

Tablo 3.1: Modelde Kullanılan Değişkenlerin Listesi ve Özellikleri

No	Değişkenin Adı	Değişkenin Rolü	Değişkenin Tipi	Eksik gözlem olan satır sayısı	Kategoriler
1	GYSani	Bağımlı	Kategorik	0	2
2	yas	Bağımsız	Sürekli	0	-
3	cinsiyet	Bağımsız	Kategorik	0	2
4	egitimyili	Bağımsız	Sürekli	0	-
5	kardessayisi	Bağımsız	Sürekli	58	-
6	medenidurum	Bağımsız	Kategorik	0	3
7	cocuksayisi	Bağımsız	Sürekli	1	-
8	kilovermek	Bağımsız	Kategorik	21	2
9	belcevresi	Bağımsız	Sürekli	3	-
10	kalçacevresi	Bağımsız	Sürekli	3	-
11	BKI	Bağımsız	Sürekli	2	-
12	sigara	Bağımsız	Kategorik	0	2
13	fizikselhast	Bağımsız	Kategorik	3	2
14	psktani	Bağımsız	Kategorik	0	5
15	antidepresan	Bağımsız	Kategorik	25	2
16	GYA_puan	Bağımsız	Sürekli	0	-
17	BSQ_puan	Bağımsız	Sürekli	0	-
18	ort_scl	Bağımsız	Sürekli	12	-

3.2. Yöntem

Veri analizinin ilk aşamasında, klinik tanı alan ve almayan gruplarda yapılan ölçümlere ait tanımlayıcı değerler ortalama \pm SD, sayı ve % frekanslar olarak tablo halinde verilmiştir. Ayrıca 17 değişken bakımından klinik tanı alan ve almayan gruplar arasındaki farklılıklar tek değişkenli analizlerden student t-testi veya ki-kare analizlerinden uygun olanı ile incelenmiştir.

İkinci aşamada, çeşitli demografik ve klinik özelliklerden yararlanılarak gece yeme alışkanlığı tanısını koymak amacıyla multiple binary lojistik regresyon ve doğrusal olmayan DVM yöntemleri kullanılmıştır. Multiple binary lojistik regresyon modeli, günümüzde pratik uygulamalarda en yaygın kullanılan veri madenciliği yöntemi olduğu için doğrusal olmayan DVM yönteminin sonuçlarıyla karşılaştırmak amacıyla tercih edilmiştir. Doğrusal olmayan DVM yöntemine ait karar fonksiyonunun oluşturulmasında, pratikte yaygın kullanılan 4 çekirdek fonksiyondan yararlanılmıştır.

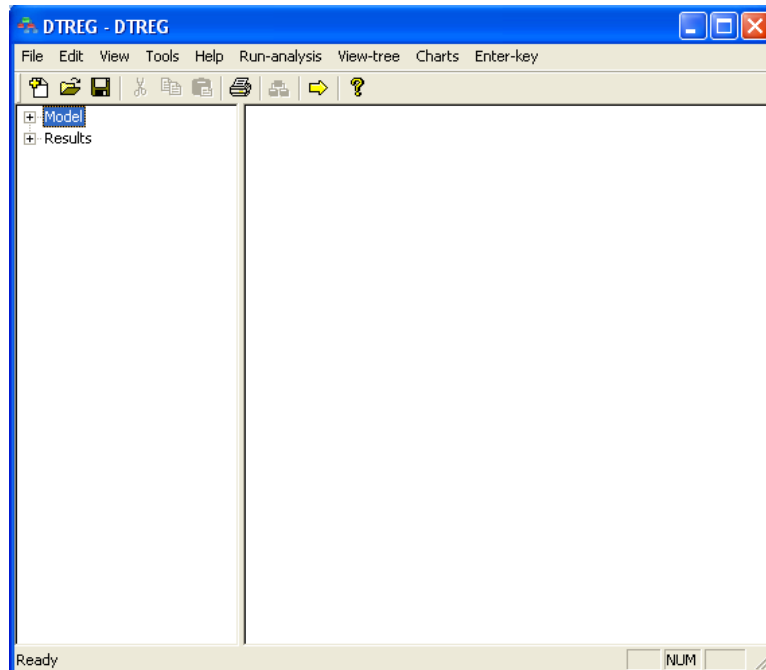
Eksik/kayıp verilerin tahmininde *surrogate deęişkenlerden* yararlanılmıştır. Tüm veriler içinde 102 kayıp veri bulunmuştur. Model parametrelerinin tahmininde, ızgara arama ve yapı arama yöntemleri kombine edilerek kullanılmıştır. Bu işlem esnasında önce ızgara arama yöntemi uygulanmış ve ızgara arama yöntemi sonlandığında, bu yöntemin bulunduğu en iyi nokta etrafında, dar bir arama aralığında yapı arama yöntemi ile en iyi nokta bulunmuştur. Izgara arama yöntemi global optimum noktanın yakınında bir bölge bulur ve daha sonra yapı arama yöntemi bu bölgede arama yaparak global optimum noktayı bulur. Optimum parametre değerlerinin seçilmesinde duyarlılık ve seçicilik değerlerini maksimize eden kombinasyon dikkate alınmıştır. Ayrıca parametre tahmin sürecinde 4 katlı çapraz geçerlilik yöntemi tercih edilmiştir.


İncelenen modeller eğitim setinde geliştirilmiş ve hem eğitim hem de test setinde model başarıları hesaplanmıştır. Eğitim ve test setlerinin oluşturulmasında 10-katlı çapraz geçerlilik yöntemi kullanılmıştır.


Elde edilen farklı modellerin performanslarının karşılaştırılmasında bağımlı iki oran arasındaki farka ait t testi kullanılmıştır.

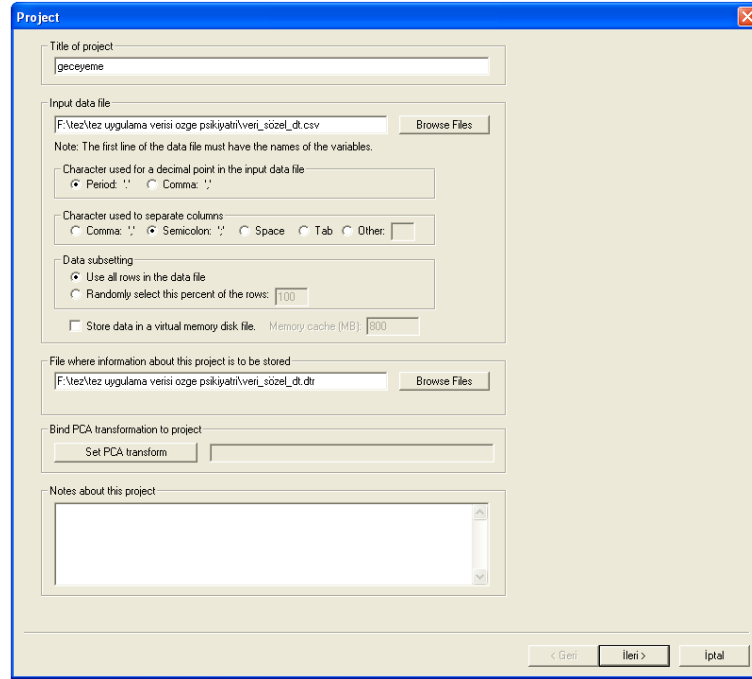
DVM yöntemi için DTREG paket programı kullanılmış, DVM analizi için programda yer alan menüler ve analiz aşamaları aşağıda adım adım verilmiştir:

Adım 1: DTREG programı açıldığında aşağıdaki ekran gözükür:

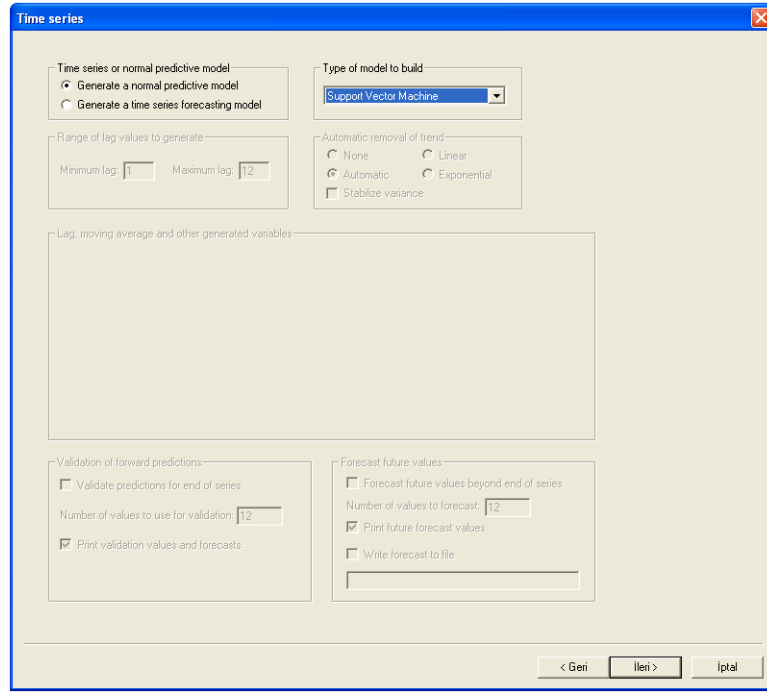


Bu ekranda,  ikonu tıklanarak veri seti programa girilir.

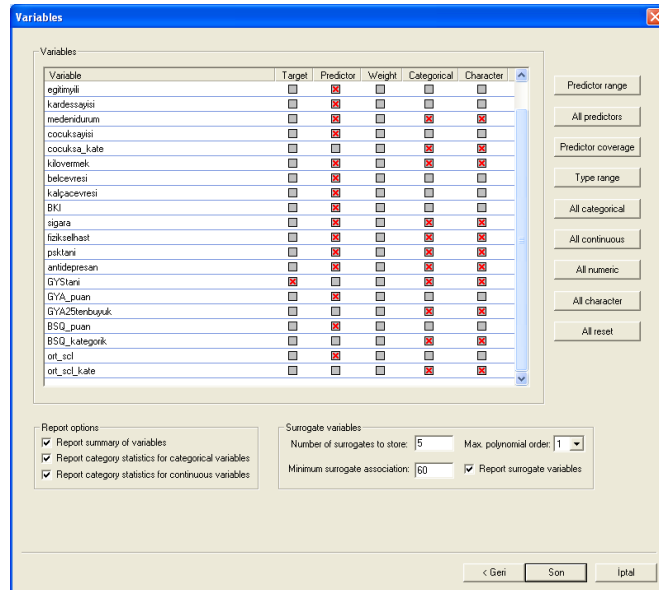
Adım 2: Veriyi programa okutmak için, çeşitli programlarda veriyi uygun şekilde kaydetmek gerekmektedir. Bunlardan biri, Excel programıdır ve Excel’de uygun şekilde girilmiş olan verinin, DTREG programında okutulabilmesi için verinin Excel’de "csv (virgülle ayrılmış)" şeklinde kaydedilmesi gerekmektedir. “.csv” uzantısıyla kaydedilmiş veri dosyası,  ikonu kullanılarak kayıt edilen yerden çağırılabilir ve bu ikon tıklandığında aşağıdaki pencere görülür:



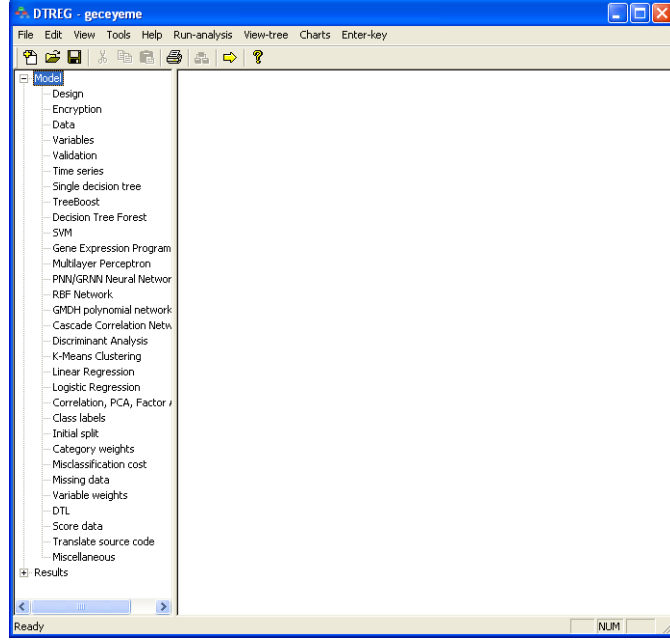
Adım 3: Açılan bu pencerede ilk görülen “Title of project” satırına, dosyaya verilmek istenen herhangi bir isim yazılır. “Input data file” satırında ise, *browse files* sekmesi tıklanarak, veri bulunduğu yerden seçilerek programa okutulur. Daha sonra *ileri* sekmesi tıklanır ve aşağıdaki pencere görülür:



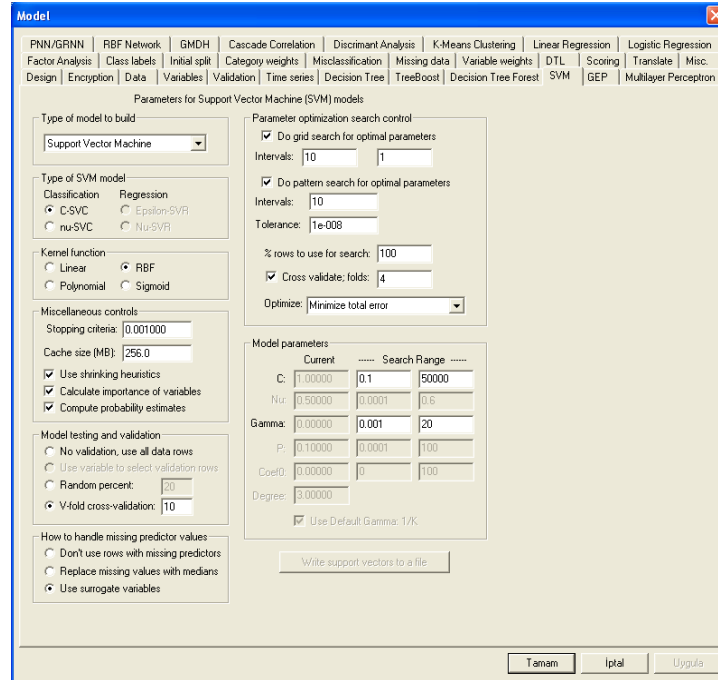
Burada “Type of the model to built” seçeneğinden *support vector machine* seçilir ve ileri sekmesi tıklanır:




Adım 4: Bu adımda, tahmin edici değişkenler ve sonuç değişkeni yukarıdaki pencerede görüldüğü gibi tanımlanır ve son sekmesi tıklanır:

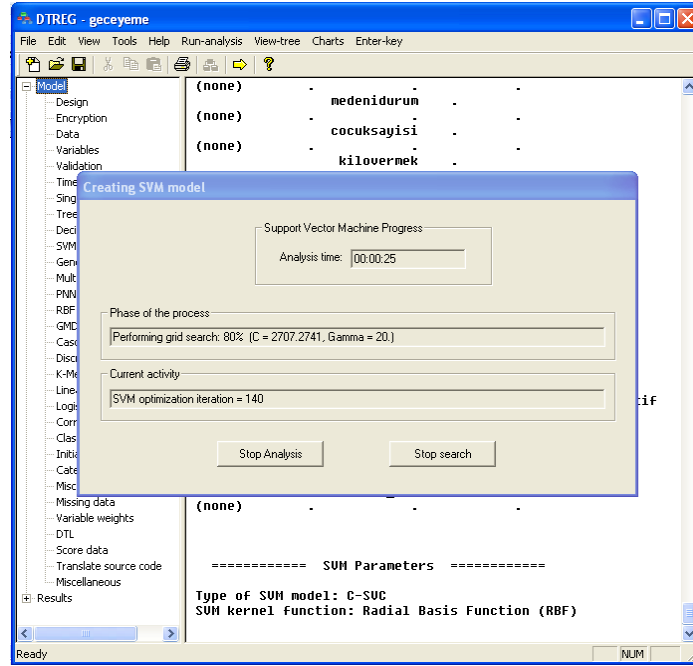


Yukarıda, ekranın sol tarafında görülen yöntemlerden SVM seçildiğinde aşağıdaki pencere görülür:



Bu pencerede ve pencerenin üstünde görülen sekmelerde uygun seçimler yapılır. Oluşturulan karar fonksiyonunun görülebilmesi ve yeni gelen bir verinin hangi sınıfa ait olduğunun belirlenebilmesi için *scoring* sekmesi kullanılır. Uygun seçimler yapıldıktan sonra *tamam* sekmesi tıklanır ve gelen ekrandan  ikonu veya *run-analysis* seçeneği

tıklanarak program çalıştırılır. Program çalıştırıldığında, analiz sürecinde ekranda aşağıdaki pencere görülür:



Bu adımlardan sonra, DTREG programı, istenilenleri bir output dosyasında görüntüler.

4. BULGULAR

Çalışmaya alınan bireylerden elde edilen, sayısal yapıdaki değişkenlerin tanımlayıcı istatistikleri Tablo 4.1’de verilmiştir. Tablo incelendiğinde, klinik olarak gece yeme sendromu tanısı alan ve almayan gruplarda yaş, eğitim yılı, kardeş sayısı, bel çevresi, kalça çevresi ve beden kitle indeksi (BKİ) ortalamaları arasında istatistiksel olarak anlamlı bir fark bulunmazken (p değerleri sırasıyla 0.702, 0.420, 0.686, 0.734, 0.186, 0.247), GYA puanı, BSQ puanı ve SCL ortalama puanı bakımından, gece yeme sendromu tanısı alan ve almayan gruplar arasında istatistiksel olarak anlamlı bir fark bulunmuştur (her bir $p < 0.001$).

