



**T. C.
ULUDAĞ ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
BİYOİSTATİSTİK ANABİLİM DALI**

**SINIFLANDIRMA PROBLEMİNİN ÇÖZÜMLENMESİNDE
YAPAY SINIR AĞLARI İLE DİSKRİMİNANT ANALİZİNİN
KARŞILAŞTIRILMASI VE BİR UYGULAMA**

Deniz SİĞİRLİ

(YÜKSEK LİSANS TEZİ)

Bursa-2006



T. C.
ULUDAĞ ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
BİYOİSTATİSTİK ANABİLİM DALI

SINIFLANDIRMA PROBLEMİNİN ÇÖZÜMLENMESİNDE
YAPAY SINIR AĞLARI İLE DİSKRİMİNANT ANALİZİNİN
KARŞILAŞTIRILMASI VE BİR UYGULAMA

Deniz SİĞİRLİ

(YÜKSEK LİSANS TEZİ)

Danışman: Yrd. Doç. Dr. Bülent EDİZ

Bursa-2006

İÇİNDEKİLER

TÜRKÇE ÖZET.....	II
İNGİLİZCE ÖZET.....	III
GİRİŞ.....	1
GENEL BİLGİLER.....	3
1. Yapay Sinir Ağları.....	3
2. Diskriminant Analizi.....	25
GEREÇ VE YÖNTEM.....	42
BULGULAR.....	44
1. Diskriminant Analizi Sonuçları.....	45
2. Yapay Sinir Ağı Modeli ile Sınıflandırma Sonuçları.....	49
TARTIŞMA VE SONUÇ.....	52
KAYNAKLAR.....	55
TEŞEKKÜR.....	58
ÖZGEÇMİŞ.....	59

ÖZET

Bu çalışmada, yapay sinir ağı modellerinin açıklanması, diskriminant analizi yöntemi ile bazı yapay sinir ağı modelleri arasındaki ilişkilerin incelenmesi ve bir uygulama üzerinde yapay sinir ağı modelleri ile diskriminant analizi yönteminin doğru sınıflandırma performanslarının karşılaştırılması amaçlanmıştır.

Uludağ Üniversitesi Tıp Fakültesi Uygulama ve Araştırma Hastanesi Gastroenteroloji servisine başvuran ve yapılan tetkik-incelemeler ile kronik karaciğer parenkim hastalığı düşünülerek karaciğer biyopsisi uygulanması sonucunda Hepatit B ve C virüslerine bağlı kronik hepatit veya siroz tanısı alan hastalara ilişkin geriye dönük olarak toplanmış olan veri setine diskriminant analizi ve yapay sinir ağları uygulanmıştır.

Yapılan analizler sonucunda, diskriminant analizi için elde edilen modelin genel doğruluk yüzdesi % 93,94 olarak, yapay sinir ağları tekniği için ise % 100 olarak hesaplanmıştır. Çalışmaya alınan veri seti için yapay sinir ağlarının diskriminant analizine göre daha yüksek “doğru sınıflandırma oranına” sahip olduğu görülmüştür.

Anahtar Kelimeler: Diskriminant analizi, yapay sinir ağları, doğru sınıflandırma oranı.

SUMMARY

COMPARISON OF ARTIFICIAL NEURAL NETWORKS AND DISCRIMINANT ANALYSIS IN SOLVING CLASSIFICATION PROBLEM AND AN APPLICATION

In this study, it is aimed to describe the artificial neural network models, to examine the relationships between the discriminant analysis method and artificial neural networks and to compare the classification effectivities of discriminant analysis and artificial neural networks on a real data set.

Discriminant analysis and artificial neural networks have been applied to a data set, which have been collected retrospectively from the patients that had been administered to the Uludag University Medical Faculty Gastroentorolgy service and had been diagnosed with chronic hepatitis related to hepatitis B and C viruses, according to the result of biopsy application which have been applied in contemplation of chronic liver parenchymal disease after the axaminations.

As a result of the analysis, correct classification ratio of the model obtained from the discriminant analysis and artificial neural network was calculated respectively as 93.94% and 100%. It have been found that, artificial neural networks has got greater “true classification rate” than the discriminant analysis for this data set.

Key words: Discriminant analysis, artificial neural networks, correct classification ratio.

GİRİŞ

Bilimin temel yöntem bilimlerinden bir tanesi olan sınıflandırma, insanoğlu tarafından üstlenilen en eski bilimsel uğraşlardan biridir. Nesnelerin sınıflandırılması belki de tüm bilimsel çabaların temelini oluşturan bir çalışmadır.

Sınıflandırma problemi üzerinde pek çok yaklaşım bulunmaktadır. Sınıflandırmayla ilgili üç ana tarihsel kol, istatistiksel yöntemler, makine öğrenme yöntemleri ve yapay sinir ağları olarak tanımlanabilir. Bu üç kol üzerinde farklı profesyonel ve akademik gruplar tarafından durulmuş ve farklı konular üzerinde yoğunlaşmıştır (1).

İstatistiksel yöntemler, makine öğrenimi ve yapay sinir ağlarının ortak amaçları olmakla birlikte, kullandıkları sınıflandırma teknikleri farklıdır. İstatistikte kullanılan sınıflandırma teknikleri klasik ve modern teknikler olmak üzere iki ana başlıkta incelenebilirler. Klasik teknikler; Fisher tarafından ortaya atılan ve doğrusal diskriminant kurallarına dayanan teknikler olurken, modern tekniklerin en eskisi Fix ve Hodges tarafından geliştirilen ve parametrik olmayan bir teknik olan yoğunluk tahminidir (1, 2).

Diğer modern teknikler; k-en yakın komşu sınıflandırma, projeksiyon izleme (projection pursuit) sınıflandırma, saf bayes (naive bayes) sınıflandırma ve olasılıksal sinir ağlarıdır. Ayrıca çok değişkenli toplamsal regresyon eğrileri (Multivariate Adaptive Regression Splines-MARS) ve değişimli koşullu beklenti (The Alternating Conditional Expectation-ACE) algoritmaları, istatistikçiler tarafından geliştirilmiş sınıflandırma algoritmalarıdır (3).

Özellikle tıp ve biyoloji alanında yapılan çalışmalarda, incelenen olaylar çoğunlukla birden fazla değişkenin etkisi altında kalmakta, dolayısıyla incelenen olayı etkileyen birden çok değişkenin birlikte ele alınması gerekmektedir. Bu durum, çok değişkenli istatistiksel çözümlerinin kullanımını kaçınılmaz kılmaktadır (4).

Çok değişkenli istatistiksel teknikler, Fisher'in doğrusal diskriminant analizi üzerindeki çalışmalarına kadar uzanmaktadır. Diskriminant analizi, araştırmacılar ve uygulamacılar tarafından sınıflandırma modelleri geliştirmede kullanılan ilk çok değişkenli istatistiksel sınıflandırma yöntemi ve en yaygın olarak kullanılan metodoloji olmuştur. Fisher'in çalışmasından yaklaşık olarak on yıl sonra Smith, doğrusal diskriminant analizinden, karesel diskriminant analizini elde etmiştir (5-7).

Sınıflandırma üzerindeki diğer bir yaklaşım olan yapay sinir ağları (YSA) ile bilim dünyası 1940'lı yıllarda tanıştı. Bu alanda yapılan ilk çalışmalar beyin hücrelerinin işlevlerinin ve birbirleri ile haberleşme şekillerinin ortaya çıkarılmasını amaçlamaktaydı

(8). Beynin bilgi işleme yöntemine benzer biçimde YSA, bir öğrenme sürecinden sonra bilgiyi toplama, hücreler arasındaki bağlantılarda bulunan ağırlıklar ile bu bilgiyi saklama ve genelleme yapma yani bu bilgiyi ileriki zamanlarda da kullanabilme yeteneğine sahip paralel dağılmış bir işlemcidir (9). Öğrenme süreci, arzu edilen amaca ulaşmak için YSA ağırlıklarının yenilenmesini sağlayan öğrenme algoritmalarını ihtiva eder.

Yapay sinir ağları bilgi sınıflama ve bilgi yorumlamanın da içinde bulunduğu çok değişik problemlerin çözümünde kullanılmasının yanı sıra sağlık bilimleri, finans, endüstri, eğitim ve benzeri alanlarda başarıyla uygulanmaktadır (10).

Sınıflandırma problemlerinde, sinir ağının görevi tüm giriş örneklerinin belirli bir sınıfa ait olup olmadığını belirlemektir (11).

1990'lı yılların başlangıcında yayınlanan bazı kaynaklarda, bazı yapay sinir ağı modelleri ile bazı istatistiksel tekniklerin benzer hatta bazılarının aynı olduğuna dikkat çekilmiştir. Sonraki çalışmalar ise bunun tesadüfi olmadığını, bu iki alanın birbirleriyle ileri derecede ilişkili olduğunu göstermiştir. Perseptron, çok katmanlı perseptron gibi bazı YSA modellerinin istatistiksel uygulamalar için faydalı olabileceği ve aynı şekilde tahmin ölçütü, güven aralıkları, tanımlayıcı yöntemler gibi bazı istatistiksel tekniklerin de yapay sinir ağı uygulamalarına uygulanabileceği bazı bilim adamları tarafından ileri sürülmektedir (12).

Örnek sınıflandırma ve tanıma problemlerinde istatistiksel teknikler ve YSA modellerinin kullanımı için uygun teknik ve modellerin karşılaştırılması gerekmektedir. Bu bakımdan diskriminant analizi, lojistik regresyon ve benzeri istatistiksel tekniklere uygun yapay sinir ağı modellerinin belirlenmesi, onların avantajları ve dezavantajlarının açıklanması büyük önem taşımaktadır (13).

Bu çalışmanın amacı, yapay sinir ağı modellerinin açıklanması, diskriminant analizi yöntemi ile bazı yapay sinir ağı modelleri arasındaki ilişkilerin incelenmesi ve bir uygulama üzerinde yapay sinir ağı modelleri ile diskriminant analizi yönteminin sınıflandırma performanslarının incelenmesidir.

GENEL BİLGİLER

1. Yapay Sinir Ağları

Yapay zeka dallarından biri olan yapay sinir ağları, günümüzde sadece yapay zeka çalışmalarında değil, bir çok bilim dalında direkt olarak kullanılmaktadır (1).

Son zamanlarda neredeyse tüm sınıflandırma problemleri için tercih edilen bir çözüm olarak dikkat çekmekte olan yapay sinir ağları (14), beynin çalışma ilkelerinin bilgisayarlar üzerinde taklit edilmesi fikri ile ortaya çıkmış ve ilk çalışmalar beyni oluşturan biyolojik hücrelerin, yani nöronların matematiksel olarak modellenmesi üzerinde yoğunlaşmıştır (8).