Tablo 4.1: Sayısal Özelliklere Ait Tanımlayıcı İstatistikler

	Tanı Aldı (n=336)	Tanı Almadı (n=97)	
	Ort ± Std Sapma	Ort ± Std Sapma	P Değeri
Yaş	36.6 ± 10.6	38.02 ± 12.6	0.702
Eğitim Yılı	8.03 ± 3.3	8.25 ± 3.8	0.420
Kardeş Sayısı	3.9 ± 1.8	4.07 ± 2.07	0.686
Bel Çevresi	91.6 ± 12.9	90.9 ± 15.1	0.734
Kalça Çevresi	107.2 ± 11.1	105.7 ± 10.9	0.186
BKİ	27.6 ± 5.5	27.0 ± 5.1	0.247
GYA Puanı	27.6 ± 7.4	15.6 ± 5.6	<0.001
BSQ Puanı	93.6 ± 44.0	67.2 ± 33.3	<0.001
SCL Ortalama Puanı	1.8 ± 0.8	1.2 ± 0.7	<0.001

Çalışmaya alınan bireylerden elde edilen, kategorik yapıdaki değişkenlerin tanımlayıcı istatistikleri Tablo 4.2’de verilmiştir. Tablo incelendiğinde, cinsiyet, çocuk sayısı, kilo vermek, fiziksel hastalık ve antidepresan kullanımı ile gece yeme sendromu tanısı alıp almama durumu arasında istatistiksel olarak anlamlı bir ilişki bulunmazken (p değerleri sırasıyla 0.369, 0.849, 0.303, 0.169, 0.936), medeni durum, sigara kullanımı,

psikiyatrik tanı ile gece yeme sendromu tanısı alıp almama durumu arasında istatistiksel olarak anlamlı bir ilişki (p değerleri sırasıyla 0.036, 0.003, 0.003) bulunmuştur.

Tablo 4.2: Kategorik Özelliklere Ait Tanımlayıcı İstatistikler

	Kategori	Tanı Aldı		Tanı Almadı		P Değeri
		Sayı	%	Sayı	%	
Cinsiyet	Kadın	65	67.0	241	71.7	0.369
	Erkek	32	33.0	95	28.3	
Medeni Durum	Bekar	25	25.8	91	27.1	0.036
	Evli	59	60.8	226	67.3	
	Dul	13	13.4	19	5.7	
Çocuk Sayısı	0	33	34.0	130	38.8	0.849
	1	18	18.6	60	17.9	
	2	29	29.9	84	25.1	
	3	10	10.3	32	9.6	
	4 ve daha fazla	7	7.2	29	8.7	
Kilo Vermek	Evet	52	55.9	159	49.8	0.303
	Hayır	41	44.1	160	50.2	
Sigara	Kullanmıyor	51	52.6	231	68.8	0.003
	Kullanıyor	46	47.4	105	31.3	
Fiziksel Hastalık	Yok	55	57.3	217	65.0	0.169
	Var	41	42.7	117	35.0	
Psikiyatrik Tanı	Majör Depresyon	32	33.0	62	18.5	0.003
	Anksiyete Bozukluğu	29	29.9	157	46.7	
	Bipolar Afektif Bozukluk	14	14.4	43	12.8	
	Psikotik Bozukluk	7	7.2	40	11.9	
	Diğer	15	15.5	34	10.1	
Antidepresan	Almıyor	22	25.6	81	25.2	0.936
	Alıyor	64	74.4	241	74.8	

Gece yeme sendromu tanısı koymayı etkileyen faktörlerin lojistik regresyonla incelenmesi Tablo 4.3'te verilmiştir. Tablo incelendiğinde; yaş, cinsiyet, eğitim yılı, kardeş sayısı, medeni durum, çocuk sayısı, kilo vermek, kalça çevresi, BKİ, sigara, fiziksel hastalık, psikiyatrik tanı, antidepresan ve ort_scl değişkenlerinin tanı koymada etkili olmadığı bulunurken; bel çevresi, GYA_puan, BSQ_puan değişkenlerinin tanı koymada etkili değişkenler olduğu görülmüştür.

Tablo 4.3: Gece Yeme Sendromunu Etkileyen Faktörlerin Lojistik Regresyon ile İncelenmesi

Değişken	Kategori	OR	P Değeri	OR İçin %95 Güven Aralığı	
				Alt Sınır	Üst Sınır
Yaş		1.031	0.097	0.990	1.068
Cinsiyet	Kadın	0.573	0.224	0.233	1.406
	Erkek	Referans Kategori			
Eğitim Yılı		0.969	0.535	0.876	1.071
Kardeş Sayısı		0.986	0.877	0.820	1.185
Medeni Durum	Bekar	2.096	0.241	0.608	7.225
	Evli	Referans Kategori			
	Dul	1.197	0.760	0.377	3.804
Çocuk Sayısı	0	1.429	0.653	0.302	6.762
	1	1.398	0.658	0.317	6.166
	2	1.794	0.393	0.469	6.863
	3	2.424	0.259	0.520	11.288
	4 ve daha fazla	Referans Kategori			
Kilo Vermek	Evlet	Referans Kategori			
	Hayır	1.005	0.989	0.494	2.045
Bel Çevresi		0.929	0.002	0.887	0.974
Kalça Çevresi		1.049	0.113	0.989	1.112
BKİ		1.062	0.330	0.941	1.197
Sigara	Kullanmıyor	0.866	0.686	0.432	1.736
	Kullanıyor	Referans Kategori			
Fiziksel Hastalık	Yok	1.149	0.717	0.541	2.439
	Var	Referans Kategori			
Psikiyatrik Tanı	Majör Depresyon	1.431	0.614	0.356	5.756
	Anksiyete Bozukluğu	0.673	0.557	0.180	2.518
	Bipolar Afektif Bozukluk	1.593	0.516	0.391	6.483
	Psikotik Bozukluk	Referans Kategori			
	Diğer	0.927	0.917	0.221	3.890
Antidepresan	Almıyor	0.716	0.575	0.223	2.300
	Alıyor	Referans Kategori			
GYA_puan		1.314	p<0.001	1.231	1.402
BSQ_puan		1.012	0.020	1.002	1.022
ort_scl		0.837	0.495	0.501	1.396

Gece yeme sendromu tanısı koymada, lojistik regresyon modelinin başarısı veya performans ölçüleri Tablo 4.4'te verilmiştir. Tablo incelendiğinde, hem eğitim seti hem de test seti için modelin genel doğru sınıflandırma başarısı, seçicilik ve negatif tahmini değeri yüksek, duyarlılık ve pozitif tahmini değeri ise nispeten daha düşük bulunmuştur. Ayrıca ROC eğrisinin altında kalan alan eğitim seti için 0.927467, test seti için 0.889359 olarak elde edilmiştir. Bu sonuçlara göre, lojistik regresyon modelinin tanı

koyma başarısının yüksek olduğu söylenebilir. Tabloda görülen *F-Ölçümü değeri*, pozitif tahmini değer (PTD) ile duyarlılığın harmonik ortalamasıdır ve aşağıdaki gibi hesaplanmaktadır:

$$F - \text{Ölçümü} = 2 \cdot \frac{PTD \cdot \text{Duyarlılık}}{PTD + \text{Duyarlılık}}$$

Tablo 4. 4: On Yedi Değişken İçeren Lojistik Regresyon Modelinin Sınıflama Başarıları

Klinik Tanı (Gerçek Durum)	Model Tahmini			
	Eğitim (Training)		Test (Validation)	
	Var	Yok	Var	Yok
Var	67	30	61	36
Yok	14	322	22	314
Doğruluk	% 89.84		% 86.61	
Duyarlılık	% 69.07		% 62.89	
Seçicilik	% 95.83		% 93.45	
Pozitif tahmini değer (PTD)	% 82.72		% 73.49	
Negatif tahmini değer (NTD)	% 91.48		% 89.71	
F-Ölçümü (F-Measure)	0.7528		0.6778	
ROC eğrisinin altında kalan alan	0.927467		0.889359	

Eğitim setinde tanı amacıyla lojistik regresyon modeline 17 değişken alındığı zaman, modele ait “yükselme ve kazanç” değerleri Tablo 4.5’ teki gibi elde edilmiştir. “Yükselme ve kazanç” tabloları, bir tahmin modelinin değerini ölçmek için yararlı bir araçtır. Bu tablolarda temel olarak, bazı hedef kategorilerde tahmin edilen hedef değerleri, katıksızlığa (purity) göre azalan bir şekilde sıralanır ve daha sonra her bir *bindeki* (bölümdeki) belirlenen hedef kategoriye ait deneğin oranı, genel oranla karşılaştırılır. “Yükseltme ve kazanç” değerleri deneklerin en iyi % 10, % 20, vb. kısımları çıkarıldığında modelin ne kadar iyileşme sağlayacağını gösterir. Tablodaki sayıların çoğu, hedef kategoriye ait deneklerin genel yüzdesiyle ilişkilidir. Tablonun altında verilen bu değere bakıldığında, “GYS_tanı=tanı aldı grubuna ait deneklerin yüzdesinin % 22.40 olduğu görülmektedir. Ancak bu değer, eğitim setinden

hesaplanmıştır. Eğitim ve test seti için hesaplanan yüzdeler az da olsa bir farklılık gösterebilir.

Tablo 4.5: On Yedi Değişken Yardımıyla Kurulan Lojistik Regresyon Modelinin Yükselme ve Kazanç Değerleri

Bin İndeksi	Binlerin Sınıf Yüzdeleri	Birikimli Popülasyon Yüzdeleri	Birikimli Sınıf Yüzdeleri	Birikimli Kazanç	Popülasyon Yüzdeleri	Sınıf Yüzdeleri	Yükselme (Lift)
1	93.18	10.16	42.27	4.16	10.16	42.27	4.16
2	65.91	20.32	72.16	3.55	10.16	29.90	2.94
3	22.73	30.48	82.47	2.71	10.16	10.31	1.01
4	20.45	40.65	91.75	2.26	10.16	9.28	0.91
5	9.09	50.81	95.88	1.89	10.16	4.12	0.41
6	0.00	60.97	95.88	1.57	10.16	0.00	0.00
7	6.82	71.13	98.97	1.39	10.16	3.09	0.30
8	2.27	81.29	100.00	1.23	10.16	1.03	0.10
9	0.00	91.45	100.00	1.09	10.16	0.00	0.00
10	0.00	100.00	100.00	1.00	8.55	0.00	0.00

Ortalama kazanç = 2.085

GYS_tanı = tanı aldı sınıfındaki bireylerin yüzdesi = % 22.40

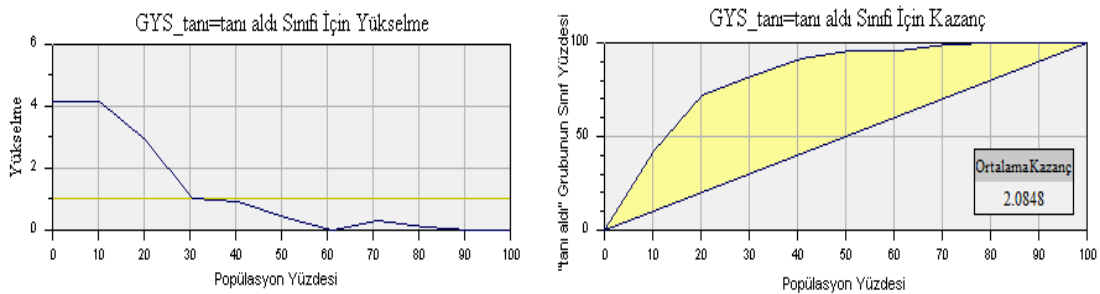
Eğitim setine ait bu tablodaki sonuçlar değerlendirildiğinde;

- *Bin indeksi:* Binler 1 ile belirlenen bir üst sınır değeri arasında numaralandırılır. Burada üst sınır 10 olarak belirlenmiştir. İlk *bin*, belirlenen hedef kategori (tanı aldı) için tahmin edilen en yüksek olasılığa sahip veri satırlarını temsil etmektedir.
- *Binlerin sınıf yüzdeleri:* Bu değer, hedef değişkenin belirlenen kategorisine ait deneklerin *bin*deki yüzdesidir. Bu uygulamada hedef değişken GYS_tanı değişkenidir ve aşağıdaki yükseltme ve kazanç tablosu GYS_tanı değişkeninin “tanı aldı” kategorisi için oluşturulmuştur.
- *Birikimli popülasyon yüzdeleri:* Bu değer, bulunduğu satırdaki *bine* düşen herhangi bir kategoriye ait tüm bireylerin birikimli yüzdesidir.
- *Birikimli sınıf yüzdeleri:* Bu değer, belirlenen kategoriye ait bireylerin birikimli yüzdesidir.
- *Birikimli kazanç:* Bu değer, birikimli sınıf yüzdesinin, birikimli popülasyon yüzdesine bölünmesiyle elde edilen orandır. Örneğin, tabloda 2. *bin* için birikimli kazanç değeri, $72.16/20.32 = 3.55$ olarak bulunmuştur.

- *Popülasyon Yüzdeleri:* Bu değer, bulunduğu satırdaki *bine* düşen tüm bireylerin yüzdesidir ve yaklaşık olarak “100/*bin* sayısı” olarak değer alır.
- *Sınıf yüzdeleri:* Bu değer, bulunduğu satırdaki *bine* düşen, belirlenen kategoriye ait bireylerin yüzdesidir. Örneğin, tabloda “tanı aldı” kategorisine ait tüm bireylerin %42.27’si ilk *binde* yer almaktadır.
- *Yükselme:* Bu değer, *bindeki* belirlenen kategoriye ait bireylerin yüzdesinin (sınıf yüzdesi), *bindeki* bireylerin toplam yüzdesine (popülasyon yüzdesi) bölünmesiyle elde edilir. Örneğin, tablonun ilk satırı için, yükselme değeri, $42.27/10.16 = 4.16$ olarak bulunmuştur.

Şekil 4.1’de, tablodaki değerlere ilişkin yükselme ve kazanç grafikleri görülmektedir. Birikimli kazanç ve yükselme grafikleri, tahmin modeli kullanmanın avantajının grafiksel olarak gösterimini sunmaktadır. Grafiklerdeki düz çizgiler tahmin modeli kullanılmadan elde edilen sınıf tahminlerini, eğri çizgi ise tahmin modeli kullanıldığında elde edilen sınıf tahminlerini göstermektedir.

Yükselme grafiği, tahmin modeli kullanıldığında elde edilecek başarının bir göstergesidir ve yükselme grafiğindeki düz çizgi (baseline) ve eğri çizgi (yükselme eğrisi) arasındaki alan ne kadar büyükse, modelin tahmin başarısı o kadar iyi olacaktır. Bu grafik, model kullanıldığında elde edilen tahmin başarısının, kullanılmadığında elde edilen tahmin başarısına göre kaç kat değişim göstereceğini ifade etmektedir. Kazanç tablosundaki çizgiler arasında kalan boyalı alan ise modelin iyileşmesini (kazanç) göstermektedir.



Şekil 4. 1: On Yedi Değişken Yardımıyla Kurulan Lojistik Regresyon Modelinin Yükselme ve Kazanç Grafikleri

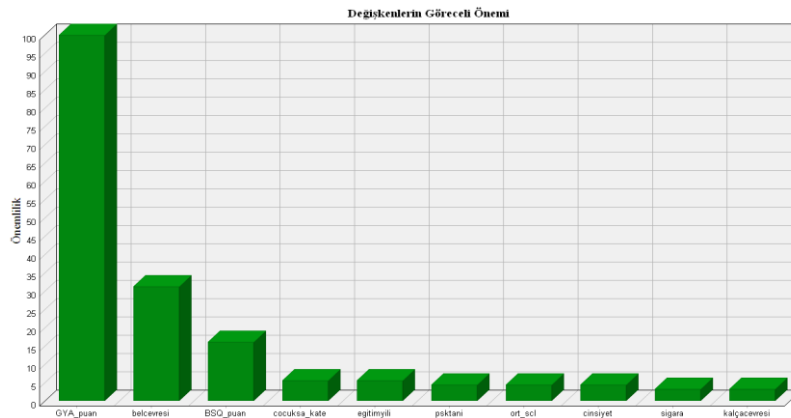
Modele alınan ve tanı koymada etkili bulunan değişkenlerin önemlilik dereceleri incelendiğinde, tanı koymada en etkili değişken GYA_puan olarak bulunmuş, daha sonra, sırasıyla bel çevresi ve BSQ_puan değişkenleri önemli değişkenler olarak elde

edilmiştir. Diğer değişkenlerin ise anlamlı olmasa da tanı koymada bir miktar etkileri olduğu söylenebilir. Tablo 4.6’da bu değişkenler ve önem dereceleri verilmiştir. Tablo değerlendirildiğinde, lojistik regresyon modeline göre tanı koyma başarısı anlamlı bulunan değişkenlerin büyüklük açısından ilk 3 sırada önem derecesine sahip olduğu söylenebilir. BSQ_puan’dan sonra gelen değişkenlerin önem derecelerinde keskin bir azalma gözlenmektedir. Tabloda sadece önem dereceleri sıfırdan büyük değişkenler yer almaktadır.

Tablo 4.6: Lojistik Regresyon Modeline Alınan Değişkenlerin Önem Dereceleri

Değişken	Önem Derecesi
GYA_puan	100.000
Bel Çevresi	31.395
BSQ_puan	16.274
Çocuk Sayısı	5.814
Eğitim Yılı	5.814
Psikiyatrik Tanı	4.651
ort_scl	4.651
Cinsiyet	4.651
Sigara	3.488
Kalça Çevresi	3.488
Medeni Durum	2.326
BKİ	2.326
Kardeş Sayısı	2.326
Yaş	2.326
Antidepresan	1.163
Fiziksel Hastalık	1.163

Şekil 4.2’de değişkenlerin önem dereceleri grafiksel olarak görünmektedir.



Şekil 4.2: Değişkenlerin Önem Dereceleri

Lojistik regresyon yöntemi ile elde edilen modelde, katkısı anlamlı bulunan üç değişkenle analiz tekrarlandığında Tablo 4.7’de verilen sınıflama başarıları elde edilmiştir. Tablo incelendiğinde, hem eğitim seti hem de test seti için doğruluk, seçicilik ve negatif tahmini değer yüksek, duyarlılık ve pozitif tahmini değer ise nispeten daha düşük bulunmuştur. Ayrıca ROC eğrisinin altında kalan alan eğitim seti için 0.912126, test seti için 0.906726 olarak elde edilmiştir. Bulunan bu değerler ve ROC eğrisinin altında kalan alan birlikte incelendiğinde, modelin tanı koyma başarısının yüksek olduğu söylenebilir.

Tablo 4.7: Üç Değişken İçeren Lojistik Regresyon Modelinin Sınıflama Başarıları

Klinik Tanı (Gerçek Durum)	Model Tahmini			
	Eğitim (Training)		Test (Validation)	
	Var	Yok	Var	Yok
Var	64	33	62	35
Yok	16	320	17	319
Doğruluk	% 88.68		% 87.99	
Duyarlılık	% 65.98		% 63.92	
Seçicilik	% 95.24		% 94.94	
Pozitif tahmini değer (PTD)	% 80.00		% 78.48	
Negatif tahmini değer (NTD)	% 90.65		% 90.11	
F-Measure	0.7232		0.7045	
ROC eğrisinin altında kalan alan	0.912126		0.906726	

Tablo 4.8’de lojistik regresyon yöntemi ile 3 değişken kullanıldığında eğitim setinden elde edilen modele ait “yükselme ve kazanç” değerleri görülmektedir.

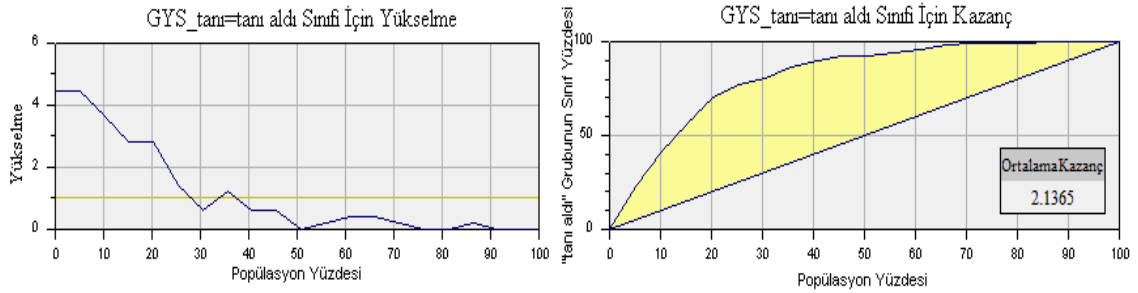
Tablo 4.8: Üç Değişken Yardımıyla Kurulan Lojistik Regresyon Modelinin Yükselme ve Kazanç Değerleri

Bin Index	Binlerin Sınıf Yüzdeleri	Birikimli Popülasyon Yüzdeleri	Birikimli Sınıf Yüzdeleri	Birikimli Kazanç	Popülasyon Yüzdeleri	Sınıf Yüzdeleri	Lift
1	90.91	10.16	41.24	4.06	10.16	41.24	4.06
2	63.64	20.32	70.10	3.45	10.16	28.87	2.84
3	22.73	30.48	80.41	2.64	10.16	10.31	1.01
4	20.45	40.65	89.69	2.21	10.16	9.28	0.91
5	6.82	50.81	92.78	1.83	10.16	3.09	0.30
6	6.82	60.97	95.88	1.57	10.16	3.09	0.30
7	6.82	71.13	98.97	1.39	10.16	3.09	0.30
8	0.00	81.29	98.97	1.22	10.16	0.00	0.00
9	2.27	91.45	100.00	1.09	10.16	1.03	0.10
10	0.00	100.00	100.00	1.00	8.55	0.00	0.00

Ortalama kazanç = 2.045

GYS_tanı = tanı aldı sınıfındaki bireylerin yüzdesi = % 22.40

Şekil 4.3'te, tablodaki değerlere ilişkin yükselme ve kazanç grafikleri görülmektedir.



Şekil 4.3: On Yedi Değişken Yardımıyla Kurulan Lojistik Regresyon Modelinin Yükselme ve Kazanç Grafikleri

a) On yedi belirleyici nitelik yardımıyla DVM yönteminden elde edilen sonuçlar

On yedi belirleyici nitelik (değişken) yardımıyla DVM yöntemi uygulandığında, parametre tahmini için yapılan arama sırasında değerlendirilen noktalarının sayısı 144, arama ile optimize edilen en iyi duyarlılık ve seçicilik değeri 0.708927 ve modelde kullanılan destek vektörlerin sayısı 207 olarak bulunmuştur. Durdurma kriteri ϵ (stopping criterion), 0.001 olarak belirlenmiş ve tahmin edilen parametre değerleri; hata maliyet değeri, $C = 0.10846446$ ve RTF parametresi olan gamma değeri, $\gamma = 0.02714418$ olarak bulunmuştur (Tablo 4.9).

Tablo 4.9: On Yedi Belirleyici Nitelik Yardımıyla DVM Yönteminden Tahmin Edilen Optimum Parametre Değerleri

		Tahmin Edilen Parametreler		
Değerlendirilen Nokta Sayısı	Optimize Edilen En Büyük Duyarlılık ve Seçicilik Değeri	ϵ	C	γ
144	0.708927	0.001	0.10846446	0.02714418

Tablo 4.10’da modelin tahmin ettiği çeşitli olasılık değerlerinin, kesim noktası olarak alındığı durumlardaki model performansının göstergeleri yer almaktadır. Bu tablo incelendiğinde aşağıdaki sonuçlar elde edilmiştir:

- *Vakaların oranı:* Verilen bir olasılığın kesim değeri olarak kabul edildiği durumda hedef kategoriye atanan (bu veride hedef kategori "tanı aldı" grubudur) vakaların oranını gösterir. Örneğin tablodaki eğitim setinde 0.10 olasılık değeri kesim değeri olarak kabul edilirse vakaların %50.61’i seçilen hedef kategoriye atanır. Olasılık kesim değeri %80’ e çıkarıldığı zaman vakaların %7.14’ü hedef kategoriye atanır geriye kalan denekler diğer kategoriye atanır. Eğer kesim değeri %56 alınırsa vakaların %15.71’i hedef kategoriye atanacaktır.
- *Hata:* Belirli bir kesim değeri için hatalı sınıflandırılan bireylerin oranıdır. Bu değer küçük olması istenir çünkü bir kesim değeri için toplumdaki vaka / kontrol dağılımına uygun bir vaka oranı elde edildiği zaman ve hata oranının da en küçük olduğu durumda o kesim değerine uygun kesim değeri adı verilir.
- *Katışıklık oranı (Impurity):* Hedef kategoriye ait bireyler olarak sınıflandırılan ancak gerçek kategorisi ile seçilen kategorisi farklı olan deneklerin oranıdır. Bir başka ifadeyle gerçekte diğer kategori içerisinde olup da hedef kategorideymiş gibi yanlış sınıflandırılan vakaların oranıdır. Örneğin tablodaki eğitim setinde, olasılık kesim değeri 0.10 alınırsa, vakaların %59.39’u gerçekte tanı almayan grupta olmasına rağmen, tanı alan gruptaymış gibi sınıflandırılır. Olasılık kesim değeri büyüdükçe katışıklık değeri küçülür.

- *Kayıp değer:* Gerçek sınıfı hedef kategori iken diğer kategoriye yanlış sınıflandırılan bireylerin oranıdır. Örneğin tablodaki eğitim setinde, 0.56 olasılık değeri kesim değeri olarak alınırsa gerçekte tanı almış olan bireylerin %39.18' i tanı almadı olarak yanlış sınıflandırılmıştır.

Bu açıklamalara göre katışıklık oranı ve kayıp değer seçilen hedef kategoriye göre yanlış pozitiflik veya yanlış negatiflik oranını ifade eder. Birinin yanlış pozitiflik oranını göstermesi durumunda, diğeri yanlış negatiflik oranını göstermektedir.

Tablonun altında verilen çıktılar incelendiğinde aşağıdaki sonuçlar elde edilmiştir:

- *ROC eğrisi altında kalan alan:* Bu alan modelin tahmin olasılıkları yardımıyla çizilen ROC eğrisi altında kalan alandır ve bu değer eğitim seti için 0.915378 ve test seti için 0.516456 olarak bulunmuştur.
- *Yanlış sınıflamayı minimize eden kesim noktası:* Bu değer, veri setinin tümü için toplam yanlış sınıflama hatasını minimize edecek olasılık kesim noktasıdır ve bu değer eğitim seti için 0.564818 ve test seti için 0.516456 olarak bulunmuştur.
- *Ağırlıklandırılmış yanlış sınıflamayı minimize eden kesim noktası:* Bu değer, ağırlıklandırılmış yanlış sınıflama hatasını minimize edecek olasılık kesim noktasıdır ve bu değer eğitim seti için 0.564818 ve test seti için 0.861392 olarak bulunmuştur.
- *Yanlış sınıflamayı dengeleyen kesim noktası:* Bu değer, her bir hedef kategori için yanlış sınıflanan bireylerin sayılarının oranlarını yaklaşık olarak eşitleyecek olasılık kesim noktasıdır ve bu değer eğitim seti için 0.212006 ve test seti için 0.239606 olarak bulunmuştur.

Bu ölçüler değerlendirildiğinde en uygun olasılık kesim değeri eğitim ve test verileri için sırasıyla 0.56 ve 0.52 olarak belirlenmiş ve bu değer üstünde olasılık değerine sahip olan bireyler tanı alan grupta kabul edilmiştir. Bu sınıflama sonucunda elde edilen model başarı ölçütleri Tablo 4.12'de görülmektedir.

Tablo 4.10: On Yedi Belirleyici Nitelik Kullanılarak DVM Yöntemi ile Elde Edilmiş Kesim Noktaları

Olasılık Kesim Değeri	Eğitim Seti				Test Seti			
	Vakaların Oranı	Hata	Katışıklık (Impurity)	Kayıp (Loss)	Vakaların Oranı	Hata	Katışıklık (Impurity)	Kayıp (Loss)
0.00	1.0000	0.7760	0.7760	0.0000	1.0000	0.7760	0.7760	0.0000
0.02	0.8761	0.6521	0.7443	0.0000	0.8960	0.6719	0.7500	0.0000
0.04	0.7333	0.5139	0.6977	0.0103	0.7915	0.5721	0.7199	0.0103
0.06	0.6337	0.4235	0.6574	0.0309	0.6617	0.4562	0.6754	0.0412
0.08	0.5642	0.3679	0.6275	0.0619	0.5851	0.4026	0.6526	0.0928
0.10	0.5061	0.3190	0.5939	0.0825	0.5407	0.3675	0.6327	0.1134
0.12	0.4783	0.2959	0.5751	0.0928	0.5062	0.3330	0.6076	0.1134
0.14	0.4437	0.2659	0.5472	0.1031	0.4680	0.3133	0.5954	0.1546
0.16	0.3598	0.1958	0.4608	0.1340	0.4293	0.2792	0.5642	0.1649
0.18	0.3424	0.1819	0.4384	0.1417	0.4126	0.2733	0.5597	0.189
0.20	0.3206	0.1705	0.4165	0.1649	0.3897	0.2581	0.5437	0.2062
0.22	0.3035	0.1534	0.3836	0.1649	0.3608	0.2328	0.5121	0.2143
0.24	0.2862	0.1407	0.3544	0.1753	0.3417	0.2146	0.4863	0.2165
0.26	0.2741	0.1301	0.3287	0.1787	0.3175	0.1951	0.4544	0.2268
0.28	0.2528	0.1258	0.3057	0.2165	0.2991	0.1790	0.4247	0.2320
0.30	0.2364	0.1186	0.2770	0.2371	0.2793	0.1757	0.4135	0.2688
0.32	0.2300	0.1168	0.2670	0.2474	0.2667	0.1674	0.3937	0.2784
0.34	0.2256	0.1185	0.2661	0.2608	0.2544	0.1644	0.3827	0.2990
0.36	0.2186	0.1193	0.2605	0.2784	0.2437	0.1651	0.3791	0.3245
0.38	0.2078	0.1225	0.2556	0.3096	0.2322	0.1559	0.3534	0.3299
0.40	0.2033	0.1225	0.2502	0.3196	0.2170	0.1594	0.3512	0.3715
0.42	0.1943	0.1221	0.2377	0.3386	0.2064	0.1655	0.3581	0.4087
0.44	0.1897	0.1181	0.2207	0.3402	0.1986	0.1639	0.3488	0.4227
0.46	0.1823	0.1156	0.2027	0.3512	0.1870	0.1570	0.3209	0.4330
0.48	0.1803	0.1176	0.2049	0.3601	0.1815	0.1534	0.3054	0.4372
0.50	0.1778	0.1155	0.1948	0.3608	0.1732	0.1524	0.2933	0.4536
0.52	0.1758	0.1175	0.1970	0.3698	0.1624	0.1494	0.2702	0.4709
0.54	0.1636	0.1112	0.1553	0.3831	0.1550	0.1568	0.2832	0.5042
0.56	0.1571	0.1086	0.1327	0.3918	0.1441	0.1511	0.2469	0.5155
0.58	0.1450	0.1104	0.1083	0.4227	0.1394	0.1539	0.2484	0.5322
0.60	0.1407	0.1107	0.0972	0.4330	0.1336	0.1551	0.2421	0.5481
0.62	0.1303	0.1168	0.0886	0.4700	0.1199	0.1638	0.2487	0.5979
0.64	0.1282	0.1166	0.0810	0.4742	0.1162	0.1633	0.2386	0.6052
0.66	0.1261	0.1164	0.0733	0.4784	0.1081	0.1613	0.2098	0.6186

Tablo 4.10 (Devam): On Yedi Belirleyici Nitelik Kullanılarak DVM Yöntemi ile Elde Edilmiş Kesim Noktaları

Olasılık Kesim Değeri	Eğitim Seti				Test Seti			
	Vakaların Oranı	Hata	Katışıklık (Impurity)	Kayıp (Loss)	Vakaların Oranı	Hata	Katışıklık (Impurity)	Kayıp (Loss)
0.68	0.1231	0.1194	0.0751	0.4919	0.1003	0.1607	0.1843	0.6349
0.70	0.1105	0.1274	0.0627	0.5377	0.0925	0.1595	0.1510	0.6495
0.72	0.1050	0.1329	0.0660	0.5621	0.0873	0.1552	0.1059	0.6517
0.74	0.0954	0.1425	0.0726	0.6051	0.0811	0.1614	0.1139	0.6790
0.76	0.0868	0.1511	0.0798	0.6434	0.0731	0.1678	0.1151	0.7113
0.78	0.0841	0.1538	0.0824	0.6557	0.0711	0.1668	0.0975	0.7136
0.80	0.0714	0.1614	0.0616	0.7010	0.0673	0.1706	0.103	0.7306
0.82	0.0676	0.1577	0.0097	0.7010	0.0596	0.1782	0.1162	0.7647
0.84	0.0653	0.1587	0.0000	0.7084	0.0503	0.1876	0.1379	0.8065
0.86	0.0620	0.1620	0.0000	0.7231	0.0414	0.1918	0.1115	0.8357
0.88	0.0564	0.1676	0.0000	0.7481	0.0369	0.1918	0.0626	0.8457
0.90	0.0467	0.1773	0.0000	0.7916	0.0307	0.1947	0.0219	0.8660
0.92	0.0308	0.1932	0.0000	0.8625	0.0269	0.1971	0.0000	0.8799
0.94	0.0238	0.2002	0.0000	0.8937	0.0170	0.2070	0.0000	0.9239
0.96	0.0174	0.2066	0.0000	0.9223	0.0078	0.2162	0.0000	0.9653
0.98	0.0065	0.2175	0.0000	0.9710	0.0018	0.2222	0.0000	0.9921
1.00	0.0000	0.2240	0.0000	1.0000	0.0000	0.2240	0.0000	1.0000

Eğitim seti için ROC eğrisi altında kalan alan = 0.915378

Eğitim seti için yanlış sınıflamayı minimize eden kesim noktası = 0.564818

Eğitim seti için ağırlıklandırılmış yanlış sınıflamayı minimize eden kesim noktası = 0.564818

Eğitim seti için yanlış sınıflamayı dengeleyen kesim noktası = 0.212006

Test seti için ROC eğrisi altında kalan alan = 0.861392

Test seti için yanlış sınıflamayı minimize eden kesim noktası = 0.516456

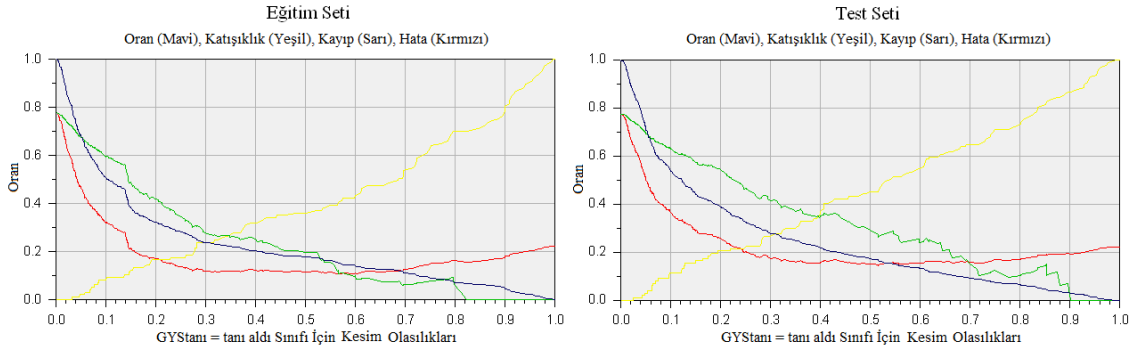
Test seti için ağırlıklandırılmış yanlış sınıflamayı minimize eden kesim noktası = 0.516456

Test seti için yanlış sınıflamayı dengeleyen kesim noktası = 0.239606

Şekil 4.4'te, yukarıda elde edilen eğitim ve test setlerine ait kesim noktası değerleri için oluşturulan olasılık kesim noktası grafiği görülmektedir. Bu grafik, yalnızca 2 kategorili hedef değişkene sahip sınıflama problemleri için çizilebilmektedir.

Olasılık kesim noktası grafiği, değişen olasılık kesim noktası değerlerinin, seçilen hedef kategoriye sınıflanan bireylerin oranını nasıl etkileyeceğini göstermektedir. Grafiğin yatay ekseninde 0.00 ile 1.0 arasında değişen kesim noktası olasılıkları, dikey ekseninde ise oranlar görülmektedir. Grafikte yer alan renkli çizgiler tabloda açıklanmış olan;

“vakaların oranı”, “hata”, “katışıklık” ve “kayıp” sütunlarına ait değerlerin grafiksel gösterimidir ve bu renkli çizgiler yardımıyla oluşturulan olasılık kesim noktası grafiği, katışıklık ve kayıp değer arasındaki ödünleşimin görsel bir ifadesini sunmaktadır.



Şekil 4 4: On Yedi Belirleyici Nitelik Kullanılarak Elde Edilmiş Eğitim ve Test Verileri için Kesim Noktası Grafikleri

Tablo 4.11’de, hedef kategorinin “tanı aldı” olarak seçilmesi durumunda elde edilen olasılık düzenlemesi yer almaktadır. Olasılık düzenleme tablosu, hedef kategorinin tahmin edilen olasılık değerinin dağılımını gösterir ve tahmin edilen olasılıkların doğruluk ölçümleri için ortalamalar üretir. Tabloda tahmin edilen olasılık değerlerinin her 0.1’lik aralığına ait 10 satır bulunmaktadır (0.00-0.10, 0.10-0.20, vb). Eğitim veya test setindeki her bir birey bu aralıklara hedef kategorinin seçimine göre atanır. Hedef değişkenin “tanı aldı” olarak seçildiği bu uygulama için, 0.00-0.10 olasılık aralığında toplam bireylerin % 49.19’i (213 birey) “tanı aldı” grubuna sınıflandırılmıştır.