Bir yapay sinir ağı, birbirlerine ağırlıklı bağlantılarla bağlanmış işlem birimlerinden oluşan, biyolojik sinir ağlarının karakteristiklerine benzer karakteristiklere sahip bir bilgi işleme sistemidir (11).

YSA ile ilgili ilk çalışmalar 1940' lı yıllarda Warren McCulloch ve Walter Pitts tarafından McCulloch-Pitts nöronlarının oluşturulmasıyla başlamıştır. Bu modelde, herhangi bir çıkışı üretmek için, sinir ağına yerleştirilen nöronlar mantıksal fonksiyonları birleştirerek yapılandırılmıştır (9). Donald Hebb 1949 yılında, günümüzde yapay sinir ağları için Hebb kuralı olarak bilinen ilk öğrenme yasasını tasarlamıştır (15). 1958 yılında Block, Minsky ve Papert ve 1962 yılında Frank Rosenblatt, perseptronlar adı verilen çok geniş bir yapay sinir ağı sınıfını tanıtmışlardır. Perseptron öğrenme kuralı, Bernard Widrow ve onun öğrencisi olan Ted Hoff, çıktı birimi ile ağ girdisi arasındaki farkı azaltmak için ağırlıkları değiştirerek, çok katmanlı ağlar için geri beslemenin habercisi olacak olan Widrow-Hoff (delta) öğrenme kuralını geliştirdiler. Teuvo Kohonen, 1972 yılında yapmış olduğu ilk çalışmasında, birleştirici hafızalı sinir ağlarıyla ilgilenmiştir. On yıl sonraki çalışması, kendiliğinden organize olabilen nitelik haritalarının gelişimiydi ve bu haritalar, kümelenmiş birimler için topolojik bir yapıyı kullanmıştır. Sinir ağlarının gelişiminde katkısı olan önemli bilim adamlarından birisi de John Hopfield'dir. David Tank ile birlikte Hopfield, uyarlanabilir aktivasyonlar ve sabit ağırlıklara dayalı günümüzde Hopfield ağları olarak bilinen ağ modelini geliştirmişlerdir (16).

Yapay sinir ağları biyolojik sinir ağlarının modellenmesiyle oluşmuşlardır. İnsanoğlunda öğrenme işlemi, beyindeki her bir nöronun diğer nöronlara hızlandırıcı (heyecanlandırıcı) ya da yavaşlatıcı (engelleyci) sinyaller gönderdiği milyonlarca nöronun etkileşimi sürecinde gerçekleşir. Beyindeki her bir nöron, hücre gövdesi (soma), bir ya da

daha çok dentrit ve akson olmak üzere üç tip bileşenden meydana gelmektedir (17). Sinaps adı verilen bağlantılarla bir başka nöron ile iletişim kurulur.

1.1. Yapay Sinir Ağı Modelinin Temel Özellikleri

Yapay sinir ağları insanın idrak etmesi ve biyolojik nöron yapısının matematiksel modelinin aşağıdaki kurallar varsayılarak genelleştirilmesi sonucunda oluşturulmuştur:

- Bilgi işleme nöron (işlem elemanı) adı verilen birimlerde gerçekleşir.
- Sinyaller, bir nörondan diğerine bağlantılar aracılığıyla iletilir.
- Her bir bağlantının, gönderilen sinyal ile çarpılan bir ağırlık değeri vardır.
- Her bir nöron, bir çıktı sinyali elde etmek için ağ girdisine (ağırlıklarla çarpılmış sinyallerin toplamı) bir aktivasyon fonksiyonu uygular. Bu, genellikle doğrusal olmayan bir fonksiyondur (11).

Yapay sinir ağlarında, bir girdi dizisi ağa girdikten sonra, birinci katmandaki her bir nöron girdi dizisinin bir elemanını alır. Her nöron, analiz etmek istediğimiz ya da bir tahmin edici olarak kullanmak istediğimiz bir numunenin bir özelliğine ya da bir karakteristiğine kodlanır veya karşılık gelir. Birimler katmanlar biçiminde organize olurlar. Katmandaki diğer nöronlarla paralel bir şekilde ağırlıklarla çarpılan girdiler aktivasyon fonksiyonunda işlenir ve son katmandaki nöronlara bir tek çıktı olarak iletilir. Elde edilen sonuç, girdinin niteliklerini temsil eden bir çıktı dizisidir. Zamanla girdiler ve uyarlanabilir ağırlıklar değiştiği için ağ buna adapte olur ve öğrenir (9).

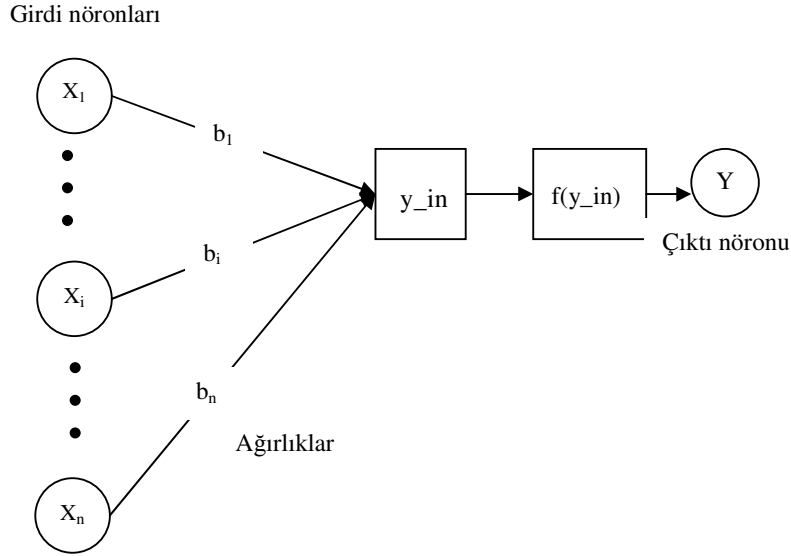
Herhangi bir yapay sinir ağının üç temel özelliği nöronlar, ağ mimarisi ve öğrenme algoritması ya da eğitim kurallarıdır (18).

1.2. Nöron Modeli

Hesaplama elemanları, düğüm, birim, işlem elemanı, yapay sinir hücresi olarak da ifade edilebilen nöronlar, bir ağın temel elemanlarıdır.

Her nöronun bir iç durumu vardır ve bu iç durum aktivasyon ya da aktivasyon seviyesi olarak adlandırılır. Bu seviye, alınan giriş değerlerinin bir fonksiyonudur. Herhangi bir nöron, kendi aktivasyonunu, genelde sinyal şeklinde diğer nöronlara gönderir. Ama bu sinyal birden fazla nörona aynı anda gönderilebilir (13).

Bulunduğu katmana göre bir nöron girdi nöronu (girdi birimi), çıktı nöronu (çıkı birimi) ve gizli nöron (gizli katman birimi) olarak adlandırılmaktadır. Bir girdi nöronu, istatistiksel modeldeki bağımsız değişkene karşılık gelmektedir (17).



Şekil-1: YSA Nöronunun Yapısı

Şekil-1’de, $X_1, X_2, \dots, X_i, \dots, X_n$ girdi nöronlarını, Y ise çıktı nöronunu göstermektedir. Bu nöronların aktivasyonları yani çıkış sinyalleri, sırasıyla $x_1, x_2, \dots, x_i, \dots, x_n$ ’ dir. Bağlantılar üzerindeki ağırlıklar $X_1, \dots, X_i, \dots, X_n$ nöronlarından Y nöronuna doğru sırasıyla $b_1, \dots, b_i, \dots, b_n$ ’ dir. y_{in} ile gösterilen net ağırlıklı girdi değeri, $X_1, \dots, X_i, \dots, X_n$ ’den Y ’ye giden ağırlıklı sinyallerin toplamıdır. Yani;

$$y_{in} = b_1x_1 + b_2x_2 + \dots + b_nx_n \quad [1]$$

biçiminde elde edilir. $f(.)$ aktivasyon fonksiyonu olmak üzere, y_{in} değerine aktivasyon fonksiyonunun uygulanmasıyla, Y nöronunun aktivasyonu olan y değeri aşağıdaki eşitlikte görüldüğü gibi elde edilir (8, 19).

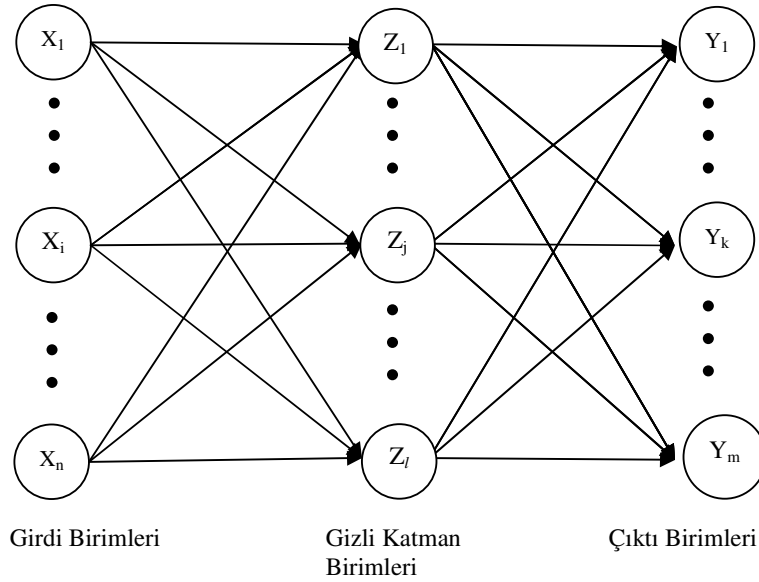
$$y = f(y_{in}) \quad [2]$$

1.3. Ağ Mimarisi

Nöronların katmanlar içindeki yerleşimleri ve diğer katmanlardaki nöronlarla olan bağlanma şekilleri ağ mimarisi olarak adlandırılmaktadır. Ağ yapısı katmanlar şeklinde yerleşmiş nöronlardan oluşmaktadır. Nöronlar girdileri diğer işlem elemanlarından gelen heyecanlandırıcı ya da engelleyici ağırlıklı bağlantılar yoluyla alırlar. Heyecanlandırıcı ağırlıklar genellikle pozitif, engelleyici ağırlıklar ise negatif değere sahiptir. Bütün sistemin davranışını belirleyen temel faktörler aktivasyon fonksiyonları ve sinyallerin gönderildiği bağlantılar üzerindeki ağırlıklardır (19).

Yapay sinir ağında ilk katman girdi katmanıdır ve dışarıdan gelen verilerin yapay sinir ağına alınmasını sağlar. Son katman ise bilgilerin dışarıya iletiildiği çıktı katmanıdır. Girdi ve çıktı katmanlarının arasında bulunabilecek bir ya da daha fazla sayıdaki katmana ise gizli katman adı verilir. Gizli katmanlar giriş uzayını keyfi bölgelere ayırarak, karmaşık problemlerin çözümünde gerekli gücü sağlar (17). Bir yapay sinir ağında gizli katman olması gerekmediği gibi, birden fazla gizli katman da bulunabilir (9).

Şekil-2’de bir gizli katmana, n adet girdi birimine, l adet gizli katman birimine ve m adet çıktı birimine sahip bir YSA gösterilmektedir.

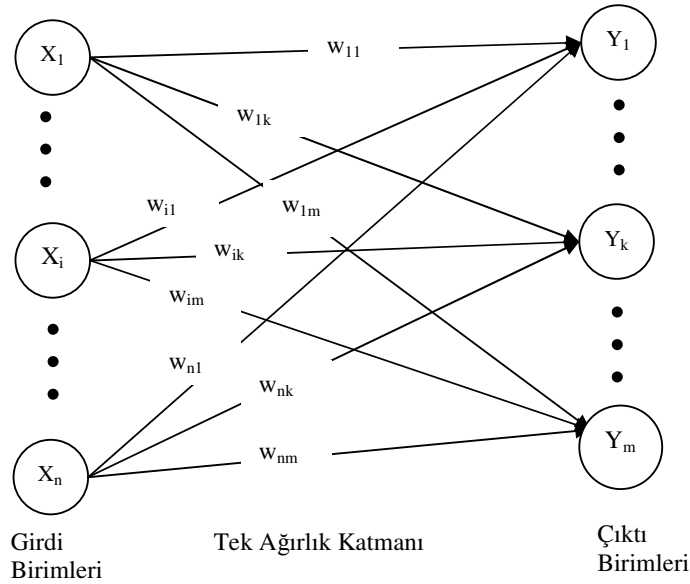


Şekil-2: Genel bir yapay sinir ağı modeli

Yapay sinir ağı, genel olarak tek katmanlı veya çok katmanlı yapay sinir ağı olarak sınıflandırılırlar. Katman sayısını belirlerken, girdi katmanı sayılmaz, çünkü bu katman üzerinde hiçbir hesaplama işlemi yoktur. Bir ağ içindeki katman sayısı nöronları bağlayan ağırlıklı bağlantıların katman sayısına eşittir (11).

1.3.1. Tek Katmanlı Yapay Sinir Ağı

Tek katmanlı yapay sinir ağlarında bir tane ağırlıklı bağlantı katmanı bulunur. Çoğu kez birimler, sinyalleri alan girdi birimi ve ağın cevabının alınacağı çıktı birimi olmak üzere ikiye ayrılmaktadır. Tipik bir tek katmanlı YSA Şekil-3'de verilmiştir.



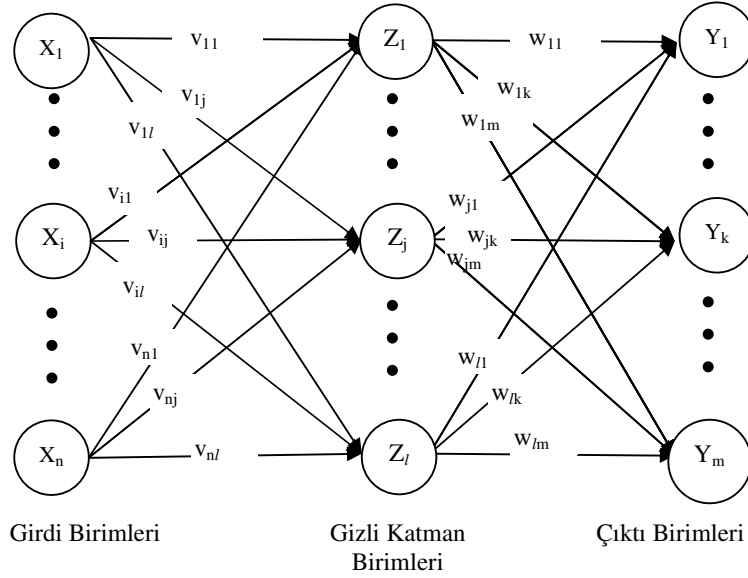
Şekil-3: Tek katmanlı bir yapay sinir ağı

Tek katmanlı yapay sinir ağı için bir çıkış birimini etkileyen ağırlıklar başka bir çıkış birimini etkilemezler. Sinir ağından alınacak cevabın yorumuna göre, aynı mimarili ağ farklı problemlerin çözümü için kullanılabilir.

1.3.2. Çok Katmanlı Yapay Sinir Ağları

Çok katmanlı yapay sinir ağlarında, girdi birimleri ile gizli birimlerin arasında bir ya da birden fazla katman bulunmaktadır. Genellikle, gizli ve çıktı birimleri arasında ağırlıklı bağlantı katmanı bulunan bu tip ağ mimarileri tek katmanlı ağ mimarilerine göre daha

karmaşık problemleri çözebilir. Eğitilmeleri çok daha zor olabilen çok katmanlı ağlar, bazı tek katmanlı ağların problemi tam olarak çözecek şekilde eğitilemediği durumlarda son derece başarılı olabilirler. Tipik birçok katmanlı ağ Şekil-4’de gösterilmiştir.



Şekil-4: İki katmanlı bir yapay sinir ağı

1.4. Öğrenme Kuralları

Yapay sinir ağlarında bilgi, ağdaki bağlantıların ağırlıklarında depolanır. Öğrenme, sistemin bir bütün olarak istenilen işlevi yerine getirecek şekilde ağırlıklarının ayarlanması sürecidir. YSA’da toplam ağ hatası istenilen düzeye erişinceye kadar eğitim devam eder (13).

Ağırlıkların ayarlanmasında genel olarak danışmanlı ve danışmansız olmak üzere iki tür öğrenme çeşidinden söz edilebilir.

1.4.1. Danışmanlı Öğrenme

Bu tip öğrenmede, yapay sinir ağlarına örnek olarak bir doğru çıktı verilir (19). İstenilen ve gerçek çıktı arasındaki farka (hataya) göre nöronlar arasındaki bağlantıların ağırlıkları, en uygun çıktıyı elde etmek için sonradan düzenlenebilir. Bu sebeple danışmanlı öğrenme algoritmasının bir “danışmana (öğretmene)” ihtiyacı vardır. Widrow-Hoff tarafından geliştirilen delta kuralı ve Rumelhart ve McClelland tarafından geliştirilen

genelleştirilmiş delta kuralı ve geriye beslemeli öğrenme algoritması danışmanlı öğrenme algoritmalarına örnek olarak verilebilir.

1.4.2. Danışmansız Öğrenme

Danışmansız öğrenmede ağ, girdi olarak verilen örnekten elde edilen çıktı bilgisine göre sınıflandırmayı kendi kendine geliştirmektedir. Bu öğrenme algoritmalarında, hiç bir hedef vektörü verilmez. Öğrenme sürecinde sadece giriş bilgileri verilir. Ağ daha sonra bağlantı ağırlıklarını aynı özellikleri gösteren desenler oluşturmak üzere ayarlar. Kohonen'in kendiliğinden organize olabilen haritaları (Kohonen's self-organizing maps) ve adaptif rezonans teorisi (adaptive resonance theory) danışmansız öğrenmeye örnek olarak verilebilir (20).

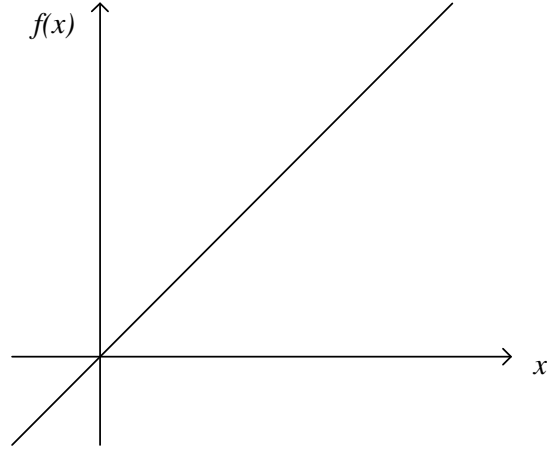
1.5. Aktivasyon Fonksiyonları

Daha önceden de bahsedildiği gibi, bir yapay sinir ağında yapılması gereken en temel işlemler, ağırlıklı girdi değerlerinin toplanması ve bir aktivasyon fonksiyonu uygulanmasıdır.

Yapay sinir ağı uygulamalarında yaygın olarak kullanılan bazı aktivasyon fonksiyonları aşağıda verilmiştir.

1.5.1. Özdeşlik Fonksiyonu

Genellikle girdi değerleri için kullanılan aktivasyon fonksiyonu Şekil-5'de gösterilen ve tüm x 'ler için $f(x) = x$ olan özdeşlik fonksiyonudur.



Şekil-5: Özdeşlik fonksiyonu

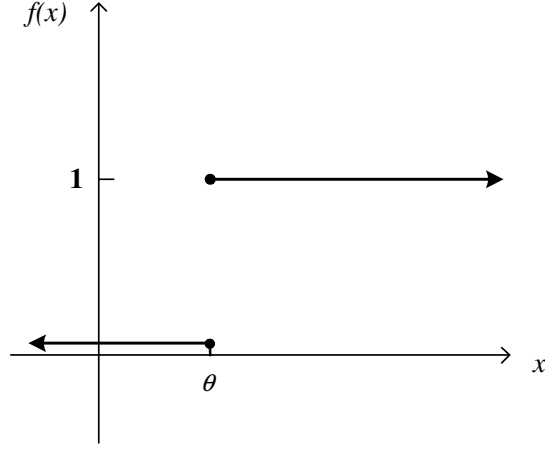
Belirli bir tabakadaki tüm nöronlar için aynı aktivasyon fonksiyonu kullanılabilir. Tek katmanlı ağlara göre çok katmanlı ağların avantajlarından yararlanabilmek için, doğrusal olmayan aktivasyon fonksiyonlarının kullanılması gerekmektedir.

1.5.2. θ Eşik Değeri ile Verilen İkili Adım (Heaviside) Fonksiyonu

Şekil-6'da gösterilen θ eşik değeri ile verilen ikili adım fonksiyonu,

$$f(x) = \begin{cases} 0 & , \quad x < \theta \text{ ise} \\ 1 & , \quad x \geq \theta \text{ ise} \end{cases} \quad [3]$$

formülü ile verilebilir.



Şekil-6: İkili adım fonksiyonu

Tek katmanlı ağlar genellikle, adım fonksiyonunun giriş değerlerini değeri ikili (0 ya da 1 veya -1 ya da 1) olan çıkış değerlerine çevirmek için kullanılmaktadırlar.