Düzenleme tablosundaki *Tahmin Edilen Değer* sütunu, aralıktaki tahmin edilen olasılığa sahip bütün veri satırları için tahmin edilen ortalama olasılık değeridir. Bu değer genellikle, bulunduğu olasılık aralığının orta noktasına yakın bir değerdir. *Gerçek Değer* sütunu ise, belirlenen hedef kategoriye ait satırların gerçek oranıdır.

Eğer kurulan model doğruysa, tahmin edilen olasılık değeri ile gerçek değer birbirine eşit olacaktır. Olasılık düzenleme tablosu, bu karşılaştırmanın yapılmasına olanak sağlamaktadır. Örneğin, tahmin edilen olasılık değerleri aralığının 0.10-0.20 olduğu 2. satıra bakılırsa, bu aralıktaki bütün veri setinin % 18.71’ini oluşturan 81kişinin olduğu görülmektedir. Bu bireyler için tahmin edilen ortalama olasılık değeri, aralığın hemen hemen orta noktası olan 0.1455 değeridir. Eğer model çok iyiye, bu bireylere ait gerçek olasılık değerinin 0.1455 olması beklenir. Ancak *Gerçek Değer* sütununa bakıldığında,

bu deęerin 0.0988 olduęu grlmektedir. Buradan, bu aralık iin tahmin edilen olasılıęın, gerek durumu overestimate eęiliminde olduęu sonucuna varılabilir. Bu bilgiler kullanılarak, tahmin edilen olasılıkları gerek olasılıkların daha doęru tahminlerine haritalayan bir olasılık dzenleme dzeltme fonksiyonu geliřtirmek mmkn olacaktır.

Tablo 4.11: On Yedi Belirleyici Nitelik Kullanıldıęında DVM Yntemi ile Elde Edilen Olasılık Dzenlemesi

Tahmin Edilen Olasılık Aralıęı	Eęitim Seti				Test Seti			
	Sayı	%	Tahmin Edilen Deęer	Gerek Deęer	Sayı	%	Tahmin Edilen Deęer	Gerek Deęer
0.00-0.10	213	49.19	0.0420	0.0376	198	45.73	0.0439	0.0556
0.10-0.20	81	18.71	0.1455	0.0988	66	15.24	0.1423	0.1364
0.20-0.30	36	8.31	0.2519	0.1944	48	11.09	0.2467	0.1250
0.30-0.40	14	3.23	0.3507	0.5714	27	6.24	0.3535	0.3704
0.40-0.50	12	2.77	0.4382	0.3333	19	4.39	0.4501	0.4211
0.50-0.60	17	3.93	0.5528	0.4118	19	4.39	0.5419	0.5263
0.60-0.70	13	3.00	0.6570	0.7692	17	3.93	0.6451	0.5882
0.70-0.80	17	3.93	0.7558	0.9412	11	2.54	0.7400	0.6364
0.80-0.90	10	2.31	0.8654	0.8000	16	3.70	0.8451	0.8750
0.90-1.00	21	4.85	0.9448	10.000	14	3.23	0.9470	0.9286

Eęitim seti iin aęırlıklandırılmıř ortalama olasılık hatası = 0.046151

Eęitim seti iin aęırlıklandırılmıř ortalama kare olasılık hatası = 0.073006

Test seti iin aęırlıklandırılmıř ortalama olasılık hatası = 0.029309

Test seti iin aęırlıklandırılmıř ortalama kare olasılık hatası = 0.047111

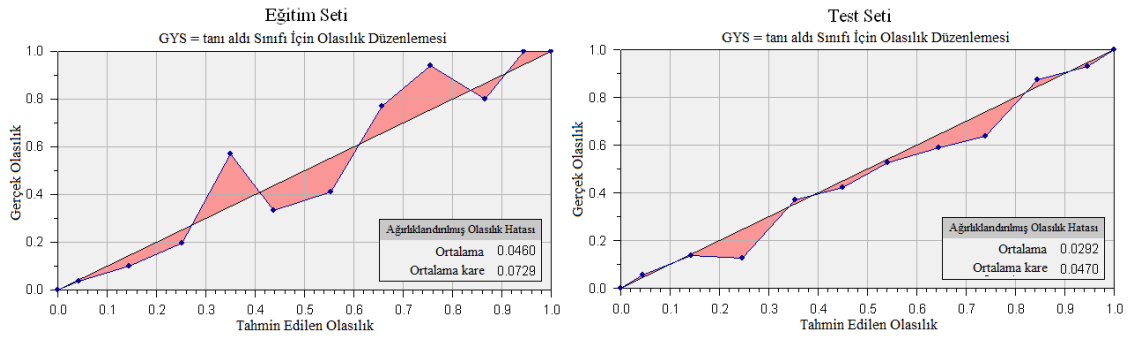
Tablonun altında yer alan bilgiler ise řyle aıklanabilir:

- *Aęırlıklandırılmıř ortalama olasılık hatası:* Bu deęer, her bir aralıęa dřen satır sayıları ile aęırlıklandırılmıř, tahmin edilen olasılık ve gerek durum arasındaki ortalama hatadır. Yine olasılık aralıęı 0.10-0.20 olan 2. satır ele alınırsa, burada hata=0.1455-0.0988=0,0467 olarak elde edilir ve tm bireylerin %18.71'i bu aralıęta yer almaktadır. Bu durumda bu hata, toplam ortalama hatanın %18.71'ine katkıda bulunmaktadır.
- *Aęırlıklandırılmıř ortalama kare olasılık hatası:* Bu deęer aęırlıklandırılmıř ortalama olasılık hatası ile aynı řekilde hesaplanmaktadır. Ancak, aęırlıęı ile

çarpılmadan önce, hatanın karesi alınır ve toplam hataya eklenir. Toplam hata kareler hesaplandıktan sonra, bu değerin karekökü bulunur. Verilen olasılık düzenleme tablosunun altındaki değer, bu karekök değeridir.

Sınıflama başarısından çok olasılık tahminlerinin önemli olduğu problemlerde, ortalama hata ve ortalama kare hata değerleri modelin kalitesini çok iyi ortaya koyan genel göstergelerdir.

Şekil 4.5'te, elde edilen değerler için olasılık düzenleme grafiği görülmektedir. Bu grafik hedef değişken için tahmin edilen olasılığın nasıl bir dağılıma sahip olduğunu göstermektedir ve tahmin edilen olasılıkların doğruluğunu ölçmek için bir araçtır. Olasılık düzenleme grafiği yalnızca, hedef değişkenin iki sınıfa sahip olduğu sınıflama problemleri için oluşturulabilir.



Şekil 4.5: On Yedi Belirleyici Nitelik Kullanılarak Elde Edilmiş Eğitim ve Test Verileri için Olasılık Düzenleme Grafikleri

Grafikteki yatay eksenle gözlemlerin tahmin edilen olasılıkları, dikey eksenle ise olayın meydana gelme sıklığına dayanan gerçek olasılık değerleri yer almaktadır. Örneğin Şekil 4.5'te verilen grafikte, 0.4 ve 0.5 arasındaki ortalama tahmin edilen olasılık değeri 0.45 civarında iken, gerçek olasılık değeri 0.90 civarındadır ve bu da hatanın bir göstergesidir. Grafikteki kırmızı taralı alan, tahmin edilen olasılık ve gerçek olasılık değerleri arasındaki farkı gösteren hata değeridir.

Belirlenen kesim noktalarına göre, gece yeme sendromu tanısı alan (Var=Tanı alan) ve almayan (Yok=tanı almayan) grupları doğru sınıflayabilmek için modelin sınıflama başarıları Tablo 4.11'de verilmiştir. Tablo incelendiğinde, doğruluk, seçicilik ve negatif tahmini değerlerin hem eğitim, hem de test verilerinde yüksek, duyarlılık ve pozitif tahmini değerlerin ise her iki grupta da nispeten daha düşük olduğu görülmüştür. Ayrıca ROC eğrisinin altında kalan alan eğitim seti için 0.915378, test seti için 0.861392 olarak

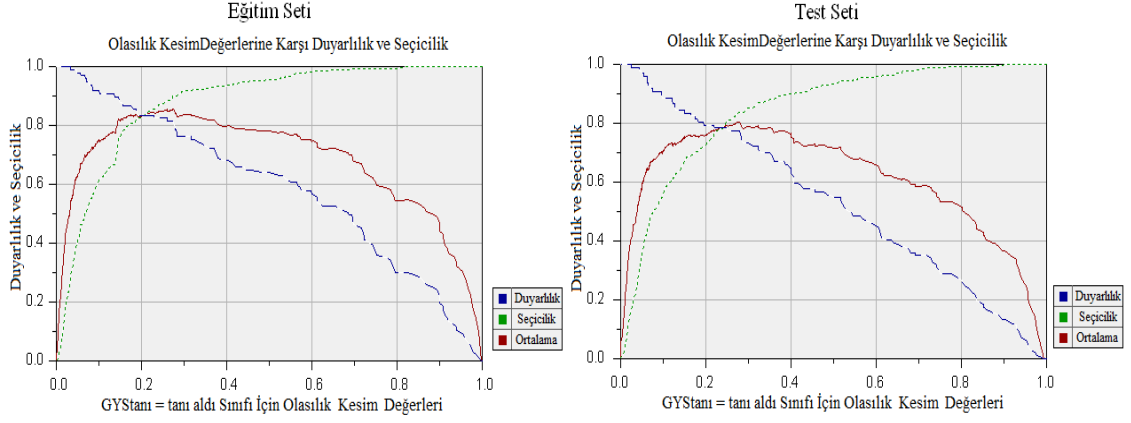
bulunmuştur. Bulunan bu değerler ve ROC eğrisinin altında kalan alanlar birlikte incelendiğinde, modelin tanı koyma başarısının oldukça iyi derecede olduğu görülmüştür.

Tablo 4.12: On Yedi Belirleyici Nitelik Kullanılarak DVM Yöntemi ile Elde Edilmiş Eğitim ve Test Verilerinde Modelin Sınıflama Başarıları

Klinik Tanı (Gerçek Durum)	Model Tahmini			
	Eğitim (Training)		Test (Validation)	
	Var	Yok	Var	Yok
Var	62	35	54	43
Yok	16	320	23	313
Doğruluk	% 88.22		% 84.76	
Duyarlılık	% 63.92		% 55.67	
Seçicilik	% 95.24		% 93.15	
Pozitif tahmini değer (PTD)	% 79.49		% 70.13	
Negatif tahmini değer (NTD)	% 90.14		% 87.92	
F-Measure	0.7086		0.6207	
ROC eğrisinin altında kalan alan	0.915378		0.861392	

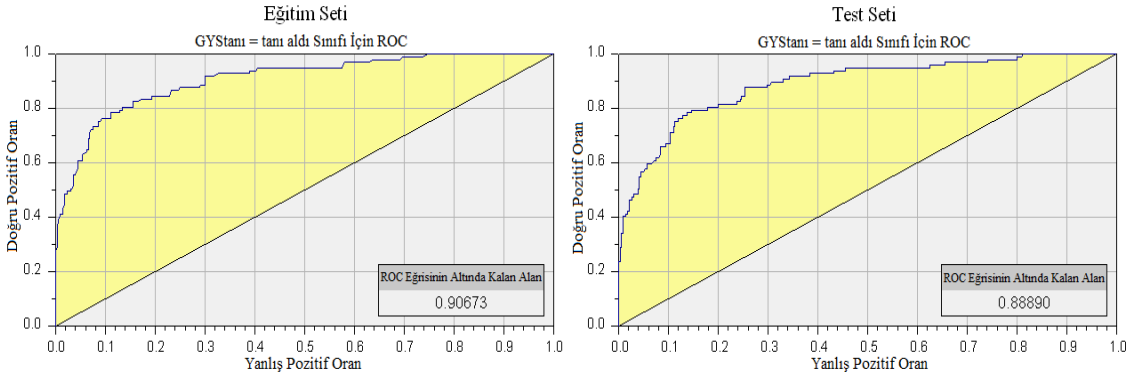
Tablo 4.12’de verilen değerlere ait duyarlılık-seçicilik ve ROC eğrisi grafikleri Şekil 4.6 ve Şekil 4.7’de verilmiştir.

Şekil 4.6 incelendiğinde, eğitim ve test setleri için hesaplanan olasılık kesim değerlerine karşı çizilen duyarlılık ve seçicilik değerlerine ait grafikler görülmektedir. Bu grafikler, hedef değişkenin iki sınıfa sahip olduğu sınıflama problemleri için uygun grafiklerdir ve bireyleri tanı aldı veya tanı almadı şeklinde sınıflamak için duyarlılık ve seçiciliğin, olasılık kesim noktası değiştirilerek nasıl ayarlanabileceğini göstermektedir.



Şekil 4.6: On Yedi Belirleyici Nitelik Kullanılarak Elde Edilmiş Eğitim ve Test Verileri için Duyarlılık ve Seçicilik Grafikleri

Şekil 4.7 incelendiğinde, eğitim ve test setleri için hesaplanan değerlerden elde edilen ROC eğrisi grafikleri görülmektedir. Grafikler incelendiğinde, hem eğitim, hem de test veri seti için ROC eğrisinin altında kalan alanların oldukça büyük (1'e yakın) olduğu görülmektedir ve bu nedenle modelin tahmin başarısının yüksek olduğu söylenebilir.



Şekil 4.7: On Yedi Belirleyici Nitelik Kullanılarak Elde Edilmiş Eğitim ve Test Verileri için Roc Eğrisi Grafikleri

On yedi belirleyici nitelik içeren DVM modeline ait “yükselme ve kazanç” değerleri Tablo 4.13’de verilmiştir.

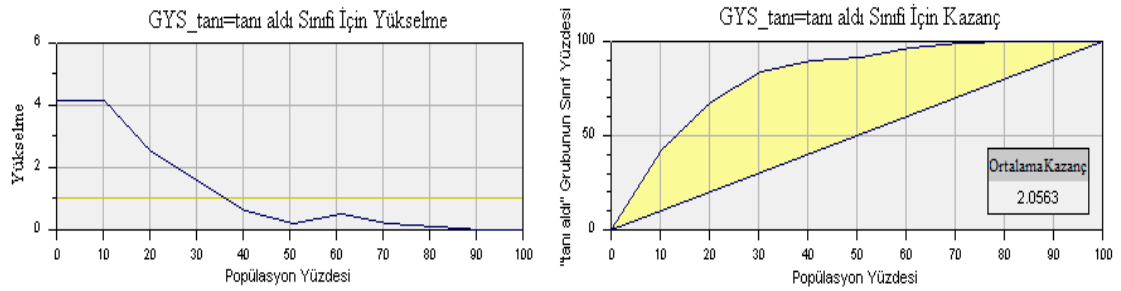
Tablo 4.13: DVM Yöntemi ile 17 Belirleyici Nitelik Kullanılarak Elde Edilen Yükselme ve Kazanç Değerleri

Bin İndeksi	Kesim Noktası Olasılığı	Binlerin Sınıf Yüzdeleri	Birikimli Popülasyon Yüzdeleri	Birikimli Sınıf Yüzdeleri	Birikimli Kazanç	Popülasyon Yüzdeleri	Sınıf Yüzdeleri	Yükselme
1	0.68147	81.82	10.16	37.11	3.65	10.16	37.11	3.65
2	0.43750	45.45	20.32	57.73	2.84	10.16	20.62	2.03
3	0.27539	43.18	30.48	77.32	2.54	10.16	19.59	1.93
4	0.18546	6.82	40.65	80.41	1.98	10.16	3.09	0.30
5	0.12086	18.18	50.81	88.66	1.74	10.16	8.25	0.81
6	0.07050	9.09	60.97	92.78	1.52	10.16	4.12	0.41
7	0.05069	9.09	71.13	96.91	1.36	10.16	4.12	0.41
8	0.03704	4.55	81.29	98.97	1.22	10.16	2.06	0.20
9	0.01798	2.27	91.45	100.00	1.09	10.16	1.03	0.10
10	0.00000	0.00	100.00	100.00	1.00	8.55	0.00	0.00

Ortalama kazanç = 2.056

GYS_tanı = tanı aldı sınıfındaki bireylerin yüzdesi = % 22.40

Şekil 4.8’de, tablodaki değerlere ilişkin yükselme ve kazanç grafikleri görülmektedir.



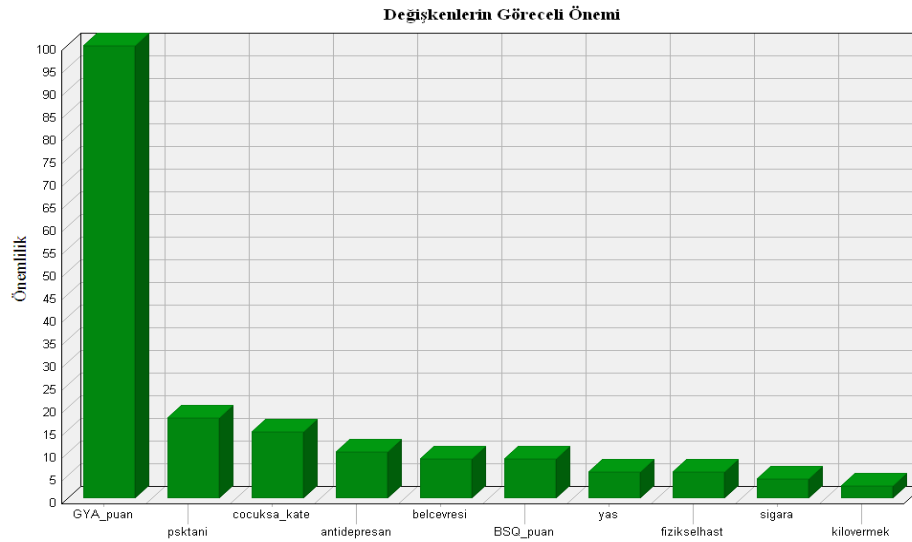
Şekil 4.8: DVM Yöntemi ile 17 Belirleyici Nitelik Kullanılarak Elde Edilen Yükselme ve Kazanç Grafikleri

Modele alınan ve tanı koymada etkili bulunan belirleyici niteliklerin önemlilik dereceleri incelendiğinde, sırasıyla GYA_puan, çocuk sayısı ve psikiyatrik tanı niteliklerinin en büyük değeri aldığı ve tanıdaki etkilerinin anlamlı olduğu söylenebilir. Bunların dışında, BSQ_puan, medeni durum ve eğitim yılının tanı koymadaki etkisi belirgin düzeyde daha düşük olmakla birlikte bu etkinin dolaylı bir etki olabileceği düşünülebilir. Tablo 4.14’te bu belirleyici nitelikler ve önem dereceleri verilmiştir. Tabloda sadece önem dereceleri sıfırdan büyük belirleyici nitelikler yer aldığı için, geriye kalan 11 belirleyici niteliğin tanıda hiçbir etkisi olmadığı söylenebilir.