1.5.3. Sigmoid Fonksiyonları

Sigmoid fonksiyonları, oldukça kullanışlı aktivasyon fonksiyonlarıdır. Sigmoid fonksiyonları içinde lojistik ve hiperbolik tanjant fonksiyonları en yaygın olarak kullanılanlarıdır. Özellikle yapay sinir ağı modellerinden biri olan ve uygulamada sıkça kullanılan geriye beslemeli öğrenme algoritmalarında bu fonksiyonların kullanımı diğerlerine göre daha avantajlıdır. Çünkü fonksiyonun belirli bir noktadaki değeri ile onun türevinin değeri arasındaki ilişki öğrenme zamanındaki hesap yükünü azaltmaktadır.

1.5.3.1. Lojistik Fonksiyon

Lojistik fonksiyon, değerleri 0 ile 1 arasında değişen bir sigmoid fonksiyonudur ve yapay sinir ağları için aktivasyon fonksiyonu olarak sıkça kullanılmaktadır. Fonksiyonun aralık değerini vurgulamak için bu fonksiyona ikili sigmoid adı verilmekle birlikte lojistik sigmoid adı da verilmektedir.

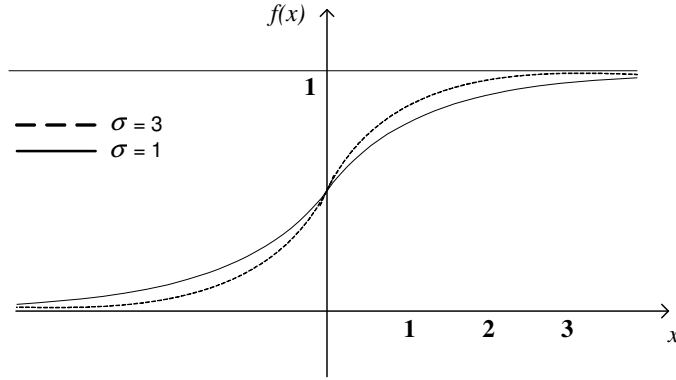
Şekil-7'de fonksiyonun adım parametresi olan α 'nın farklı değerleri için ikili sigmoid eğrileri görülmektedir. Lojistik fonksiyon,

$$f(x) = \frac{1}{1 + e^{-\alpha x}} \quad [4]$$

veya türevi olan

$$f'(x) = \sigma f(x)[1 - f(x)] \quad [5]$$

formülleri ile hesaplanmaktadır.



Şekil-7: İkili sigmoid fonksiyonu ($\sigma = 1$ ve $\sigma = 3$ için)

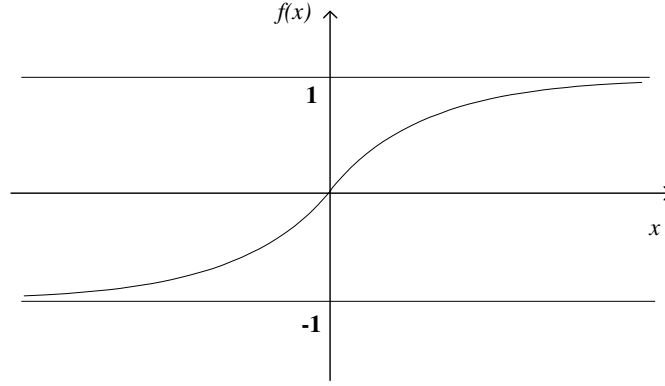
Lojistik sigmoid fonksiyonu istenen değer aralığına göre ölçeklenebilir ve böylece probleme uygun bir fonksiyon haline gelebilir. En yaygın kullanılan aralık, -1 ile 1 aralığıdır. Bu sigmoid fonksiyonu, iki kutuplu sigmoid olarak adlandırılır. Şekil-8'de $\sigma = 1$ için gösterilen iki kutuplu sigmoid fonksiyonu,

$$g(x) = 2f(x) - 1 = \frac{1 - e^{-\sigma x}}{1 + e^{-\sigma x}} \quad [6]$$

ve onun türevi olan

$$g'(x) = \frac{\sigma}{2} [1 + g(x)][1 - g(x)] \quad [7]$$

formülleriyle hesaplanmaktadır.



Şekil-8: İki kutuplu sigmoid fonksiyonu ($\sigma = 1$ için)

1.5.3.2. Hiperbolik Tanjant Fonksiyonu

Hiperbolik tanjant fonksiyonu da istenilen çıkış aralığı -1 ile 1 arasında ise aktivasyon fonksiyonu olarak kullanılmaktadır. İki kutuplu sigmoid fonksiyonuna çok yakın olan bu fonksiyon,

$$h(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{1 - e^{-2x}}{1 + e^{-2x}} \quad [8]$$

veya türevi olan

$$h'(x) = [1 + h(x)][1 - h(x)] \quad [9]$$

formülleri ile hesaplanmaktadır.

Eğer ikili veri 0 ile 1 arasında değişiyorsa bunu iki kutuplu biçime sokarak iki kutuplu sigmoid ya da hiperbolik tanjant fonksiyonunda kullanmak genelde tercih edilen bir yöntemdir (13).

1.6. Tek Katmanlı Yapay Sinir Ağı Modelleri

Tek katmanlı yapay sinir ağları, yapay sinir ağlarının önemli bir sınıfını oluştururken, aynı zamanda daha karmaşık yapıda bulunan çok katmanlı yapay sinir ağlarına da ışık tutarlar (13).

Örnek sınıflandırma problemlerinde, her giriş vektörü (örnek, numune) belirli sınıflara ait olabilir ya da olmayabilir. Basit olarak, bir sınıfa üye olma sorusu göz önünde bulundurulur. Çıkış birimi için +1 cevabının alınmasıyla örneğin o sınıfa üye olduğu, -1 cevabı alınmasıyla, örneğin o sınıfa üye olmadığı belirlenir. Bu tip durumlarda, her bir sınıf için bir çıkış birimi vardır.

Tek katmanlı sinir ağlarının eğitilmesinde üç önemli yöntem aşağıdaki gibidir.

- Hebb Kuralı
- Perseptron Öğrenme Kuralı
- Delta Kuralı

Gerçek dünyada karşılaşılan birçok problem, daha karmaşık mimarileri ve karmaşık eğitim kurallarını gerektirir ve genel olarak, tek katmanlı yapay sinir ağları bu tip problemleri çözmede yeterli değildir. Ancak şartlar bu ağları kullanmak için elverişli ise, doğru sonuçlar alınabilmesi mümkündür (16).

1.6.1. Hebb Kuralı

Hebb kuralı, bir yapay sinir ağı için, en eski ve en basit öğrenme kuralı olarak bilinir. Hebb, öğrenmenin, sinaps uzunluklarını (ağırlıkları) değiştirerek meydana geleceğini önermiştir. Hebb'e göre, eğer birbiri ile bağlı iki nöronun her ikisi de aynı zamanda "aktif" ise, bu nöronlara uygun ağırlıkların artırılması gerekmektedir. Benzer olarak, eğer her iki nöron aynı zamanda "pasif" ise, ağırlıkların artırılması gerekir. Hebb eğitiminde, iki nöron arasındaki bağlantı, bu nöronların öğrenmeleri sırasındaki aktivasyon değerlerinin arasındaki korelasyon miktarıyla orantılıdır (9).

1.6.2. Perseptron Öğrenme Kuralı

Perseptronlar, yapay sinir ağlarının öğrenebilir niteliğini taşıyan ilk modelidir. Hebb kuralından daha yetenekli bir öğrenme kuralıdır. Perseptron tekrarlı öğrenme algoritmasıdır ve çözümün varlığı durumunda yakınsama niteliğine sahiptir. Bu, perseptron modelinin en önemli niteliklerinden biridir.

Sınıflandırma problemlerinde eşitlik [10]'da verilen eşik değerli aktivasyon fonksiyonu kullanılır.

$$f(y_{in}) = \begin{cases} -1, & y_{in} < -\theta \text{ ise} \\ 0, & -\theta \leq y_{in} \leq \theta \text{ ise} \\ 1, & y_{in} > \theta \text{ ise} \end{cases} \quad [10]$$

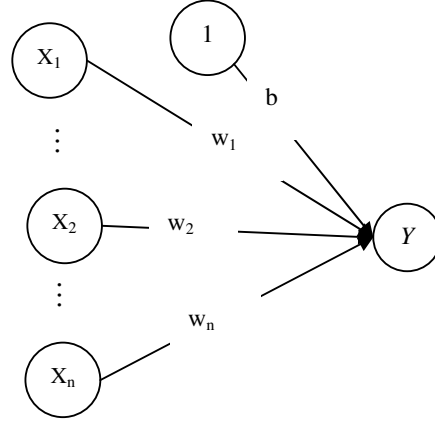
Çıktı biriminin aktivasyonu $y = f(y_{in})$ şeklinde hesaplanır.

Birleştirici birimden cevap birimine giden bağlantıların ağırlıkları perseptron öğrenme kuralı ile ayarlanır. Her eğitim girişi için, sinir ağı, çıkış biriminin cevabını hesaplar. Daha sonra sinir ağı, bu örnek için çıkış değeri ile hedeflenen çıkış arasındaki farkı karşılaştırarak bir hata oluşup oluşmadığını tespit eder. Yapay sinir ağı, hesaplanmış çıkış değeri “0” ve hedef değeri “-1” olan örnek için hatayı ayırt edemez, buna karşıt olarak hesaplanmış çıkış değeri “+1” ve hedef değeri “-1” olan örnek için hatayı ayırt edebilir. Bu durumlarda, hedef verinin işareti yönünde ağırlıkların işareti değiştirilmelidir. Bununla birlikte çıkış birimine “0” olmayan sinyaller gönderen bağlantıların ağırlıkları ayarlanmalıdır. Eğer belirli bir eğitim giriş örneğinde hata oluşuyorsa, ağırlıklar eşitlik [11]’deki gibi değiştirilmelidir.

$$w_i(\text{yeni}) = w_i(\text{eski}) + \alpha t x_i \quad [11]$$

Burada hedef değeri t, “+1” ya da “-1”dir ve α öğrenme oranı katsayısıdır. Eğer hata oluşmadıysa ağırlıklar değiştirilmemelidir. Eğitim işlemi hata oluşmayıncaya kadar devam etmelidir. Bu kuralın amacı, ağıın tam olarak doğru cevap veremediği eğitim örnekleri için ağırlıkları ayarlamaktır. Ayrıca, eğitim sonunda bu ağı sınırsız sayıdaki eğitim adımları için ağırlıkların değerlerini bulmalıdır (11).

Şekil 9’da perseptronun mimarisi gösterilmiştir. Burada X_1, \dots, X_n girdi birimleri, Y çıktı birimi ve 1 sapma sinyalidir. b sapma ağırlığı, w_i ($i = 1, \dots, n$) ağırlıklardır.