Tablo 4.14: On Yedi Belirleyici Nitelik Kullanıldığında DVM ile Elde Edilmiş, Değişkenlerin Önem Dereceleri

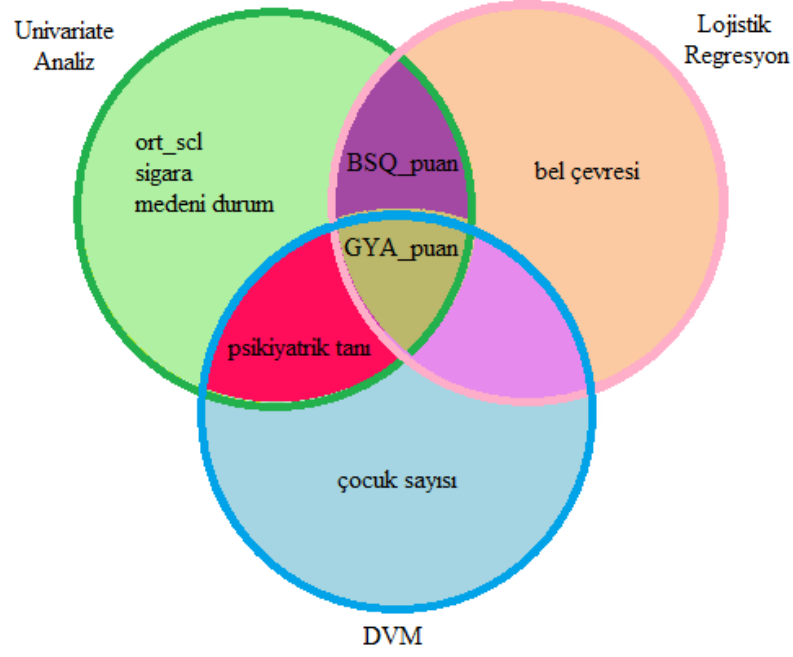
Belirleyici Nitelik	Önem Derecesi
GYA_puan	100.000
Çocuk Sayısı	15.385
Psikiyatrik Tanı	15.385
BSQ_puan	7.692
Medeni Durum	5.128
Eğitim Yılı	5.128

Şekil 4.9’da, önemli bulunan bu değişkenlerin önem derecelerini göreceli olarak gösteren grafik yer almaktadır.



Şekil 4.9: DVM Yönteminde Kullanılan 17 Belirleyici Nitelikten Önemli Bulunanlar

Şekil 4.10’da tek değişkenli analiz, lojistik regresyon ve DVM yönteminin sonuçları ilişkili olarak verilmiştir. Burada, her üç yaklaşımla da anlamlı bulunan GYA_puan değişkeni kesin risk faktörü olarak düşünülebilir. İkinci sıradaki risk faktörü, tek değişkenli analiz ve DVM yöntemlerinin kesişimindeki psikiyatrik tanı; üçüncü sıradaki risk faktörü, tek değişkenli analiz ve lojistik regresyon yöntemlerinin kesişimindeki BSQ_puan; dördüncü sıradaki risk faktörü, DVM yöntemi ile anlamlı bulunan çocuk sayısı, beşinci sıradaki risk faktörü lojistik regresyon ile anlamlı bulunan bel çevresi ve son sıradaki risk faktörleri ise, tek değişkenli analiz ile anlamlı bulunan ort_scl, sigara ve medeni durum değişkenleridir.



Şekil 4.10: Tek Değişkenli (Univariate) Analiz, Lojistik Regresyon ve DVM Yöntemi ile Anlamlı Bulunan Değişkenler

b) Üç belirleyici nitelik yardımıyla DVM yönteminden elde edilen sonuçlar

On yedi belirleyici nitelik kullanılarak DVM modeli geliştirildiğinde bu niteliklerden 6 tanesinin önemlilik dereceleri sıfırdan farklı bulunmuştur. Ancak bu önemlilik derecelerindeki ani düşüş dikkate alındığında 6 nitelikten 3'ünün önemlilik dereceleri belirgin düzeyde yüksek bulunmuştur. Önemlilik dereceleri yüksek olan bu 3 nitelik yardımıyla DVM modeli yeniden oluşturulmuş ve aşağıdaki sonuçlar elde edilmiştir.

Tablo 4.15'te, parametre tahmini sürecinde elde edilen değerler görülmektedir. Burada, parametre tahmini için yapılan arama sırasında değerlendirilen noktalarının sayısı 155, arama ile optimize edilen en iyi duyarlılık ve seçicilik değeri 0.752736 ve modelde kullanılan destek vektörlerin sayısı 193 olarak bulunmuştur. Durdurma kriteri ϵ (stopping criterion), 0.001 olarak belirlenmiş ve tahmin edilen parametre değerleri; hata maliyet değeri, $C = 137.018273$ ve RTF parametresi olan gamma değeri, $\gamma = 0.00106154$ olarak bulunmuştur.

Tablo 4.15: Üç Belirleyici Nitelik Kullanıldığı Durumda Parametre Tahmini Sürecinde Elde Edilen Değerler

		Tahmin Edilen Parametreler		
Değerlendirilen Nokta Sayısı	Optimize Edilen En Büyük Duyarlılık ve Seçicilik Değeri	ε	C	γ
155	0.752736	0.001	137.018273	0.00106154

Tablo 4.16’da modelin tahmin ettiği çeşitli olasılık değerlerinin kesim noktası olarak alındığı durumlardaki, model performansının göstergeleri yer almaktadır. Bu tablo incelendiğinde aşağıdaki sonuçlar elde edilmiştir:

- *Vakaların oranı:* Tablodaki eğitim setinde, 0.10 olasılık değeri kesim değeri olarak kabul edilirse vakaların %49.28’i seçilen hedef kategoriye atanır. Olasılık kesim değeri %80’ e çıkarıldığı zaman vakaların %8.52’si hedef kategoriye atanır geriye kalan denekler diğer kategoriye atanır. Eğer kesim değeri %41 alınırsa vakaların yaklaşık %20.00’si hedef kategoriye atanacaktır.
- *Katışıklık oranı (Impurity):* Tablodaki eğitim setinde, olasılık kesim değeri 0.10 alınırsa, vakaların %57.82’si gerçekte tanı almayan grupta olmasına rağmen, tanı alan gruptaymış gibi sınıflandırılır. Olasılık kesim değeri büyüdükçe katışıklık değeri küçülmektedir.
- *Kayıp değer:* Tablodaki eğitim setinde, 0.58 olasılık değeri kesim değeri olarak alınırsa gerçekte tanı almış olan bireylerin %44.33’ i tanı almadı olarak yanlış sınıflandırılmıştır.

Tablonun altında verilen çıktılar incelendiğinde aşağıdaki sonuçlar elde edilmiştir:

- *ROC eğrisi altında kalan alan:* Modelin tahmin olasılıkları yardımıyla çizilen ROC eğrisi altında kalan alan değeri, eğitim seti için 0.906726 ve test seti için 0.888899 olarak bulunmuştur.
- *Yanlış sınıflamayı minimize eden kesim noktası:* Veri setinin tümü için toplam yanlış sınıflama hatasını minimize edecek olasılık kesim noktası olan bu değer eğitim seti için 0.405945 ve test seti için 0.567276 olarak bulunmuştur.
- *Ağırlıklandırılmış yanlış sınıflamayı minimize eden kesim noktası:* Ağırlıklandırılmış yanlış sınıflama hatasını minimize edecek olasılık kesim

noktası olan bu değer eğitim seti için 0. 405945 ve test seti için 0. 567276 olarak bulunmuştur.

- *Yanlış sınıflamayı dengeleyen kesim noktası:* Bu değer, her bir hedef kategori için yanlış sınıflanan bireylerin sayılarının oranlarını yaklaşık olarak eşitleyecek olasılık kesim noktası olan bu değer eğitim seti için 0. 204053 ve test seti için 0. 230630 olarak bulunmuştur.

Tablo 4.16: Üç Belirleyici Nitelik Kullanıldığında DVM Yöntemi ile Elde Edilen Kesim Noktaları

Olasılık Kesim Değeri	Eğitim Seti				Test Seti			
	Deneklerin Oranı	Hata	Safılık (Impurity)	Kayıp (Loss)	Deneklerin Oranı	Hata	Safılık (Impurity)	Kayıp (Loss)
0.00	1.0000	0.7760	0.7760	0.0000	1.0000	0.7760	0.7760	0.0000
0.02	0.8351	0.6111	0.7318	0.0000	0.8099	0.5951	0.7291	0.0206
0.04	0.7144	0.4996	0.6929	0.0206	0.7192	0.5137	0.7014	0.0412
0.06	0.5950	0.3941	0.6429	0.0515	0.6073	0.4064	0.6501	0.0515
0.08	0.5551	0.3542	0.6172	0.0515	0.5411	0.3494	0.6158	0.0722
0.10	0.4928	0.3011	0.5782	0.0722	0.4888	0.3018	0.5795	0.0825
0.12	0.4378	0.2528	0.5328	0.0871	0.4510	0.2732	0.5545	0.1031
0.14	0.3853	0.2213	0.4965	0.1340	0.4116	0.2430	0.5230	0.1237
0.16	0.3838	0.2198	0.4945	0.1340	0.3875	0.2291	0.5066	0.1464
0.18	0.3502	0.1954	0.4592	0.1546	0.3713	0.2258	0.5024	0.1753
0.20	0.3488	0.1941	0.4571	0.1546	0.3565	0.2156	0.4882	0.1856
0.22	0.3132	0.1677	0.4100	0.1753	0.3364	0.1971	0.4599	0.1889
0.24	0.3120	0.1665	0.4078	0.1753	0.3233	0.1870	0.4427	0.1959
0.26	0.2858	0.1495	0.3696	0.1959	0.3089	0.1773	0.4244	0.2062
0.28	0.2847	0.1485	0.3673	0.1959	0.3055	0.1739	0.4179	0.2062
0.30	0.2585	0.1364	0.3306	0.2275	0.2960	0.1644	0.3993	0.2062
0.32	0.2576	0.1373	0.3317	0.2315	0.2748	0.1524	0.3696	0.2268
0.34	0.2567	0.1382	0.3329	0.2356	0.2701	0.1523	0.3672	0.2371
0.36	0.2306	0.1267	0.2889	0.2680	0.2518	0.1477	0.3485	0.2677
0.38	0.2297	0.1258	0.2862	0.2680	0.2420	0.1473	0.3414	0.2887
0.40	0.2288	0.1249	0.2835	0.2680	0.2405	0.1458	0.3374	0.2887
0.42	0.1981	0.1275	0.2564	0.3423	0.2301	0.1539	0.3476	0.3299
0.44	0.1973	0.1283	0.2575	0.3459	0.2146	0.1430	0.3111	0.3402
0.46	0.1965	0.1291	0.2585	0.3495	0.2060	0.1473	0.3139	0.3689
0.48	0.1797	0.1275	0.2314	0.3835	0.1989	0.1458	0.3033	0.3814
0.50	0.1789	0.1283	0.2324	0.3870	0.1801	0.1363	0.2564	0.4021

Tablo 4.16 (Devam): Üç Belirleyici Nitelik Kullanıldığında DVM Yöntemi ile Elde Edilen Kesim Noktaları

Olasılık Kesim Değeri	Eğitim Seti				Test Seti			
	Deneklerin Oranı	Hata	Safılık (Impurity)	Kayıp (Loss)	Deneklerin Oranı	Hata	Safılık (Impurity)	Kayıp (Loss)
0.52	0.1781	0.1291	0.2334	0.3905	0.1785	0.1346	0.2496	0.4021
0.54	0.1544	0.1290	0.1922	0.4433	0.1761	0.1357	0.2492	0.4097
0.56	0.1536	0.1282	0.1880	0.4433	0.1660	0.1359	0.2346	0.4330
0.58	0.1528	0.1274	0.1837	0.4433	0.1583	0.1350	0.2189	0.4481
0.60	0.1290	0.1319	0.1432	0.5065	0.1515	0.1372	0.2134	0.4679
0.62	0.1281	0.1328	0.1442	0.5105	0.1430	0.1411	0.2100	0.4959
0.64	0.1273	0.1337	0.1452	0.5144	0.1316	0.1386	0.1755	0.5156
0.66	0.1196	0.1322	0.1159	0.5281	0.1245	0.1407	0.1651	0.5361
0.68	0.1187	0.1331	0.1168	0.5321	0.1224	0.1385	0.1506	0.5361
0.70	0.1178	0.1339	0.1176	0.5361	0.1199	0.1365	0.1349	0.5371
0.72	0.1032	0.1424	0.1045	0.5876	0.1176	0.1387	0.1374	0.5470
0.74	0.1021	0.1414	0.0954	0.5876	0.1143	0.1421	0.1415	0.5621
0.76	0.0893	0.1432	0.0474	0.6202	0.1052	0.1445	0.1219	0.5876
0.78	0.0871	0.1432	0.0358	0.6251	0.1017	0.1455	0.1136	0.5977
0.80	0.0852	0.1435	0.0271	0.6302	0.0994	0.1433	0.0940	0.5979
0.82	0.0839	0.1448	0.0275	0.6359	0.0862	0.1516	0.0803	0.6459
0.84	0.0711	0.1575	0.0325	0.6927	0.0841	0.1538	0.0824	0.6555
0.86	0.0696	0.1591	0.0332	0.6998	0.0718	0.1614	0.0643	0.6999
0.88	0.0587	0.1653	0.0000	0.7381	0.0660	0.1627	0.0350	0.7158
0.90	0.0475	0.1765	0.0000	0.7878	0.0592	0.1695	0.0390	0.7461
0.92	0.0361	0.1880	0.0000	0.8391	0.0432	0.1809	0.0000	0.8073
0.94	0.0236	0.2004	0.0000	0.8947	0.0334	0.1906	0.0000	0.8508
0.96	0.0132	0.2108	0.0000	0.9412	0.0241	0.1999	0.0000	0.8923
0.98	0.0059	0.2181	0.0000	0.9736	0.0060	0.2180	0.0000	0.9731
1.00	0.0000	0.2240	0.0000	1.0000	0.0000	0.2240	0.0000	1.0000

Eğitim seti için ROC eğrisi altında kalan alan = 0.906726

Eğitim seti için yanlış sınıflamayı minimize eden kesim noktası = 0. 405945

Eğitim seti için ağırlıklandırılmış yanlış sınıflamayı minimize eden kesim noktası = 0. 405945

Eğitim seti için yanlış sınıflamayı dengeleyen kesim noktası = 0. 204053

Test seti için ROC eğrisi altında kalan alan = 0.888899

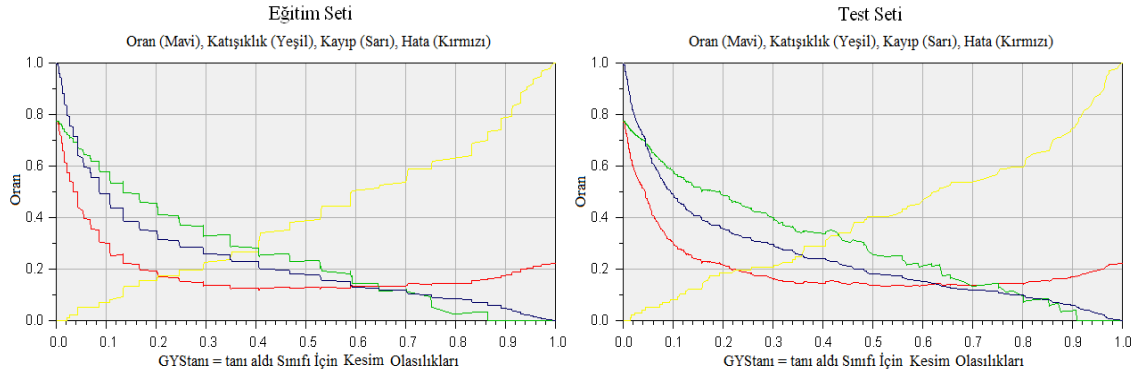
Test seti için yanlış sınıflamayı minimize eden kesim noktası = 0. 567276

Test seti için ağırlıklandırılmış yanlış sınıflamayı minimize eden kesim noktası = 0. 567276

Test seti için yanlış sınıflamayı dengeleyen kesim noktası = 0. 230630

Bu ölçüler değerlendirildiğinde en uygun olasılık kesim değeri eğitim ve test verileri için sırasıyla 0.41 ve 0.57 olarak belirlenmiş ve bu değer üstünde olasılık değerine sahip olan bireyler tanı alan grupta kabul edilmiştir.

Şekil 4.11’de yukarıda elde edilen kesim noktası değerleri için oluşturulan kesim olasılıkları grafiği görülmektedir.



Şekil 4.11: Üç Belirleyici Nitelik Kullanılarak Elde Edilmiş Eğitim ve Test Verileri için Kesim Noktası Grafikleri

Tablo 4.17’de, GYS_tanı hedef değişkeni için, hedef kategorinin “tanı aldı” olarak seçilmesi durumunda elde edilen olasılık düzenlemesi yer almaktadır. Tablo incelendiğinde, 0.00-0.10 olasılık aralığında toplam bireylerin % 50.58’i (219 birey) “tanı aldı” grubuna sınıflandırıldığı görülmektedir

Tabloda, tahmin edilen olasılık değerleri aralığının 0.10-0.20 olduğu 2. satıra bakılırsa, bu aralıkta bütün veri setinin % 14.32’sini oluşturan 62 kişinin olduğu görülmektedir. Bu bireyler için tahmin edilen ortalama olasılık değeri, aralığın hemen hemen orta noktası olan 0.1318 değeridir. Eğer model başarılı bir model ise, bu bireylere ait gerçek olasılık değerinin 0.1318 olması beklenir. Ancak *Gerçek Değer* sütununa bakıldığında, bu değer 0.1290 olduğu görülmektedir.

Tablonun altında verilen bilgilere göre aşağıdaki açıklamalar yapılabilir.

- *Ağırlıklandırılmış ortalama olasılık hatası:* Olasılık aralığı 0.10-0.20 olan 2. satır ele alınır, burada hata $0.1318 - 0.1290 = 0.0028$ olarak elde edilir ve tüm bireylerin % 14.32’si bu aralıkta yer almaktadır. Bu durumda bu hata, toplam ortalama hatanın % 14.32’sine katkıda bulunmaktadır.

- *Ağırlıklandırılmış ortalama kare olasılık hatası:* Bu değer ağırlıklandırılmış ortalama olasılık hatası ile aynı şekilde hesaplanmaktadır. Ancak, ağırlığı ile çarpılmadan önce, hatanın karesi alınır ve toplam hataya eklenir. Toplam hata kareler hesaplandıktan sonra, bu değerın karekökü bulunur. Verilen olasılık düzenleme tablosunun altındaki değer, bu karekök değeridir.

Tablo 4. 17: Üç Belirleyici Nitelik Kullanıldığında Elde Edilen Olasılık Düzenlemesi

Tahmin Edilen Olasılık Aralığı	Eğitim Seti				Test Seti			
	Sayı	%	Tahmin Edilen Değer	Gerçek Değer	Sayı	%	Tahmin Edilen Değer	Gerçek Değer
0.00-0.10	219	50.58	0.0373	0.0320	221	51.04	0.0383	0.0362
0.10-0.20	62	14.32	0.1318	0.1290	57	13.16	0.1398	0.1754
0.20-0.30	40	9.24	0.2434	0.1750	26	6.00	0.2413	0.0769
0.30-0.40	12	2.77	0.3490	0.3333	24	5.54	0.3344	0.3333
0.40-0.50	22	5.08	0.4285	0.500	27	6.24	0.4529	0.4074
0.50-0.60	22	5.08	0.5602	0.5455	13	3.00	0.561	0.4615
0.60-0.70	4	0.92	0.6467	0.5000	14	3.23	0.6418	0.5000
0.70-0.80	15	3.46	0.7348	0.6667	8	1.85	0.7504	0.7500
0.80-0.90	16	3.70	0.8606	0.9375	18	4.16	0.8456	0.7778
0.90-1.00	21	4.85	0.9458	10.000	26	6.00	0.9473	0.9615

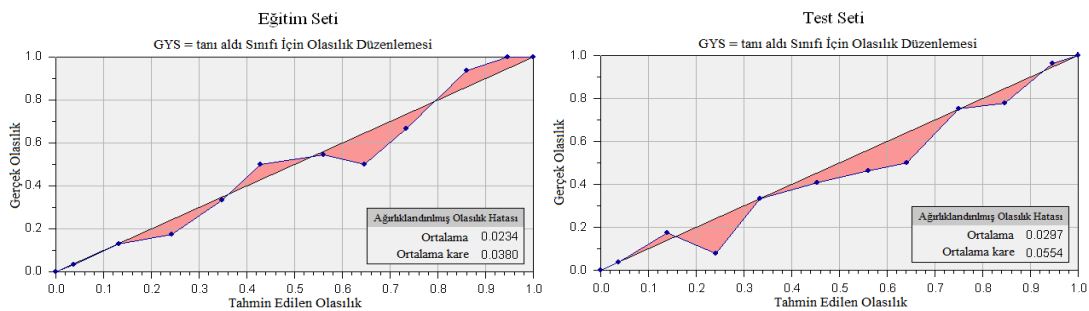
Eğitim seti için ağırlıklandırılmış ortalama olasılık hatası = 0.023426

Eğitim seti için ağırlıklandırılmış ortalama kare olasılık hatası = 0.038027

Test seti için ağırlıklandırılmış ortalama olasılık hatası = 0.029787

Test seti için ağırlıklandırılmış ortalama kare olasılık hatası = 0.055428

Şekil 4.12’de, elde edilen değerler için olasılık düzenleme grafiği görülmektedir. Şekiller incelendiğinde, hem eğitim hem de test verileri için hatayı gösteren kırmızı taralı alanın küçük ve tahminin başarılı olduğu söylenebilir.



Şekil 4.12: Üç Belirleyici Nitelik Kullanılarak Elde Edilmiş Eğitim ve Test Verileri için Olasılık Düzenleme Grafikleri

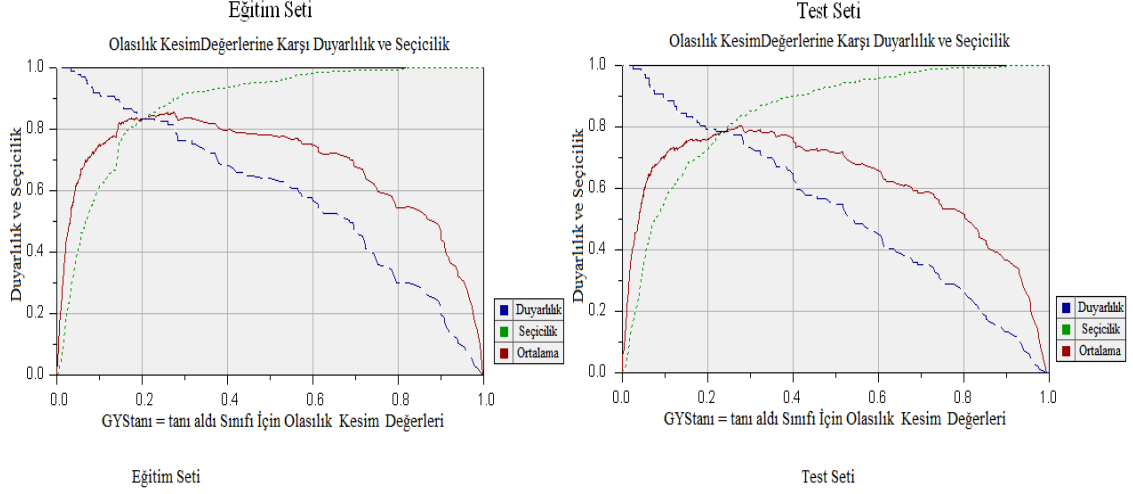
Kurulan bu DVM modeline göre hesaplanan sınıflama başarıları Tablo 4.18’de verilmiştir. Tablo incelendiğinde, doğruluk, seçicilik ve negatif tahmini değerin hem eğitim, hem de test verilerinde yüksek, duyarlılık ve pozitif tahmini değerin ise her iki grupta da nispeten daha düşük olduğu görülmüştür. Ayrıca ROC eğrisinin altında kalan alan eğitim seti için 0.906726, test seti için 0.888899 olarak bulunmuştur. Bulunan bu değerler ve ROC eğrisinin altında kalan alanlar birlikte incelendiğinde, modelin tanı koyma başarısı yüksek olarak değerlendirilmiştir.

Tablo 4.18: Önemli Bulunan Üç Belirleyici Nitelik Kullanılarak DVM Yöntemi ile Elde Edilmiş Eğitim ve Test Verilerinde Modelin Sınıflama Başarıları

Klinik Tanı (Gerçek Durum)	Model Tahmini			
	Eğitim (Training)		Test (Validation)	
	Var	Yok	Var	Yok
Var	60	37	58	39
Yok	18	318	21	315
Doğruluk	% 87.30		% 86.14	
Duyarlılık	% 61.86		% 59.79	
Seçicilik	% 94.64		% 93.75	
Pozitif tahmini değer (PTD)	% 76.92		% 73.42	
Negatif tahmini değer (NTD)	% 89.58		% 88.98	
F-Measure	0. 6857		0. 6591	
ROC eğrisinin altında kalan alan	0. 906726		0. 888899	

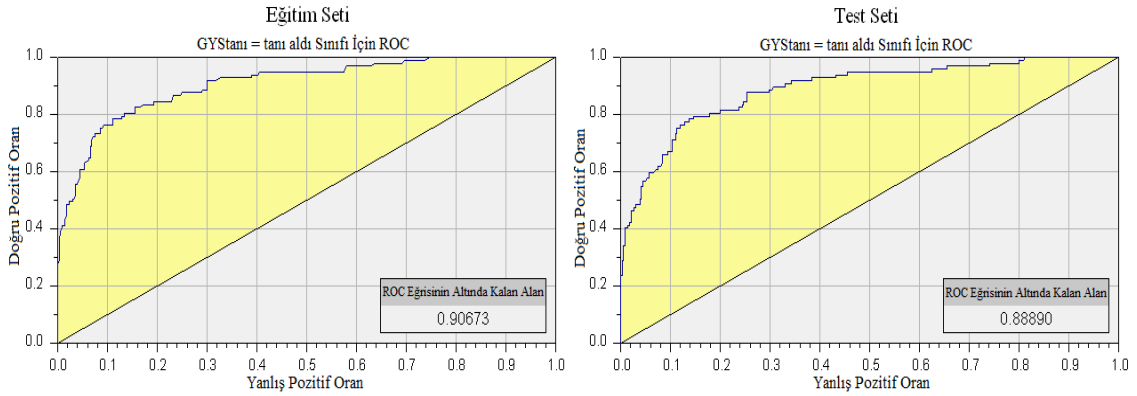
Tablo 4.18’de verilen değerlere ait duyarlılık-seçicilik ve ROC eğrisi grafikleri Şekil 4.13 ve Şekil 4.14’te verilmiştir.

Şekil 4.13 incelendiğinde, eğitim ve test setleri için hesaplanan olasılık kesim değerlerine karşı çizilen duyarlılık ve seçicilik değerlerine ait grafikler görülmektedir.



Şekil 4.13: Üç Belirleyici Nitelik Kullanılarak Elde Edilmiş Eğitim ve Test Verileri için Duyarlılık ve Seçicilik Grafikleri

Şekil 4.14 incelendiğinde, eğitim ve test setleri için hesaplanan değerlerden elde edilen ROC eğrisi grafikleri görülmektedir. Grafikler incelendiğinde, hem eğitim, hem de test veri seti için ROC eğrisinin altında kalan alanların oldukça büyük (1'e yakın) olduğu görülmektedir ve bu nedenle modelin tahmin başarısının yüksek olduğu söylenebilir.



Şekil 4.14: Üç Belirleyici Nitelik Kullanılarak Elde Edilmiş Eğitim ve Test Verileri için ROC Eğrisi Grafikleri

Tablo 4.19'da DVM yöntemi ile eğitim setinden elde edilen modele ait "yükselme ve kazanç" değerleri görülmektedir.

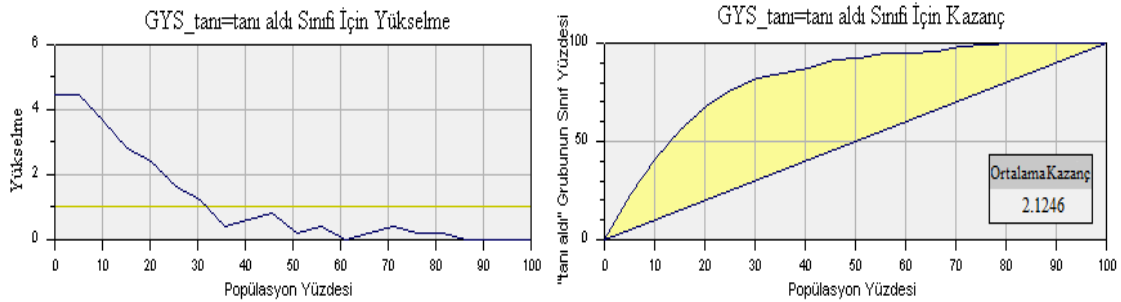
Tablo 4.19: DVM Yöntemi ile 3 Belirleyici Nitelik Kullanılarak Elde Edilen Yükseltme ve Kazanç Değerleri

Bin İndeksi	Kesim Noktası Olasılığı	Binlerin Sınıf Yüzdeleri	Birikimli Popülasyon Yüzdeleri	Birikimli Sınıf Yüzdeleri	Birikimli Kazanç	Popülasyon Yüzdeleri	Sınıf Yüzdeleri	Yükseltme
1	0.80095	88.64	10.16	40.21	3.96	10.16	40.21	3.96
2	0.46329	50.00	20.32	62.89	3.09	10.16	22.68	2.23
3	0.28898	36.36	30.48	79.38	2.60	10.16	16.49	1.62
4	0.15488	18.18	40.65	87.63	2.16	10.16	8.25	0.81
5	0.09242	11.36	50.81	92.78	1.83	10.16	5.15	0.51
6	0.05992	4.55	60.97	94.85	1.56	10.16	2.06	0.20
7	0.04352	2.27	71.13	95.88	1.35	10.16	1.03	0.10
8	0.01967	4.55	81.29	97.94	1.20	10.16	2.06	0.20
9	0.01004	4.55	91.45	100.00	1.09	10.16	2.06	0.20
10	0.00000	0.00	100.00	100.00	1.00	8.55	0.00	0.00

Ortalama kazanç = 1.984

GYSt_tanı = tanı aldı sınıfındaki bireylerin yüzdesi = % 22.40

Şekil 4.15'te, tablodaki değerlere ilişkin yükseltme ve kazanç grafikleri görülmektedir.



Şekil 4.15: Üç Belirleyici Nitelik Kullanılarak Elde Edilmiş Yükseltme ve Kazanç Grafikleri

Yukarıda bahsedilen 17 değişkenli lojistik regresyon, 3 değişkenli lojistik regresyon, 17 belirleyici nitelikli DVM ve 3 belirleyici nitelikli DVM yöntemleri kullanılarak elde edilen eğitim setlerine ait sonuçların karşılaştırması Tablo 4.20'de verilmiştir. Tablo incelendiğinde, tanı yapmak için bulunan modellerin tanı başarıları (performansları) arasında anlamlı düzeyde fark bulunmamıştır (tüm p değerleri > 0.05).

Tablo 4.20: İncelenen Model Performanslarının Karşılaştırması

		Doğruluk	Duyarlılık	Seçicilik
Karşılaştırma 1	Lojistik_17	% 89.84	% 69.07	% 95.83
	Lojistik_3	% 88.68	% 65.98	% 95.24
P Değeri		0.8572	0.7932	0.937
Karşılaştırma 2	Lojistik_17	% 89.84	% 69.07%	% 95.83
	DVM_17	% 88.22	%63.92%	% 95.24
P Değeri		0.8010	0.6594	0.937
Karşılaştırma 3	Lojistik_17	% 89.84	% 69.07	% 95.83
	DVM_3	% 87.30	% 61.86	% 94.64
P Değeri		0.6912	0.8562	0.8744
Karşılaştırma 4	Lojistik_3	% 88.68	% 65.98	% 95.24
	DVM_17	% 88.22	% 63.92	% 95.24
P Değeri		0.9424	0.8586	1.000
Karşılaştırma 5	Lojistik_3	% 88.68	% 65.98	% 95.24
	DVM_3	% 87.30	% 61.86	% 94.64
P Değeri		0.828	0.7192	0.9368
Karşılaştırma 6	DVM_17	% 88.22	% 63.92	% 95.24
	DVM_3	% 87.30	% 61.86	% 94.64
P Değeri		0.8846	0.8562	0.9368

Bu sonuçlara ilaveten, doğrusal, polinomiyal ve sigmoid çekirdek fonksiyonlar kullanılarak on yedi doğal nitelik yardımıyla DVM yönteminin karar fonksiyonları yeniden oluşturulmuş ve Tablo 4.21, 4.22 ve 4.23'te bu karar fonksiyonlarının tanı

başarıları verilmiştir. Tablolar incelendiğinde, elde edilen sonuçların RTF ile elde edilen sonuçlar ile benzerlik gösterdiği görülmüştür.

Tablo 4. 21: Eğitim ve Test Verilerinde Doğrusal Çekirdek Fonksiyon ve 17 Belirleyici Nitelik Kullanılarak Elde Edilen Modelin Tanı Başarıları

Klinik Tanı (Gerçek Durum)	Model Tahmini			
	Eğitim (Training)		Test (Validation)	
	Var	Yok	Var	Yok
Var	63	34	49	48
Yok	13	323	18	318
Doğruluk	% 89.15		% 84.76	
Duyarlılık	% 64.95		% 50.52	
Seçicilik	% 96.13		% 94.64	
Pozitif tahmini değer (PTD)	% 82.89		% 73.13	
Negatif tahmini değer (NTD)	% 90.48		% 86.89	
F-Measure	0. 7283		0. 5976	
ROC eğrisinin altında kalan alan	0. 919980		0. 887058	

Tablo 4. 22: Eğitim ve Test Verilerinde Polinomial Çekirdek Fonksiyon ve 17 Belirleyici Nitelik Kullanılarak Elde Edilen Modelin Tanı Başarıları

Klinik Tanı (Gerçek Durum)	Model Tahmini			
	Eğitim (Training)		Test (Validation)	
	Var	Yok	Var	Yok
Var	63	34	51	46
Yok	13	323	17	319
Doğruluk	% 89.15		% 85.45	
Duyarlılık	% 64.95		% 52.58	
Seçicilik	% 96.13		% 94.94	
Pozitif tahmini değer (PTD)	% 82.89		% 75.00	
Negatif tahmini değer (NTD)	% 90.48		% 87.40	
F-Measure	0. 7283		0. 6182	
ROC eğrisinin altında kalan alan	0. 920072		0. 881443	

Tablo 4. 23: Eğitim ve Test Verilerinde Sigmoid Çekirdek Fonksiyon ve 17 Belirleyici Nitelik Kullanılarak Elde Edilen Modelin Tanı Başarıları

Klinik Tanı (Gerçek Durum)	Model Tahmini			
	Eğitim (Training)		Test (Validation)	
	Var	Yok	Var	Yok
Var	59	38	52	45
Yok	12	324	22	314
Doğruluk	% 88.45		% 84.53	
Duyarlılık	% 60.82		% 53.61	
Seçicilik	% 96.43		% 93.45	
Pozitif tahmini değer (PTD)	% 83.10		% 70.27	
Negatif tahmini değer (NTD)	% 89.50		% 87.47	
F-Measure	0. 7024		0. 6082	
ROC eğrisinin altında kalan alan	0. 907462		0. 878068	

Kullanılan 4 farklı çekirdek fonksiyon için analiz süreleri Tablo 4.24’te verilmiştir. Tablo incelendiğinde RTF kullanıldığı durumda harcanan sürenin özellikle doğrusal ve polinomiyal çekirdek fonksiyonlara göre oldukça düşük olduğu görülmektedir.

Tablo 4. 24: Kullanılan 4 Farklı Çekirdek Fonksiyonu için Analiz Süreleri

	RTF	Doğrusal	Polinomiyal	Sigmoid
Analiz Süreleri (dk:sn:salise)	00:28.58	29:51.06	37:17.67	02:35.80

Eğitim setinde RTF, doğrusal, polinomiyal ve sigmoid çekirdek fonksiyonlarına ait sonuçların karşılaştırması Tablo 4.25’te verilmiştir. Tablo incelendiğinde, kullanılan çekirdek fonksiyonların tahmin başarı göstergeleri arasında anlamlı düzeyde fark bulunmamıştır (tüm p değerleri > 0.05).