Şekil-9: Basit bir perseptron mimarisi

Sınıflandırma problemlerinde, sinir ağının görevi tüm giriş örneklerinin belirli bir sınıfa ait olup olmadığını belirlemektir. Sınıfa ait olma çıkışın “+1” değerine, ait olmama ise çıkışın “-1” değerine uygun olmasıyla belirlenir. Sınıflandırma işlemi yapılabilmesi için ağ, tekrarlı bir teknik ile eğitilir. Girdi ve hedefler ikili veya iki kutuplu olabilir. θ eşik değeri tüm birimler için değişmezdir. Sapma ve eşik değerinin her ikisinin aynı zamanda kullanılmasına ihtiyaç duyulmaktadır. Bu işlemin algoritması aşağıda verilmiştir. Bu algoritma, ağırlıkların başlangıç değerlerine ve öğrenme oranına tam olarak duyarlı değildir.

Adım 0: Ağırlıklar ve sapmalara başlangıç değerlerini ata.

(Ağırlıkları ve sapma değeri kolaylık için “0” olarak alınabilir.)

Öğrenme oranı olan α ’yı ayarla ($0 < \alpha \leq 1$).

(kolaylık için, α 1’e eşitlenebilir.)

Adım 1: Durma koşulu yanlış iken, adım 2-6’ yı uygula.

Adım 2: Her bir **s:t** öğrenme çifti için, 3-5 adımlarını uygula.

Adım 3: Giriş birimlerinin aktivasyonlarını ayarla.

$$x_i = s_i \quad i = 1, \dots, n$$

Adım 4: Her çıktı birimi için aktivasyonları hesapla.

$$y_{in_j} = b_j + \sum_{i=1}^n x_i w_{ij} ; \quad j = 1, \dots, l \quad [12]$$

$$f(y_{in}) = \begin{cases} -1, & y_{in} < -\theta \text{ ise} \\ 0, & -\theta \leq y_{in} \leq \theta \text{ ise} \\ 1, & y_{in} > \theta \text{ ise} \end{cases} \quad [13]$$

Adım 5: Ağırlıkları ve sapmaları ayarla.

eğer $t_j \neq y_j$ ise,

$$b_j(\text{yeni}) = b_j(\text{eski}) + t_j$$

$$w_{ij}(\text{yeni}) = w_{ij}(\text{eski}) + t_j x_i$$

eğer $t_j = y_j$ ise,

$$b_j(\text{yeni}) = b_j(\text{eski})$$

$$w_{ij}(\text{yeni}) = w_{ij}(\text{eski})$$

Adım 6: Durma koşulunu uygula (Eğer adım 2’de hiç bir ağırlık değişmezse dur; aksi durumda devam et).

Algoritmada çıktı birimlerinin sayısı $m = 1$ olabilir. Örneğin, mantıksal fonksiyonları gözden geçirirken çıktı biriminin sayısının bir olduğu kabul edilir. Eğitimden sonra, ağ her bir eğitim vektörünü doğru şekilde sınıflandırır.

Sınıflandırma ile ilgili perseptron eğitim algoritmasında, bir ayırma doğrusu yerine, pozitif cevaplar bölgesini sıfır cevaplar bölgesinden ayıran $w_1 x_1 + w_2 x_2 + b > \theta$ doğrusu ve negatif cevaplar bölgesini sıfır cevaplar bölgesinden ayıran $w_1 x_1 + w_2 x_2 + b < -\theta$ doğrusu olmak üzere iki ayırma doğrusu vardır (11).

1.6.3. Delta Kuralı

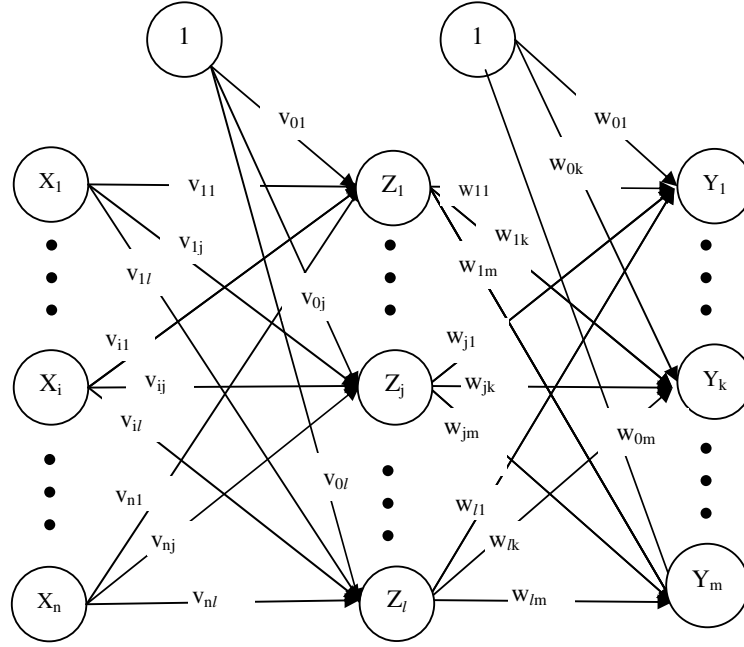
Widrow ve Hoff kuralı olarak da bilinen delta kuralı, Widrow ve Hoff tarafından 1960 yılında ortaya atılmış iteratif bir öğrenme sürecidir. Delta kuralında, tüm girdi numuneleri için çıktı ve hedef farkları karelerinin toplamının, başka bir ifadeyle, toplam hatanın küçültülmesi hedeflenmiştir. Amaç, tüm eğitim numunelerinin hatalarını en aza indirmektir. Ağırlık düzeltmeleri, çok sayıdaki eğitim numunesi ile beraber biriktirilebilir ve bu yığın güncelleştirilmesi olarak adlandırılır (9, 11, 13).

1.7. Çok Katmanlı Yapay Sinir Ağı Modelleri

Tek katmanlı ağların ayrılabilen problemlerin çözümünde başarısız oldukları görüldüğünde, bilim adamları çok katmanlı YSA modellerini incelemişlerdir. Burada önemli aşamalardan biri bu tip ağlar için akıllı bir eğitim algoritması geliştirmektir. 1986 yılında Rumelhart, Hinton ve Williams tarafından bu gerçekleştirildi (21). Standart geriye yayılım (back-propagation) olarak adlandırılan bu eğitim metodu hata kareler toplamının geriye yayılım yöntemiyle küçültülmesi fikrine dayanır ve genelleştirilmiş delta kuralını kullanır. Dolayısıyla bu yöntem her adımda hatanın küçültülmesi için, Widrow-Hoff eğitiminde olduğu gibi, gradient azalış yöntemini kullanır. Bu durumda gizli katmanda doğrusal olmayan aktivasyon fonksiyonları, örneğin lojistik sigmoid fonksiyonu ve ona uygun olarak genelleştirilmiş delta kuralı uygulanmaktadır. Bu yöntem daha iyi tahmin yapmak, sınıflandırmak ve öngörü problemleri için büyük imkanlar sağlamaktadır (22).

1.7.1. Standart Geriye Yayılım Ağ Mimarisi

Şekil-10'da bir gizli katmana sahip çok katmanlı ileri beslemeli bir ağ gösterilmiştir. Burada X girdi katmanına, Z gizli katmanına, Y çıktı katmanına ait birimlerdir. Gizli katmanın j. birimine dahil olan sapma ağırlıkları v_{0j} ($j = 1, \dots, l$), çıktı katmanının k. birimine dahil olan sapma ağırlıkları w_{0k} ($k = 1, \dots, m$) ve sapsalara uygun birimlerin girdi sinyalleri "1" olarak gösterilmiştir. Ağda sinyallerin yayılımı girdi birimlerinden gizli birimlere, sonra ise gizli birimlerden çıktı birimlerine doğru yönelmiştir. Bu nedenle ağ ileri beslemeli çok katmanlı ağ gibi göze alınmaktadır. Gizli birimin sayısı bir olduğundan şekilde gösterilen ağ iki katmanlı bir yapay sinir ağıdır. Gizli birimlerin sayısının birden fazla olduğu yapay sinir ağ mimarileri de vardır (11, 22).



Şekil-10: Tek gizli katmanlı ileri beslemeli çok katmanlı bir yapay sinir ağı

1.7.2. Standart Geriye Yayılım Eğitim Algoritması

Standart geriye yayılım eğitim algoritması aşağıdaki gibidir.

1. Aşama: İleri Besleme

Adım 0: Ağırlıkların ilk değerlerini başlat (küçük rasgele değerler olarak ayarla).

Adım 1: Durma koşulu yanlış iken 2-9 adımlarını uygula.

Adım 2: Tüm eğitim çiftleri için 3-8 adımlarını uygula.

Adım 3: Tüm \$X_i\$ (\$i=1, \dots, n\$) girdi birimlerinin \$x_i\$ girdi sinyallerini al ve bu sinyali üst katmanın tüm elemanlarına (gizli birimlere) dağıt.

Adım 4: Tüm \$Z_j\$ (\$j=1, \dots, p\$) gizli birimlerin ağırlıklı girdi sinyallerini topla.

$$y_in_j = b_j + \sum_{i=1}^n x_i w_{ij} ; \quad j = 1, \dots, m \quad [14]$$

Çıktı sinyalini hesaplamak için aşağıdaki gibi aktivasyon fonksiyonunu uygula.

$$z_j = f(z_in_j) \quad [15]$$

Bu sinyali üst katmanındaki tüm elemanlarına gönder.

Adım 5: Her Y_k ($k=1,\dots,m$) çıktı birimi ağırlıklı girdi sinyallerini aşağıdaki gibi topla.

$$y_{in_k} = w_{0k} + \sum_{j=1}^p z_j w_{jk} \quad [16]$$

Çıktı sinyalini hesaplamak için aktivasyon fonksiyonunu uygula.

$$y_k = f(y_{in_k}) \quad [17]$$

2. Aşama: Hatanın Hesaplanması ve Geriye Yayılması

Adım 6: Her Y_k ($k=1,\dots,m$) çıktı birimi girdi numunesine uygun hedef numunesini al ve onun hata bilgisini aşağıdaki gibi hesapla.

$$\delta_k = (t_k - y_k) \cdot f'(y_{in_k}) \quad [18]$$

Bu birim için ağırlık değişim terimini aşağıdaki gibi hesapla.

$$\Delta w_{jk} = \alpha \delta_k z_j \quad [19]$$

Sapma değişim terimini aşağıdaki gibi hesapla.

$$\Delta w_{0k} = \alpha \delta_k \quad [20]$$

δ_k değerini alt tarafındaki katmanın birimlerine gönder.

Adım 7: Her bir Z_j ($j=1,\dots,p$) gizli biriminin delta girdilerini topla:

$$\delta_{in_j} = \sum_{k=1}^m \delta_k w_{jk} \quad [21]$$

Bu değer için hata bilgi terimini hesaplamak için aktivasyon fonksiyonunun türevi ile çarp,

$$\delta_j^h = \delta_{in_j} f'(z_{in_j}) \quad [22]$$

ve ağırlık değişim terimini aşağıdaki gibi hesapla.

$$\Delta v_{ij} = \alpha \delta_j^h x_i \quad [23]$$

Sapma değişim terimini aşağıdaki gibi hesapla.

$$\Delta v_{0j} = \alpha \delta_j^h \quad [24]$$

3. Aşama: Ağırlıkların Güncellenmesi

Adım 8: Her bir çıktı biriminin sapmasını ve ağırlıklarını aşağıdaki gibi güncelle.

$$w_{jk}(\text{yeni}) = w_{jk}(\text{eski}) + \Delta w_{jk} \quad [25]$$

Her bir gizli birimin sapmasını ve ağırlıklarını aşağıdaki gibi güncelle.

$$v_{ij}(\text{yeni}) = v_{ij}(\text{eski}) + \Delta v_{ij} \quad [26]$$

Adım 9: Durma koşulunu kontrol et.

Her bir döngüde tüm eğitim vektörlerinin kümesi bir kez uygulanır. Bir geri yayılım sinir ağının eğitimi için birçok döngü gerekir. Buradaki algoritmada her eğitim numunesi kullanıldıktan sonra ağırlıklar güncellenir. Geriye yayılım algoritmasında, bir fonksiyonun eğimi, bu fonksiyonun hangi yönde daha hızlı arttığını gösterir. Eğimin ters yönü ise bu fonksiyonun hangi yönde en hızlı şekilde azaldığını gösterir (11, 13).

1.7.3. Geriye Yayılım Algoritma Çeşitleri

1.7.3.1. Momentum

Momentum geriye yayılımda, ağırlık değişiminin yönü o anki eğimle bir önceki eğimin kombinasyonu şeklindedir. Momentumu kullanmak için bir veya daha önceki eğitim numunelerinin ağırlıkları saklanmalıdır.

Örnek olarak, geri yayılımın momentumlu basit bir biçimi, t+1 eğitim adımının yeni ağırlıkları t ve t-1 eğitim adımlarındaki ağırlıkları temel alır. Momentumlu geri yayılımın ağırlık güncelleme formülü,

$$w_{jk}(t+1) = w_{jk}(t) + \alpha \delta_k z_j + \mu [w_{jk}(t) - w_{jk}(t-1)] \quad [27]$$

$$\Delta w_{jk}(t+1) = \alpha \delta_k z_j + \mu \Delta w_{jk}(t) \quad [28]$$

$$v_{ij}(t+1) = v_{ij}(t) + \alpha \delta_j x_i + \mu [v_{ij}(t) - v_{ij}(t-1)] \quad [29]$$

$$\Delta v_{ij}(t+1) = \alpha \delta_j x_i + \mu \Delta v_{ij}(t) \quad [30]$$

şeklindedir. Momentum katsayısı değeri olan μ , 0 -1 aralığında sınırlandırılmıştır.

Momentum, ağırlık değişimini kabul edilebilir ölçülerde yapmasını sağlar ve küçük bir öğrenme oranı ile birlikte kullanıldığında herhangi bir örnek için büyük hatalara

yol açacak olan cevapları engeller. Momentum kullanıldığında ağ eğimin yönünde işlem yapmaz. O anki ve ondan önceki ağırlık düzeltmesinin yönlerinin kombinasyonu yönünde işlem yapar (11, 20).

1.7.3.2. Eşlenik Eğitim Azalış Algoritması

Standart geriye yayılım algoritması, ağırlıkları eğimin ters yönünde adimsal olarak ayarlamaktadır. Bu yön, ağ performans fonksiyonunun en hızlı azaldığı yöndür. Fonksiyon eğimin ters yönü boyunca en hızlı şekilde azalsa da, bu önemli sayılabilecek hızlı bir yakınsama üretmemektedir. Eşlenik eğitim algoritmalarında, genellikle adimsal azalma yönlerinden daha hızlı yakınsamayı sağlayan eşlenik yönler boyunca bir arama yapılmaktadır.

1.7.3.3. Adapte Olabilen Öğrenme Oranları

Standart geriye yayılım algoritması, o anki ağırlıklar için olan hata yüzeyinde hatanın en hızlı azaldığı yönde ağırlıkları değiştirir. Ağırlık ayarlamasının yönünün değiştirilmesi hakkında çeşitli yöntemler önerilmiştir ve bu konuda çalışılmıştır. Bunlardan biri de, geriye yayılım ağının öğrenme oranının artırılması için öğrenme oranının eğitim sırasında değiştirilmesidir. Başlangıç ağ çıktısı ve hatası bulunur. Her döngüde ağırlıklar ve sapma o anki öğrenme oranı kullanılarak hesaplanır ve yeni ağırlıklar ve hata bulunur. Yeni hata eski hatayı geçerse yeni ağırlıklar ve sapma uygulanmaz ve buna ek olarak öğrenme oranı azaltılır (genelde 0,7 ile çarpılır), aksi halde yeni ağırlıklar ve sapma uygulanır ve öğrenme oranı artırılır (genelde 1,05 ile çarpılır).

1.7.3.4. Delta-Bar-Delta

Bu yaklaşımın temelinde, her ağırlığın kendi öğrenme oranının olması vardır. Öğrenme oranları eğitim sırasında değiştirilir. Eğer ağırlık değişimi (artma veya azalma) birkaç adım için aynı yönde ise ağırlıklar için öğrenme oranı artırılmalıdır. Bir ağırlık için oluşan hatanın kısmi türevinin işareti birkaç adım için aynı olursa ağırlık değişimi aynı yönde değiştirilmelidir, tam tersi olursa azaltılmalıdır.

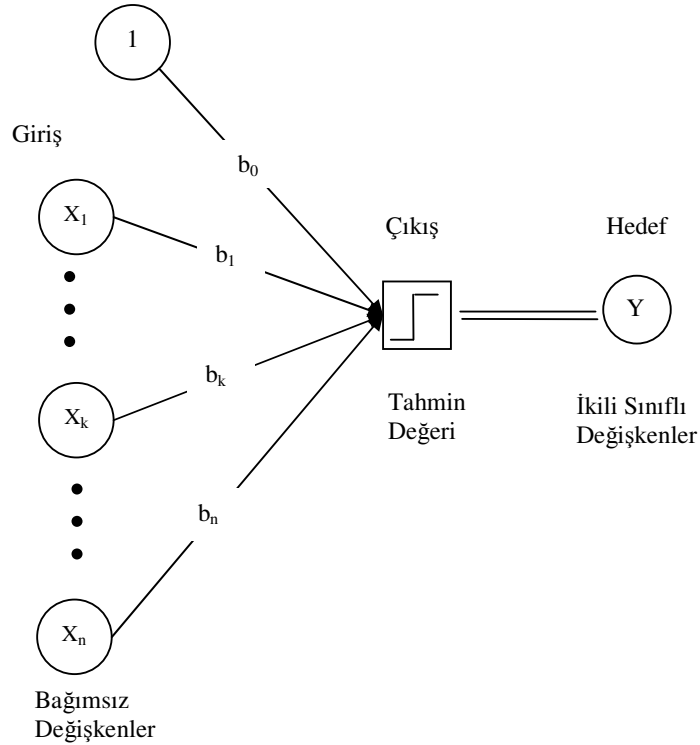
$w_{jk}(t)$, rasgele seçilmiş bir ağırlığı, $\alpha_{jk}(t)$ bu ağırlık için öğrenme oranı ve E ise örnek için hata kareleri belirtmek üzere, eşitlik [31]'e göre ağırlıklar güncellenmektedir.

$$w_{jk}(t+1) = w_{jk}(t) - \alpha_{jk}(t) \frac{\partial E}{\partial w_{jk}} \quad [31]$$

1.8. Yapay Sinir Ağları ve Bazı İstatistiksel Teknikler

Bazı yapay sinir ağı modelleri, bazı istatistiksel tekniklere benzer veya aynıdır. Örnek sınıflandırma ve tanıma problemlerinde istatistiksel teknikler ve yapay sinir ağı modellerinin kullanımı için uygun teknik ve modellerin karşılaştırılmasını gerektirmektedir. Bu bakımdan diskriminant analizi, lojistik regresyon ve benzeri istatistiksel tekniklere uygun yapay sinir ağı modellerinin belirlenmesi, onların avantaj ve dezavantajlarının açıklanması büyük önem taşımaktadır (12, 23).

Çıktı birimine eşik (adım) aktivasyon fonksiyonu uygulanan ve n sayıda girdi birimine sahip tek katmanlı bir perseptron Şekil-11'de gösterilmiştir.



Şekil -11: Eşik fonksiyonlu tek katmanlı perseptron ve doğrusal diskriminant fonksiyonu

Burada çıkış birimindeki σ işareti aktivasyonun eşik fonksiyonu olduğunu belirtir. Eşik fonksiyonu,

$$f(y_{in}) = \begin{cases} -1, & y_{in} < -\theta \text{ ise} \\ 0, & -\theta \leq y_{in} \leq \theta \text{ ise} \\ 1, & y_{in} > \theta \text{ ise} \end{cases} \quad [32]$$

şekindedir. θ eşik değeri, sıfıra eşit olursa [32] nolu fonksiyon,

$$f(y_{in}) = \begin{cases} -1, & y_{in} < 0 \text{ ise} \\ 1, & y_{in} \geq 0 \text{ ise} \end{cases} \quad [33]$$

şeklini alır.

Burada y_{in} , Y çıktı birimine dahil olan toplam ağırlıklı sinyaldir ve

$$y_{in} = b_0 + \sum_{i=1}^n b_i \cdot x_i \quad [34]$$

şeklinde hesaplanır. Bu durumda y değeri ise,

$$y = f(y_{in}) \quad [35]$$

formülü ile elde edilir.