Tablo 4. 25: DVM Yönteminde 17 Belirleyici Nitelik ve 4 Farklı Çekirdek Fonksiyon Kullanılarak ile Elde Edilen Tahmin Başarı Göstergelerinin Karşılaştırması

		Doğruluk	Duyarlılık	Seçicilik
Karşılaştırma 1	RTF	% 88.22	% 63.92	% 95.24
	Doğrusal	% 89.15	% 64.95	% 96.13
P Değeri		0.8852	0.9288	0.9058
Karşılaştırma 2	RTF	% 88.22	% 63.92	% 95.24
	Polinomial	% 89.15	% 64.95	% 96.13
P Değeri		0.8852	0.9288	0.9058
Karşılaştırma 3	RTF	% 88.22	% 63.92	% 95.24
	Sigmoid	% 88.45	% 60.82	% 96.43
P Değeri		0.9712	0.785	0.9058
Karşılaştırma 4	Doğrusal	% 89.15	% 64.95	% 96.13
	Polinomial	% 89.15	% 64.95	% 96.13
P Değeri		1.000	1.000	1.000
Karşılaştırma 5	Doğrusal	% 89.15	% 64.95	% 96.13
	Sigmoid	% 88.45	% 60.82	% 96.43
P Değeri		0.9138	0.717	0.9686
Karşılaştırma 6	Polinomial	% 89.15	% 64.95	% 96.13
	Sigmoid	% 88.45	% 60.82	% 96.43
P Değeri		0.9138	0.717	0.9686

5. TARTIŞMA ve SONUÇ

DVM metodu, diğerk bazı yöntemlere göre daha yeni ve çeşitli problemler için doğru modeller üretebilme yeteneğine sahip bir modelleme yöntemidir ve teorik istatistik temellere dayanarak, özellikle verinin doğrusal olarak ayrılammama durumunda doğru ve güçlü sonuçlar üretir^{16,107}.

Doğrusal olarak ayrılammayan verileri sınıflandırmak için DVM yöntemi, çeşitli çekirdek fonksiyonlarını kullanmaktadır ve iyi sonuçlar elde etmeyi sağlayacak olan bu fonksiyonların doğru olarak seçilmesi yöntemin en önemli unsurlarındandır. Çekirdek dönüşümü için seçilmesi gereken fonksiyona karar vermede karşılaşılabilecek zorluk, DVM'nin bir dezavantajı olarak düşünülebilir¹⁰⁸. Ancak çeşitli çalışmalarda, radyal tabanlı çekirdek fonksiyonunun birçok koşulda yüksek sınıflama doğruluğu verdiği bulunmuş ve bu nedenle DVM yöntemiyle sınıflama yapıldığında bu fonksiyonun kullanılması önerilmiştir³.

DVM yöntemi, yapay sinir ağları gibi bazı yaklaşımlardan radikal farklılıklar gösterir. Örneğin, DVM ile eğitim her zaman global bir minimum bulur ve basit geometrik yorumu daha fazla araştırma için verimli bir zemin sağlar⁶⁸.

Bunların yanı sıra DVM yönteminde bir karar fonksiyonu elde edilebilmekte ve bu fonksiyon yardımıyla yeni bir gözlem ait olduğu sınıfa sınıflandırılabilir. Verilerin dağılımı ve nitelikler arası ilişkilerin şeklinde her hangi bir varsayım gerektirmemesi ise yöntemin birçok alanda kullanımını yaygınlaştıracaktır.

Uygulamada, gece yeme sendromu tanısı koyabilmek için 17 değişkenden oluşan veri seti kullanılmıştır. Bu veri setinde, GYS_tanı değişkeni kişilerin gece yeme sendromu tanısı alıp almadığını gösteren hedef değişkendir ve burada amaç, diğerk değişkenlerden faydalanarak kişileri hedef değişkendeki gerçek gruplarına sınıflamaktır. Bu amacı gerçekleştirmek için, DVM yönteminin yanında, tek değişkenli analizler ve lojistik regresyon analizi de kullanılmıştır.

Tek değişkenli analiz ile elde edilen sonuçlarda, 17 değişkenden GYA_puan, BSQ_puan, scl_ort, medeni durum, sigara kullanımı ve psikiyatrik tanı değişkenlerinin gece yeme sendromu tanısı koymada etkili değişkenler olduğu bulunmuştur (Tablo 4.1 ve Tablo 4.2).

Lojistik regresyon analizi ile elde edilen sonuçlarda, 17 deęişkinden bel çevresi, GYA_puan ve BSQ_puan deęişkenlerinin tanı koymada etkili olduęu bulunmuştur (Tablo 4.3). Bu yöntemle elde edilen sınıflama başarılarına bakıldığında, eğitim ve test setlerinde sınıflama başarılarının yüksek olduęu görülmüştür (Tablo 4.4). Ayrıca, modele katkısı anlamlı bulunan 3 deęişkenle analiz tekrarlandığında, sınıflama başarıları aynı şekilde yüksek bulunmuştur (Tablo 4.7). Bu durumda, 3 deęişken kullanılarak da kişilere doęru tanı konulabileceęi ve dolayısıyla 17 deęişken yerine bu 3 deęişkenin tanı koymada kullanılabilceęi söylenebilir.

DVM yönteminde RTF kullanılarak elde edilen sonuçlarda, 17 belirleyici nitelikten 6 tanesi DTREG programı tarafından önemli bulunmuş ancak bu 6 nitelikten, önem derecesi 3. nitelikten sonra ani bir düşüş gösterdiği için ilk 3 tanesinin modele katkısı anlamlı olarak düşünölmüştür. Bu belirleyici nitelikler, GYA_puan, çocuk sayısı ve psikiyatrik tanı olarak elde edilmiştir. DVM yöntemiyle, 17 belirleyici nitelik yardımıyla bulunan sınıflama başarıları da yine lojistik regresyon ile benzer sonuçlar vermiş ve burada da yüksek sınıflama doęruluęu elde edilmiştir (Tablo 4.12). Ayrıca, modele katkısı anlamlı bulunan 3 belirleyici nitelikle analiz tekrarlandığında, sınıflama başarıları aynı şekilde yüksek bulunmuştur (Tablo 4.18). Bu durumda, 3 deęişken kullanılarak da kişilere doęru tanı konulabileceęi ve dolayısıyla 17 deęişken yerine bu 3 deęişkenin tanı koymada kullanılabilceęi düşünölebilir. Ancak sırasıyla 17 ve 3 belirleyici nitelik kullanılarak elde edilen optimum parametre deęerlerine bakıldığında, 17 belirleyici nitelik kullanıldığında elde edilen hata maliyetinin, 3 belirleyici nitelik kullanıldığı duruma kıyasla oldukça düşük olduęu (Tablo 4.9 ve Tablo 4.15) ve bu nedenle 17 belirleyici nitelikle elde edilen sonuçların daha başarılı sonuçlar olduęu söylenebilir.

Bunların yanı sıra DVM yönteminin kullanıldığı pratik uygulamalarda sık tercih edilen doęrusal, polinomial ve sigmoid çekirdek fonksiyonları kullanılarak yeni modeller oluşturulmuştur. Bu modellere ait performans ölçüleri, RTF çekirdek fonksiyonu kullanıldığında elde edilen modelin performansından anlamlı fark göstermemiştir. Ancak analiz sürelerine bakıldığında, RTF'nin kullanıldığı durum için bu sürenin oldukça kısa olduęu söylenebilir ve bu da RTF kullanmanın bir avantajı olarak düşünölebilir. Ayrıca, parametre tahminleri için dezavantaj olarak gösterilen zamansal

sorunun günümüz teknolojisinde ve kullanılan yüksek performanslı bilgisayarlar sayesinde artık bir problem olmaktan çıktığı söylenebilir.

Bütün bu koşullar topluca değerlendirildiğinde, modeller içerisinde en basit yapıda olan ve klinik açıdan en çok anlam taşıyan modelin seçilmesi önerilebilir. Ancak şu unutulmamalıdır, bu çalışmada yapılan karşılaştırmalar tek bir veri seti üzerinde gerçekleştirilmiştir. Bu durumda amaca uygun yöntemin seçimine yönelik hatalar söz konusu olabilir. Model seçiminde en doğru yaklaşım, genelleme yeteneği daha geniş olan, kısa sürede oluşturulabilen ve doğrusal olmayan ilişkileri dikkate alan modelin tercih edilmesidir.

Bu çalışmanın uygulama bölümünde kullanılan, 17 doğal nitelikten oluşan ve toplam 433 kişiye ait olan veri seti, veri madenciliği araştırmalarında orta büyüklükte kabul edilebilecek büyüklüktedir. Bu veri seti için lojistik regresyon ve DVM yöntemi birbirlerine benzer sonuçlar üretmiş olsalar da, DVM'nin birçok yönden lojistik regresyona göre üstünlüğü olduğu söylenebilir. DVM yöntemi, istatistiksel öğrenme teorisine dayanan yeni ve doğrusal olmayan, karmaşık yapıya sahip verileri sınıflamada daha etkin bir yöntemdir. Lojistik regresyon yönteminde, DVM'ye kıyasla aşırı uyum sorununun ortaya çıkma olasılığı yüksektir. Her iki yöntemde de amaç tahmin doğruluğunu maksimum yapmaktır, ancak lojistik regresyon verinin olabilirliğini maksimumlaştırarak bu doğruluğu sağlamaya çalışan dolaylı bir metotken, DVM ise bu doğruluğu sınıflar arasındaki marjini maksimumlaştırarak sağlayan daha dolaysız bir yöntemdir^{109,110}.

Yapılan yurt dışı çalışmalar incelendiğinde, veri madenciliği yöntemlerinin tıp alanında yoğun bir şekilde kullanıldığı görülmektedir. Bu yöntemler içerisinde de random forest ve destek vektör makineleri taşıdıkları önemli avantajlar nedeniyle, hastalıkların risk faktörlerini belirlemede klasik yöntemlere tercih edilmektedir. Elde edilen sonuçlar değerlendirildiğinde ise risk faktörlerinin başarılı bir şekilde belirlendiği ve model performanslarının oldukça yüksek olduğu gözlenmiştir^{111,112}.

DVM nin tıp alanında en yaygın kullanım alanlarından birisi de değişken sayısı ve gözlem sayısı açısından oldukça kapsamlı veri içeren gen çalışmalarıdır. Son yılların popüler konularından birisi olan gen-hastalık ilişkisinin DVM yöntemi gibi veri

madenciliği yöntemleri ile incelenmesi, hem güvenilir hem de kullanılabilir sonuçlar vermektedir^{113,114}.

Gece yeme sendromunu etkileyen risk faktörlerinin belirlenmesi amacıyla yapılan çalışmalarda, potansiyel risk faktörleri ya tek değişkenli analiz yöntemleriyle incelenmiş veya bilinen çoklu modeller kullanılmıştır^{115,116}. Bu çalışmada ise ilk defa gece yeme sendromunu etkileyen risk faktörleri DVM yöntemiyle incelenmiştir.

Konu ile ilgili dikkate alınmayan veya ölçülmeyen veya ölçülemeyen risk faktörleri söz konusu olabilir. Ayrıca dikkate alınan risk faktörlerinin, gece yeme sendromu ile doğrudan ilişkileri incelenmiş, risk faktörleri arasındaki interaksiyonlar değerlendirilmemiştir. Bu iki durum çalışmanın bir sınırlılığı olarak düşünülebilir.

Son olarak, DVM yönteminin tanı koyma veya sınıflandırma işlevinin yanı sıra tahmin ve yapı tanıma işlevlerinin de sağlık alanında önemli uygulama alanları bulacağı söylenebilir. Mesela çeşitli genetik yapılar için gelecekte ortaya çıkması muhtemel hastalıkların belirlenmesi, kişilerin fiziksel özellikleri, çevre şartları, genetik yapı özelliklerinden yararlanılarak yaşam sürelerinin tahmin edilmesi veya hastalık profillerinin çıkarılması, hastane kalite düzeyinin tahmin edilmesi, hastane gelir düzeyini etkileyen risk faktörlerinin belirlenmesi ve buna bağlı olarak gelir düzeyi artırıcı tedbirlerin alınması, poliklinikte hasta muayene sürelerinin tahmin edilmesi ve randevu sisteminin geliştirilmesi gibi birçok konuda DVM yöntemi hizmet verme yeteneğine sahiptir.

6. KAYNAKLAR

1. Koyuncugil AS, Özgülbaş N. Veri Madenciliği: Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları. Bilişim Teknolojileri Dergisi. 2009;(2)2:21-31.
2. Baykal N. Tıpta Veri Madenciliği ve Bilgi Keşfi. Tıp Bilişimi Güz Okulu. Ekim 2003
3. Aydoğan Ü. Destek Vektör Makinalarında Kullanılan Çekirdek Fonksiyonların Sınıflama Performanslarının Karşılaştırılması. 2010, Hacettepe Üniversitesi, Sağlık Bilimleri Enstitüsü, Yüksek Lisans Tezi, 71 sayfa. Ankara. (Prof. Dr. Osman Saraçbaşı Doç. Dr. Hafize Sezer)
4. Çelik M, Veri Madenciliğinde Kullanılan Sınıflandırma Yöntemleri ve Bir Uygulama. 2009, İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, Ekonometri Anabilimdalı, Yüksek Lisans Tezi, 61 sayfa, İstanbul. (Doç. Dr. Enis Sınıksaran).
5. Amasyalı MF. Makine Öğrenmesine Giriş. Yıldız Teknik Üniversitesi Makine Mühendisliği Bölümü Ders Notları.
<http://www.ce.yildiz.edu.tr/myindex.php?id=14>. Temmuz 2006.
6. Kaya E, Bulun M, Arslan A. Tıpta Veri Ambarları Oluşturma ve Veri Madenciliği Uygulamaları. Selçuk Üniversitesi, Bilgisayar Mühendisliği Bölümü, Selçuk Üniversitesi, Selçuklu Tıp Fakültesi. ab.org.tr/ab03/tammetin/96.doc. 12.01.2012.
7. Azimli M. Veri Madenciliği Tıpta Veri Madenciliği Uygulamaları. Bilişim Enstitüsü Bilişim Sistemleri Yüksek Lisans (Uzak) Programı, 18 sayfa. (Yard. Doç. Dr. Suat Özdemir).
8. Morelli R. Using a Support Vector Machine to Analyze a DNA Microarray. Department of Computer Science, Trinity College, Hartford. 2007.
9. SVM - Support Vector Machines Introduction to Support Vector Machine (SVM) Models <http://www.dreg.com/svm.htm>. 23.02.2012.
10. Karagülle F. Destek Vektör Makinelerini Kullanarak Yüz Bulma. 2008, Trakya Üniversitesi Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı, Yüksek Lisans Tezi, 63 sayfa, Edirne. (Yrd. Doç. Dr. Rembiye Kandemir).
11. Lin TS, Ngo T. Clustering High Dimensional Data Using SVM. Department of Computer Science San José State University.
csc.csudh.edu/jhan/HD%20SVM.ppt. 21.12.2011.
12. Noble WS. What is a Support Vector Machine? Nature Biotechnology 2006;24(12):1564-67.

-
13. Avşar E. Tek-Sınıf Destek Vektör Makineleri Kullanılarak Epilektik EEG İşaretlerinin Sınıflandırılması. 2009, İstanbul Üniversitesi Fen Bilimleri Enstitüsü. Elektronik ve Haberleşme Mühendisliği Anabilim Dalı, Yüksek lisans Tezi, 69 sayfa, İstanbul. (Yrd. Doç. Dr. Mustafa E. KAMASAK).
 14. Guyon I, Weston J, Barnhill S, Vapnik V. Gene Selection for Cancer Classification using Support Vector Machine. *Machine Learning*. 2002;46:389–422.
 15. Moore AW. Support Vector Machines. School of Computer Science Carnegie Mellon University. www.cs.cmu.edu/~awm. 2001.
 16. Gunn SR. Support Vector Machines for Classification and Regression. 1998, University of Southampton, Faculty of Engineering, Science and Mathematics School of Electronics and Computer Science, Technical Report, 54 sayfa.
 17. Parçacık Sürü Optimizasyonu (Particle Swarm Optimization)
[http://www.yazilimutfagi.com/MakaleOku.aspx?Makale=Par%C3%A7acIk+S%C3%BCr%C3%BC+Optimizasyonu+\(Particle+Swarm+Optimization\)&MakaleKey=1a4ceabd-5ad7-4d8f-88e6-6feb15dd8f2d](http://www.yazilimutfagi.com/MakaleOku.aspx?Makale=Par%C3%A7acIk+S%C3%BCr%C3%BC+Optimizasyonu+(Particle+Swarm+Optimization)&MakaleKey=1a4ceabd-5ad7-4d8f-88e6-6feb15dd8f2d). 06.01.2012.
 18. Kaya H, Köymen K. Veri Madenciliği Kavramı ve Uygulama Alanları. *Doğu Anadolu Bölgesi Araştırmaları*. 2008; 159-164.
 19. Zhong N, Zhou L. Methodologies for Knowledge Discovery and Data Mining: Third Pacific-Asia Conference, Pakdd-99, Beijing, China, April 26-28, 1999.
 20. Özekes S. Veri Madenciliği Modelleri ve Uygulama Alanları. *İstanbul Ticaret Üniversitesi Dergisi*. 2003; 3(3):65-82.
 21. Coşkun C, Baykal A. Veri Madenciliğinde Sınıflandırma Algoritmalarının Bir Örnek Üzerinde Karşılaştırılması. ab.org.tr/ab11/bildiri/67.odt. 21.04.2011.
 22. Tekerek A. Veri Madenciliği Süreçleri ve Açık Kaynak Kodlu Veri Madenciliği Araçları. Gazi Üniversitesi, Elektronik-Bilgisayar Eğitimi Bölümü, Ankara. 20.11.2011.
 23. Ayre LB. Data Mining for Information Professionals. <http://www.psu.edu.tr>. 10.01.2012.
 24. Kıyak E. CRISP-DM Yöntembilimi Kullanılarak Deniz Kuvvetleri Verisi Üzerinde Veri Madenciliği Sınıflandırma Tekniklerinin Karşılaştırılması. 2006, Kocaeli Üniversitesi Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı, Yüksek Lisans Tezi, 123 sayfa, Kocaeli (Yrd. Doç. Dr. Nevcihan DURU).

-
25. Duda RO, Hart PE, Stork DG. Mitpress Mit Edu Books Chapters.
<http://mitpress.mit.edu/books/chapters/0262012111chap1.pdf>. 09.06.2011.
 26. Nilsson NJ. Introduction to Machine Learning An Early Draft of A Proposed Textbook. Robotics Laboratory Department of Computer Science Stanford University, Stanford. 1998.
 27. Vasif V, Nabiye V. Yapay Zeka-Problemler-Yöntemler-Algoritmalar. Seçkin Yayıncılık. Ankara. 2003: s. 566.
 28. Tolun S. Destek Vektör Makineleri: Banka Başarısızlığının Tahmini Üzerine Bir Uygulama. 2008, İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, İşletme Anabilim Dalı, Sayısal Yöntemler Bilim Dalı, Doktora Tezi, 265 sayfa, İstanbul. (Prof. Dr. Öner ESEN).
 29. Çomak E. Destek Vektör Makinelerinin Etkin Eğitimi İçin Yeni Yaklaşımlar. 2008, Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Elektrik Elektronik Mühendisliği Anabilim Dalı, Doktora Tezi, 125 sayfa, Konya. (Prof.Dr. Ahmet Arslan)
 30. Stoutamire D. Machine Learning, Game Play, and Go. Computer and Information Science. 1991; 2: 1-90.
 31. Kaygulu MS. Supervised and Unsupervised Learning Techniques in Data Mining. 1999, Dokuz Eylül Üniversitesi Fen bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, Yüksek Lisans Tezi.
 32. Öztemel E. Yapay Sinir Ağları. Papatya Yayıncılık. 2012: p: 21-25.
 33. Şahin B. Öğrenme Modelleri
http://www.yapayzeka.somee.com/makale_detay.aspx?id_no=13. 09.05.2008.
 34. Elmas Ç. Yapay Sinir Ağları Kuram, Mimari, Uygulama. Seçkin Yayıncılık. 1. Baskı. Ankara. 2003.
 35. Özkes S, Çamurcu AY. Veri Madenciliğinde Sınıflama ve Kestirim Uygulaması. Marmara Üniversitesi Fen Bilimleri Dergisi. 2002;18:1-17.
 36. Cristianini N Taylor JS. An Introduction to Support Vector Machines and other kernel-based training methods. Cambridge University Press, 2000.
 37. Veri Madenciliği. www.deu.edu.tr/userweb/k.yaralioglu/dosyalar/ver_mad.doc. 21.08.2011
 38. Kahramanlı H. Hibit Bulanık Sinir Ağını Kullanarak Bir Sınıflandırma ve Kural Çıkarma Sisteminin Geliştirilmesi. 2008, Selçuk Üniversitesi Fen Bilimleri

-
- Enstitüsü Matematik Anabilim Dalı, Doktora Tezi, 189 sayfa, Konya. (Prof. Dr. Novruz Allahverdi).
- 39 . Takçı H. Dördüncü Saat sınıflandırma: temel kavramlar, karar ağaçları ve model değerlendirme. Veri Madenciligi Dersi – GYTE – Dr. Hidayet Takçı. 10.05.2008.
 40. Akpınar, H. Veri Tabanlarında Bilgi Keşfi ve Veri Madenciligi. İstanbul Üniv. İşletme Fakültesi Dergisi. 2000;29:1.
 41. Akman M, Genç Y, Ankaralı H. Veri madencilğine genel bakış ve Random Forests yönteminin incelenmesi: Sağlık alanında bir uygulama. Türkiye Klinikleri J Biostat 2011;3(1):36-48.
 42. Yapay Sinir Ağları. <http://www.yildiz.edu.tr/~gulez/3k1n.pdf>. 22.06.2011
 43. Saraç T. Yapay Sinir Ağları, 2004, Seminer Projesi, Gazi Üniversitesi, Ankara.
 44. Karşılığil ME. Ağ Madenciligi Yöntemleriyle internet Kullanıcılarının Sınıflandırılması. 2006, Yıldız Teknik Üniversitesi Elektrik-Elektronik Fakültesi Bilgisayar Mühendisliği Bölümü Bilgisayar Projesi, İstanbul.
 45. Öğüdücü ŞG. Veri Madenciligi Farklı Sınıflandırma Yöntemleri. <http://www.ninova.itu.edu.tr/EgitimDetay.aspx?eId=195/>. 22.11.2011.
 46. Yıldırım Ü, Çataltepe Z. Örüntü Tanıma ve Öznitelik Seçme Yöntemleri Kullanarak Kısa Zaman Sonraki Yol Trafik Hız Öngörüsü http://eski.tbd.org.tr/resimler/ekler/aa48fc4880bb0c9_ek.pdf. 02.01.2012.
 47. Çapar A, Taşdemir K, Kılıç Ö, Gökmen M. Türkçe Elyazısı Tanıma Sistemlerinde Öznitelik Çıkarma ve Sınıflandırma Yöntemlerinin Karşılaştırması <http://www.ursi.org.tr/2002-1.Ulusal%20Kongre/ursicd1/C14.pdf>. 22.11.2011.
 48. Fukunaga K. Introduction to Statistical Pattern Recognition, 2. baskı., Boston: Academic Press, 1990.
 49. Grother PJ, Candela GT, Blue JL. Fast Implementations of Nearest Neighbor Classifiers. Pattern Recognition. 1997;30:459-465.
 50. Şeker SE. KNN (K nearest neighborhood, en yakın k komşu). <http://www.bilgisayarkavramlari.com/2008/11/17/knn-k-nearest-neighborhood-en-yakin-k-komsu/>. 04.07.2011.
 51. Emel GG, Taşkın Ç. Genetik Algoritmalar ve Uygulama Alanları. Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi. 2002;21(1):129-152.
 52. Bozüyük T, Gökçe İ, Yağcı C, Akar G. Yapay Zeka Teknolojilerinin Endüstrideki Uygulamaları. 2005, Marmara Üniversitesi Teknik Bilimler Meslek Yüksek

-
- Okulu Elektrik Programı, Bitirme Tezi, 67 sayfa, İstanbul. (Yrd.Doç.Dr.Mehmet Tektaş).
53. Aytekin MA, Kalaycı TE. Gezgin Satıcı Probleminin İkili Kodlanmış Genetik Algoritmalarla Çözümünde Yeni Bir Yaklaşım. http://en.wikipedia.org/wiki/Travelling_salesman_problem#Exact_algorithms. 09.10.2011.
54. Wang X, Zhang Y. Statistical Learning Theory and State of Art in SVM. Proceeding of the Second IEEE Internationale Conference on Cognitive Informatics. 2003;1
55. Bousquet O, Boucheron S, Lugosi G. Introduction to Statistical Learning Theory. http://www.econ.upf.edu/~lugosi/mlss_slc.pdf. 13.10.2011.
56. Hurtado JE. An Examination of Methods for Approximating Implicit from the Viewpoint of Statistical Learning Theory. Structural Safety. 2004;26(3):271-293.
57. Yıldırım S. Arıza Teşhisinde Destek Vektör Makinelerinin Kullanımı. 2006, Fırat Üniversitesi Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı, Yüksek Lisans Tezi, 113 sayfa, Elazığ. (Doç.Dr. Erhan Akın).
58. Bousquet O. Statistical Learning Theory. Machine Learning Summer School, August 2003.
59. Sravanthi AK. Edited-Bootstrapped Support Vector Machines For One-Class Data Classification. Nagarjuna University, 2006, Nagarjuna University, Bachelor of Technology in Electronics and Communication Engineering, Master of Thesis, India. (Dr. Guoliang Fan).
60. Cortes C, Vapnik V. Support Vector Networks. Machine Learning, 1995;20:273-297.
61. DiMona J. Statistical Learning Theory: Classification Using Support Vector Machines. <http://www.cs.cmu.edu/~awm/tutorials>. 07.02.2012.
62. Yu PS, Chen ST, Chang IF. Support Vector Regression for Real-Time Flood Stage Forecasting. Journal of Hydrology, 2006; 328(1): 704-706.
63. Vapnik VN. An Overview of Statistical Learning Theory. IEEE Transactions On Neural Networks, 1999;10(5): 988-999.
64. Kecman V. Learning and Soft Computing. Asco Typesetters. USA. 2001: p. 145-148.

-
65. Çomak E. Destek Vektör Makineleri Çoklu Sınıf Problemleri İçin Çözüm Önerileri. 2004, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı, Yüksek Lisans Tezi, Konya. (Doç. Dr. Ahmet Arslan).
 66. KXEN Analytic Framework
www.ie.metu.edu.tr/~tayyar/Downloads/KXEN_Egitim_TR.pdf. 01-11-2011
 67. David SB. ECE 695 “Statistical Learning Theory” Lecture Notes Lecture 4. Cornell University, 2003.
 68. Burges CJC. Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery. 1998;2:121–167.
 69. Structural Risk Minimization. <http://www.svms.org/srm/>. 22.02.2012.
 70. Güner ES. Türkçe İçin Derlem Tabanlı Bir Anafor Çözümleme Çalışması. 2008, Trakya Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, Yüksek Lisans Tezi, 74 sayfa, Edirne (Yrd. Doç. Dr. Yılmaz KILIÇASLAN)
 71. Modellerde Fazla Kapsama. bilmuh.gyte.edu.tr/~htakci/vm/Model_Overfitting.doc. 19.11.2011.
 72. Schölkopf B, Smola AJ. Learning with Kernels. MIT Press. 2002.
 73. Vapnik, VN. Chervonenkis AY. Theory of Pattern Recognition: Statistical Problems of Learning. Moscow: Nauka, 1974.
 74. Vapnik VN. The Nature of Statistical Learning Theory Second Edition. Springer Verlag New York. 1999;94.
 75. Sewell M. Structural Risk Minimization. Department of Computer Science University College London, 2008.
 76. Trafalis B, İnce H. Support Vector Machine for Regression and Applications to Financial Forecasting. Neural Networks. Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks, 2000;6:348-353.
 77. Alpaydın E. Introduction To Machine Learning. MIT Press, Massachusetts, 2004.p. 22-23.
 78. Cox PG, Adhami R. Multi-class support vector machine classifier applied to hyperspectral data. Proceedings of the Thirty-Fourth Southeastern Symposium on System Theory. IEEE, 2002. p.271-274.
 79. Vapnik V. Support Vector Machines.
<http://web.cecs.pdx.edu/~mm/AIFall2011/SVMs.pdf>. 1979 (21.09.2011).

-
80. Demirci DA. Destek Vektör Makineleri ile Karakter Tanıma. 2007, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, Matematik Mühendisliği Anabilim Dalı, Yüksek Lisans Tezi, 43 sayfa, İstanbul. (Yrd. Doç. Dr. Nilgün Güler Bayazıt).
 81. Wang L. Support Vector Machines: Theory and Application, Studies in Fuzziness and Soft Computing, 2005; Volume 177, Springer.
 82. SVM - Support Vector Machines. <http://www.dtrek.com/svm.htm>. 20.09.2011.
 83. Walgampaya CK. Cost-Benefit Analysis in Multiple Time Series Prediction. 2006, University of Louisville, Department of Computer Engineering and Computer Science, Master of Science Thesis. Kentucky.
 84. Kavzoğlu T, Çölkesen İ. Destek Vektör Makineleri ile Uydu Görüntülerinin Sınıflandırılmasında Kernel Fonksiyonlarının Etkilerinin İncelenmesi. Harita Dergisi. 2010;7: 144.
 85. Mukherjee S. Classifying Microarray Data Using Support Vector Machines. Whitehead Institute for Genome Research and Center for Biological and Computational Learning at MIT. 2011.
 86. Venture D. SVM Example. <http://www.dainf.ct.utfpr.edu.br/~kaestner/Mineracao/SVM.example.pdf>. 12.03.2009.
 87. Mavroforakis ME, Theodoridis S. A Geometric Approach to Support Vector Machine (SVM) Classification. IEEE Transactions on Neural Networks. 2006;17(3).
 88. Tan Y, Wang J. A Support Vector Machine With A Hybrid Kernel and Minimal Vapnik-Chervonenkis Dimension. IEEE Transactions on Knowledge and Data Engineering, IEEE Educational Activities Department. USA, April 2004;16(4):389.
 89. Sullivan K, Luke S. Evolving Kernels for Support Vector Machine Classification. GECCO, 2007. <http://dl.acm.org/citation.cfm?id=1277292>. 10.10.2011.
 90. Demir B, Ertürk S. Hiperspektral Görüntülerde Destek Vektör Makinaları ile Sınıflandırma Öncesi Dalgacık Uzayında Gürültü Giderimi. http://kulis.kocaeli.edu.tr/pub/siu07_svm_noise_red.pdf. 21.01.2012.
 91. Erastö P. Support Vector Machines - Backgrounds and Practice. Rolf Nevanlinna Institute, Helsinki 2011. <http://ethesis.helsinki.fi/julkaisut/mat/rolfn/lt/erasto/supportv.pdf>. 21.01.2012.

-
92. Özbeyaz A, Gürsoy Mİ, Çoban R. EEG İşaretlerinin DVM İle Sınıflandırılmasında Ayarlama Ve Çekirdek Parametrelerinin PSO Algoritmasına Dayalı Optimizasyonu. 2011 IEEE 19th Signal Processing and Communications Applications Conference, 2011.
 93. Tan M. Support Vector Machine and Its Applications. www.iro.umontreal.ca/.../svm_tutorial.ppt. 22.10.2011.
 94. Liu TY. SVM and Its Applications to Text Classification <http://www.infoautoclassification.org>. 23.03.2006.
 95. Alıç E. Optimizasyona Dayalı Sınıflandırma Modelleri. <http://emrealic.wordpress.com/tag/optimizasyona-dayali-siniflandirma-modelleri/>. 22.10.2011.
 96. Lin HT, Lin CJ. A Study on Sigmoid Kernels for SVM and the Training of non-PSD Kernels by SMO-type Methods. Department of Computer Science and Information Engineering National Taiwan University Taipei 106, Taiwan. 2003.
 97. Lee B, Kim S, Seok J, Won S. Nonlinear System Identification based on Support Vector Machine using Particle Swarm Optimization. SICE-ICASE International Joint Conference, Bexco, Busan, Korea. 18-21.10.2006
 98. Kwok JT, Tsang IW. Linear Dependency Between and the Input Noise in Support Vector Regression. IEEE Transactions on Neural Networks. 2003;14:3.
 99. Smola AJ, Murata N, Schölkopf B, Müller KR. Asymptotically Optimal Choice of Loss for Support Vector Machines. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.40.5941>. 09.10.2011.
 100. Cherkassky V, Ma Y. Comparison of Model Selection for Regression. Neural Computation. 2003;15:1691–1714.
 101. Law M. A Simple Introduction to Support Vector Machines. http://www.cise.ufl.edu/class/cis4930sp11dtm/notes/intro_svm_new.pdf. 13.09.2011.
 102. Basak D, Pal S, Patranabis DC. Support Vector Regression. Neural Information Processing – Letters and Reviews. 2007;11(10):203-224.
 103. Parrella F. Online Support Vector Regression. 2007, University of Genoa, Department of Information Science, A Thesis presented for the degree of Information Science, 92 pages, Italy.

-
104. Cherkassky V, Ma Y. Practical Selection of SVM Parameters and Noise Estimation for SVM Regression. *Journal Neural Networks*. 2004;17(1):113-126.
 105. Marshall HM, Allison KC, O'Reardon JP, Birketvedt G, Stunkard AJ. Night eating syndrome among nonobese persons. *Int J Eat Disord*. 2004;35:217–222.
 106. Cooper T, Fairburn CG. The Development and Validation of the Body Shape Questionnaire. *International Journal of Eating Disorders*. 1986;6:485-494.
 107. Auria L, Moro RA. Support Vector Machines (SVM) as a Technique for Solvency Analysis. *Deutsches Institut für Wirtschaftsforschung Berlin. Discussion Papers* 811. 2008;8.
 108. Meyer D. Support Vector Machines. *R-News*. 2001;1(3).
 109. Pochet NLMM, Suykens JAK. Support vector machines versus logistic regression: improving prospective performance in clinical decision-making. *Ultrasound Obstet Gynecol*. 2006;27:607–608.
 110. Puniyani K. ML Classroom: Classification I Logistic Regression, SVMs. <http://www.slideshare.net/NYCPredictiveAnalytics/intro-to-classification-logistic-regression-svm#>. 13.12.2011.
 111. Zhou XH, Li SL, Tian F, Xie Y, Pei MY, Kang S, et al. Building a disease risk model of osteoporosis based on traditional Chinese medicine symptoms and western medicine risk factors. *Statistics in Medicine* 2012; doi: 10.1002/sim.4382.
 112. Fitzgerald AJ, Pinder S, Purushotham AD, O'Kelly P, Ashworth PC, Wallace VP. Classification of terahertz-pulsed imaging data from excised breast tissue. *Journal of Biomed Optics* 2012; 17(1): 5-11.
 113. Duttagupta R, Dirienzo S, Jiang R, Bowers J, Gollub J, Kao J, et al. Genome-Wide Maps of Circulating miRNA Biomarkers for Ulcerative Colitis. *PLoS One* 2012; 7(2): e31241. 1-13.
 114. Yi Z, Li Z, Yu S, Yuan C, Hong W, Wang Z, et al. Blood-based gene expression profiles models for classification of subsyndromal symptomatic depression and major depressive disorder. *PLoS One* 2012; 7(2): e31283 1-11.
 115. Gallant AR, Lundgren J, Drapeau V. The night-eating syndrome and obesity. *Obesity Reviews* 2012; doi: 10.1111/j.1467-789X.2011.00975.x.
 116. Jacobi C, Fittig E, Bryson SW, Wilfley D, Kraemer HC, Taylor CB. Who is really at risk? Identifying risk factors for subthreshold and full syndrome eating disorders in a high-risk sample. *Psychological Medicine* 2011; 41(9): 1939-49.

ÖZGEÇMİŞ

1985 yılında Akçakoca'da doğdu. İlkokul, ortaokul ve lise öğrenimini Ankara'da tamamladıktan sonra, 2004 yılında Gazi Üniversitesi Fen-Edebiyat Fakültesi İstatistik Bölümü'nü kazandı. 2008 yılında bu bölümden mezun oldu ve 2009 yılında Ankara Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Anabilim Dalı'nda yüksek lisans eğitimine başladı. Daha sonra 2010 yılında Düzce Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Anabilim Dalı'nda Araştırma Görevlisi olarak göreve başladı ve yüksek lisans eğitimine de aynı üniversitede devam etti. Hala Düzce Üniversitesi'nde araştırma görevlisi olarak görevini sürdürmektedir.

Tel: 0532 510 37 22

e-posta: ozgeyilmaz@duzce.edu.tr
ozge_yilmaz85@hotmail.com