Ağ belirli giriş birimleri için eğitildiği zaman $\hat{b}_0, \hat{b}_1, \dots, \hat{b}_n$ ağırlıkları belirlenir ve

$\hat{b}_0 + \sum_{i=1}^n \hat{b}_i \cdot x_i < \theta$ ile $\hat{b}_0 + \sum_{i=1}^n \hat{b}_i \cdot x_i \geq \theta$ hiper yarı düzlemleri örnekleri sınıflandırır. Başka

bir ifadeyle, eğer örnek için ağ çıktı aktivasyonu “1” ise o birinci hiper yarı düzleme (1. gruba), “-1” ise ikinci hiper yarı düzleme (2. gruba) atanır. $\theta = 0$ ise ayırıcı hiper düzlem

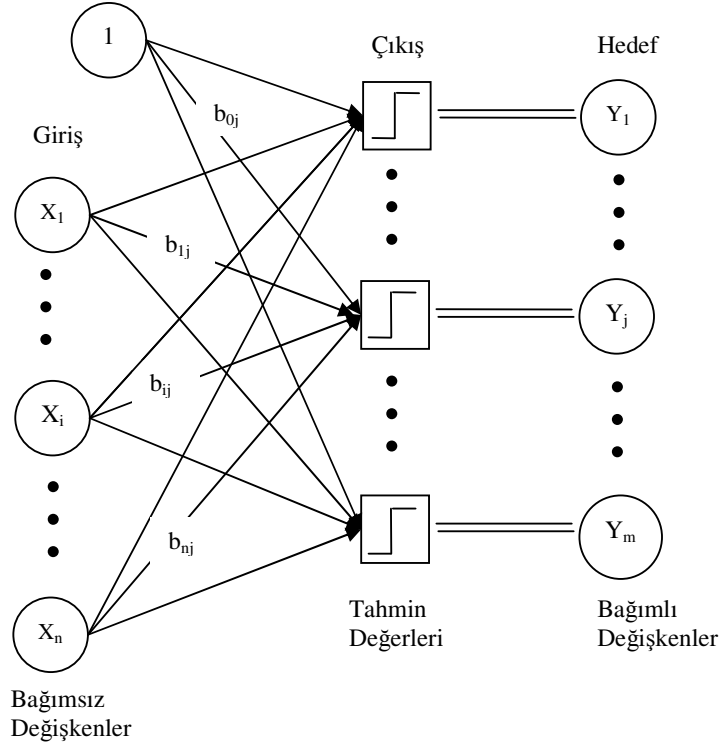
$\hat{b}_0 + \sum_{i=1}^n \hat{b}_i \cdot x_i = \theta$ olur. Bu işlem, istatistikteki doğrusal diskriminant analizinin rolünü

gerçekleştirir (17, 24).

Böylelikle, eşik fonksiyonlu perseptron ve doğrusal diskriminant fonksiyonunun her ikisi de örnekleri doğrusal ayrılabilir iki sınıfa ayırır. Böyle bir problem çözülebilirse,

perseptron yakınsama teoremine göre, perseptron algoritması uygun doğrusal diskriminant fonksiyonunun kat sayılarını yaklaşık olarak bulabilir (12).

Birden fazla çıktı birimi olan, çıktı birimlerine eşik fonksiyonu uygulanan ve n sayıda girdi birimine sahip tek katmanlı bir perseptron Şekil-12’de gösterilmiştir.



Şekil-12: Birden fazla çıktı birimi olan eşik fonksiyonlu tek katmanlı perseptron ve çoklu diskriminant fonksiyonu

Birden fazla çıktı birimi olan eşik aktivasyon fonksiyonlu perseptron, çoklu diskriminant fonksiyonudur. Bu durumda karakteristik veriler R^n uzayından bir R^q uzayına aktarılır ($q < n$) ve sınıflandırma R^q 'nin yardımıyla yapılır. Burada R^q bir q boyutlu hiper düzlemi göstermektedir (12).

2. Diskriminant Analizi

Yapılan gözlemler sonucu ortaya çıkan verilerin temel bir özeliği dolayısıyla hangi popülasyona ait olacağına karar vermek çoğu zaman mümkün değildir. Gözlemlerin sahip oldukları birden fazla özellik göz önünde tutulduğunda, bunların ait oldukları

popülasyonlara ayrılmasında somut bir belirleyici fonksiyon ya da fonksiyonlar gerekmektedir.

Diskriminant Analizi (DA), iki ya da daha fazla sayıdaki gruba ait birimler arasındaki farklılıkları maksimum yapan ve değişkenlerin doğrusal bileşiminden meydana gelen bir veya daha çok fonksiyonun belirlenmesidir (25, 26).

DA, sınıflandırma modellerinin geliştirilmesinde kullanılan, ilk çok değişkenli istatistiksel sınıflandırma yöntemidir (5).

Fisher tarafından tanıtıldığında, diskriminant analizinin amacı, iki grubu birbirinden ayırabilmektir. Fisher, değişkenlerin doğrusal bir bileşimi olan diskriminant fonksiyonunu oluşturmuştur (27). Daha sonra Welch diskriminant fonksiyonunun oluşturulmasında, Neyman-Pearson olabilirlik oranı kriterini kullanmıştır (28).

2.1. Diskriminant Analizinin Temel Varsayımları

Birimler üzerinden elde edilen ölçüm değerlerinin en azından aralıklı ölçekle ölçülmesi gerektiği diskriminant analizinde, birimlerin analiz öncesinde hangi gruplara ait olukları bilinmelidir. Grupların her biri birbirinden ayrık ve her birey sadece ve sadece bir gruba atanacak şekilde oluşturulmalıdır.

Gruplardaki birim sayılarının eşit olması gerekmemektedir. Ancak, en küçük gruptaki birim sayısının ayırıcı değişkenlerin sayısını geçmemesi gerekmektedir. Maksimum ayırıcı değişken sayısının, gruplardaki toplam birim sayısından iki eksik olması gerekmektedir.

Ayırıcı değişkenlerden herhangi birisi diğer bir ayırıcı değişkenin doğrusal bileşimi olmamalıdır. Ayırıcı değişkenlerden herhangi birisi ile diğer bir ayırıcı değişken arasında yüksek korelasyon varsa ya da bir değişken diğerlerinin bir fonksiyonu ise, bu değişken için tolerans değeri 0'a yaklaşacak ve bu durum bazı hesaplamaların yapılmasını engelleyecektir.

Diğer bir varsayım ise her grubun çok değişkenli normal dağılıma sahip popülasyonlardan çekilmiş olduğunun kabul edilmesidir.

DA, gruplara ilişkin varyans-kovaryans matrislerinin homojen olmasına karşı oldukça duyarlıdır. Grupların varyans-kovaryans matrislerinin eşit varsayılması, diskriminant fonksiyonlarını hesaplamakta kullanılan formüllerin elde edilmesini ve önem testlerinin yapılmasını kolaylaştırmaktadır (25, 26, 29, 30).

2.2. Diskriminant Fonksiyonlarının Elde Edilmesi

DA'nde birimlerin genel ortalamadan sapmalarının kareleri toplamı olan 'genel kareler toplamı ve çapraz çarpımlar matrisi', 'gruplar arası kareler toplamı ve çapraz çarpımlar matrisi' ile 'grup içi kareler toplamı çapraz çarpımlar matrisi'nin toplamından oluşmaktadır.

Genel kareler toplamı ve çapraz çarpımlar matrisi T ile gösterilmek üzere;

$$T = [t_{ij}]_{p \times p} \quad [36]$$

olup, bu matrisin elemanları aşağıdaki formüle göre hesaplanır.

$$t_{ij} = \sum_{k=1}^g \sum_{m=1}^{n_k} (x_{ikm} - \bar{x}_i)(x_{jkm} - \bar{x}_j) \quad (i = 1, 2, \dots, p; j = 1, 2, \dots, p) \quad [37]$$

Burada; g grup sayısı, x_{ikm} i. ayırıcı değişkenin k. gruptaki m. birey için değeri, n_k k. gruptaki birey sayısı, \bar{x}_i i. değişkenin genel ortalaması ve p ise değişken sayısıdır.

N, tüm gruptaki toplam birim sayısını göstermek üzere, eğer T matrisinin her elemanı (N-1)'e bölünürse "genel kovaryans matrisi" (Σ_T) elde edilir. Bu matrisin ana köşegeni üzerindeki elemanları varyansları, diğer elemanları ise kovaryansları göstermektedir. Söz konusu kovaryanslar iki değişkenin birlikte ne kadar değiştiğinin bir ölçüsü olmaktadır.

T matrisinin her elemanının, aynı sıra ve sütuna düşen iki köşegenel elemanın çarpımlarının kareköküne bölünmesiyle "genel korelasyon matrisi"(R) elde edilir. Bu matrisin elemanları,

$$r_{ij} = \frac{t_{ij}}{\sqrt{t_{ii}t_{jj}}} \quad i = 1, \dots, p; j = 1, \dots, p \quad [38]$$

şeklinde bulunur.

Grup içi kareler toplamı çapraz çarpımlar matrisi W ile gösterilmek üzere,

$$W = [w_{ij}]_{p \times p} \quad [39]$$

olup, bu matrisin elemanları aşağıdaki gibi hesaplanır.

$$w_{ij} = \sum_{k=1}^g \sum_{m=1}^{n_k} (x_{ikm} - \bar{x}_{ik})(x_{jkm} - \bar{x}_{jk}) \quad (i = 1, 2, \dots, p; j = 1, 2, \dots, p) \quad [40]$$

Burada; g grup sayısı, x_{ikm} i . ayırıcı değişkenin k . gruptaki m . birey için değeri, n_k k . gruptaki birey sayısı, \bar{x}_{ik} k . gruptaki i . değişkenin genel ortalaması ve p ise değişken sayısıdır.

W matrisinin her elemanı $(N-g)$ 'ye bölünürse grup kovaryans matrislerinin ağırlıklı bir ortalaması olan “gruplar içi kovaryans matrisi” (S_w) elde edilir.

T ve W matrislerinin farkı ($B = T - W$), “gruplar arası kareler toplamı ve çapraz çarpımlar matrisi” adı verilen B matrisi ile ölçülür. W ve B matrisleri grup içi ve gruplar arası ilişkilere ait temel bilgileri içerdiklerinden gruplar arası ayırımın yapılmasında kullanılan ölçüt veya fonksiyonda, bu matrislerin varyans miktarlarının birbirlerine olan oransal durumlarını da dikkate almak gerekir.

Bu durumda analizde kullanılacak ölçüt;

$$\lambda = \frac{v' B v}{v' W v} \quad [41]$$

şeklinde ifade edilen fonksiyonu maksimum yapan değerdir. Burada;

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_p \end{bmatrix}_{p \times 1} \quad [42]$$

$p \times 1$ boyutlu katsayılar vektörüdür. Lagrange yöntemine göre $v' W v = 1$ kısıtlaması ile λ 'nın sütun vektörü olan v 'ye göre kısmi türevi alınıp sıfıra eşitlenirse [43] nolu eşitlik elde edilir.

$$\frac{\partial \lambda}{\partial v} = \frac{2[(Bv)(v' W v) - (v' B v)(Wv)]}{(v' W v)^2} = 0 \quad [43]$$

[43] nolu eşitlik, $v'Wv = 1$ kısıtlaması kullanıldığında [44] nolu eşitlik şeklini alır.

$$Bv - \lambda Wv = 0 \quad [44]$$

Bu eşitlik W matrisinin tersi ile soldan çarpılırsa, aşağıdaki eşitlik [45] elde edilir.

$$(W^{-1}B - \lambda I)v = 0 \quad [45]$$

Özdeğer olarak adlandırılan λ 'nın köklerini bulan her λ değeri için [46] nolu denklemin çözümü aranan v vektörünü verir.

$$|W^{-1}B - \lambda I| = 0 \quad [46]$$

λ 'nın i 'inci kökü için bulunan vektör v_i ile gösterilirse, i 'inci ayırma fonksiyonu,

$$f_i = v_i' X \quad [47]$$

şeklinde ifade edilir. Buna göre, p tane ayırıcı değişken için, k . gruptaki, m . birey için ayırma fonksiyonu aşağıdaki gibi gösterilebilir:

$$f_{km} = v_1 X_{1km} + v_2 X_{2km} + \dots + v_p X_{pkm} \quad [48]$$

Burada;

X_{ikm} : i . ayırıcı değişkenin ($i=1,2,\dots,p$), k . gruptaki ($k=1,2,\dots,g$), m . birey ($m=1,2,\dots,n_k$) için değeri,

v_i : ayırma fonksiyonunun i . katsayısıdır.

Elde edilen v_i değerleri "istenilen ayırma fonksiyonunun standart olmayan ham katsayıları" olarak tanımlanır.

Katsayıların, dolayısıyla ayırma fonksiyonlarının yorumunu kolaylaştırmak için bazı dönüşümler yaparak yeni katsayılar elde edilebilmektedir. Bu katsayılar, özgün verilerin

standart hale getirilmeden aşağıdaki dönüşüm formüllerinin kullanılmasıyla bulunabilir (25).

$$u_i = v_i \sqrt{N-g} \quad \text{ve} \quad u_0 = -\sum_{i=1}^p u_i \bar{x}_i \quad [49]$$

dönüşüm formüllerinden ayırma fonksiyonu,

$$f_{km} = u_0 + u_1 X_{1km} + u_2 X_{2km} + \dots + u_p X_{pkm} \quad [50]$$

biçiminde elde edilir.

Katsayılar da yapılan düzeltme işlemi sayesinde, ayırma fonksiyonlarının eksen merkezi, sistemin merkezi olarak kabul edilen tüm ayırma fonksiyonları eksenlerinin sıfır değeri aldığı noktaya kayar. Böylece, herhangi bireyin ayırma fonksiyonu değerine bakıldığında, sistem merkezinden hangi yönde ve ne kadar uzakta olduğu da belirlenebilir (26).

Standart olmayan katsayılar her bir değişkenin sınıflandırma işlemine olan tekil katkısını göstermektedir. Ancak, özgün verilerin standart hale getirilip tekrar hesaplanması yerine aşağıdaki dönüşüm yapılarak standart katsayılar elde edilebilir (25).

$$c_i = u_i \sqrt{\frac{w_{ii}}{N-g}} = v_i \sqrt{w_{ii}} \quad [51]$$

Burada;

w_{ii} : i değişkeni için kareler toplamı

N: Tüm gruplardaki toplam birey sayısı

g: Grup sayısı

olarak alınmaktadır.

Standart katsayıların mutlak değerlerinin büyüklükleri kıyaslanarak hangi değişkenin sınıflandırıcı etkisinin, başka bir deyişle öneminin daha büyük olacağı yorumlanabilir (31).

Bir ayırıcı değişkenler ile diskriminant fonksiyonları arasındaki ilişkiyi ortaya koyan katsayılar ise, yapısal katsayılar olarak adlandırılmaktadır. Katsayının işareti ise değişkenin ilgili ayırıcı değişkenin fonksiyon üzerindeki etkisinin yönünü belirler (32).

Diskriminant fonksiyonlarının değişkenlerle ilişkisini gruplar arasında araştırılmak istenildiğinde genel kareler toplamı ve çapraz çarpımlar matrisinden yararlanarak “toplam yapısal katsayılar” aşağıdaki formüle göre elde edilebilir.

$$c_{kj}^* = \frac{v_{kj} \sqrt{t_{kk}}}{\sqrt{\sum_{i=1}^p \sum_{m=1}^p v_{ij} v_{mj} t_{im}}} \quad k = 1, \dots, p; j = 1, \dots, r \quad [52]$$

olmak üzere toplam yapısal katsayılar;

$$S_{ij} = \sum_{k=1}^p \frac{t_{ik} c_{kj}^*}{\sqrt{t_{ii} t_{kk}}} \quad i = 1, \dots, p; j = 1, \dots, r \quad [53]$$

formülüyle bulunur.

2.3. Ayırma Fonksiyonlarının Sayısı, Önem Sıraları ile Oransal Güçlerinin Belirlenmesi ve Ayırma Fonksiyonlarının Test Edilmesi

Diskriminant analizinde söz konusu grupların sayısı ikiden fazla olduğu için elde edilecek ayırma fonksiyonlarının sayısı birden fazla olacak ve maksimum ayırma fonksiyonu sayısının $\max\{r\} = \{g-1, p\}$ olacaktır.

Ayırma fonksiyonlarından önemli olanları Wilks lambda (Λ) istatistiği ile belirlenebilir. Wilks tarafından geliştirilen istatistik aşağıdaki formüle göre hesaplanmaktadır (33).

$$\Lambda_k = \prod_{i=k+1}^r \frac{1}{1 + \lambda_i} \quad k = 0, \dots, r-1 \quad [54]$$

Wilks Λ istatistiği 0 ile 1 değerleri arasında değişmektedir. Bu istatistikten elde edilen değer sıfıra ne kadar yakın ise, analizde kullanılan ayırıcı değişkenlerin gruplar arası

farklılıkları açıklamada o kadar etkili olduğunu göstermektedir. Değerin 1'e eşit olması durumunda, ayırıcı değişkenlerin gruplar arasında ayırım yapılmasında hiçbir etkisinin olmadığı, başka bir deyişle grup merkezlerinin çakışık olacağı yorumu yapılır.

Bulunan Λ istatistiğinin istatistik bakımından önemli olup olmadığına karar vermek amacıyla Bartlett tarafından geliştirilen ve yaklaşık $(p-k) \times (g-k-1)$ serbestlik dereceli ki-kare (χ^2) dağılımı gösteren test istatistiği aşağıdaki gibi ifade edilir (34).

$$\chi^2 = - \left[N - \left(\frac{p+g}{2} \right) - 1 \right] \log_e \Lambda_k \quad k = 0, \dots, r-1 \quad [55]$$

$k = 0$ için hesaplanan Λ istatistiği [55] nolu formülde yerine konulursa, bulunan χ^2 değerini, seçilen α anlamlılık seviyesinde, $(p-k) \times (g-k-1)$ serbestlik dereceli χ^2 tablo değeriyle karşılaştırmak suretiyle Λ istatistiğinin, dolayısıyla gruplar arası farklılığın önem kontrolü yapılabilir. Ayırma fonksiyonları elde edilmeden gruplar arası farklılığın seçilen anlamlılık seviyesinde önemli çıkması halinde Λ istatistiğinin ve dolayısıyla birinci ayırma fonksiyonunun istatistiksel olarak önemli olduğu anlaşılabilir, bu fonksiyonun elde edilmesine geçilir. Birinci fonksiyon elde edildikten sonra; aynı işlemler $k = 1, \dots, r-1$ için tekrar edilir. Ancak, bu işlemler sırasında Λ istatistiğinin önemsiz çıkması halinde geriye kalan fonksiyonların elde edilmesinden vazgeçilir. Çünkü bu aşamadan sonra elde edilecek fonksiyonlar da önemsiz olacağından grupları ayırmada etkili olamazlar.

Wilks Λ istatistiğini test etmek amacıyla uygulamada oldukça yaygın olarak kullanılan diğer bir test istatistiği, Rao tarafından geliştirilen F oranı olup buna ilişkin oran aşağıdaki şekilde ifade edilir.

$$F_\Lambda = \frac{1 - \Lambda^{1/c}}{\Lambda^{1/c}} \cdot \frac{n_2}{n_1} \quad [56]$$

Burada;

$$c = \sqrt{\frac{p^2 (g-1)^2 - 4}{p^2 (g-1)^2 - 5}} \quad [57]$$

$$n_1 = p(g-1) \quad n_2 = c \left[N - \frac{p+g+2}{2} \right] - \frac{p(g-1)}{2} + 1 \quad [58]$$

şeklinde tanımlanmaktadır.

F_Λ değeri, $p > 2$ ve $g > 3$ olduğu durumlarda n_1 ve n_2 serbestlik dereceli F dağılımı gösterir (29). Yukarıdaki formüle göre hesaplanacak olan F_Λ değeri, seçilen anlamlılık seviyesindeki n_1 ve n_2 serbestlik dereceli F tablo değeri ile karşılaştırılarak Λ istatistiğinin önemli olup olmadığı kontrol edilir. Λ istatistiğinin önemli bulunması durumunda, gruplar arası farklılığın ve dolayısıyla ayırma yönteminde açıklandığı gibi elde edilen özdeğerler arasında $\lambda_1 > \lambda_2 > \dots > \lambda_r$ sırası sözkonusu olduğundan, birinci özdeğerden elde edilen ayırma fonksiyonunun en fazla ayırım gücüne sahip olduğu kanıtlanmış olur (25).

2.4. Diskriminant Analizinde Sınıflandırma Kuralları

2.4.1. İki Grup Olması Durumunda Sınıflandırma

$X = (X_1, X_2, \dots, X_p)'$, birey üzerinden elde edilmiş ölçüm değerlerine karşılık gelen p boyutlu X rasgele vektörü, Π_1 popülasyonundan gelen bir rasgele vektör ise olasılık yoğunluk fonksiyonu $f_1(x)$, Π_2 popülasyonundan gelen bir rasgele vektör ise olasılık yoğunluk fonksiyonu $f_2(x)$ ile gösterilsin. X vektörünün alabileceği tüm değerler p boyutlu R^p örneklem uzayında yer almak üzere, sınıflandırma problemi bu uzayı öyle R_1 ve R_2 bölgelerine ayırır ki, X vektörü R_1 bölgesine düştüğünde X gözlem vektörünün elde edildiği birey G_1 grubuna, R_2 bölgesine düştüğünde ise G_2 grubuna sınıflandırılır. $R_1 \cup R_2 = R$ olup, R_1 ve R_2 ayrıktır.

q_1 ve q_2 , sırasıyla bireyin G_1 ve G_2 gruplarına ait olma olasılıklarını gösterebilir. Yani, karıştırılmış iki grubun birleşmesinden oluşan tek bir gruptan rasgele alınan bir birimin G_1 'den ve G_2 'den gelme olasılıkları q_1 ve q_2 olup, bu olasılıklar önsel olasılıklar olarak ifade edilmektedirler.

Gerçekte G_1 grubuna ait olan bir bireyin G_1 grubuna atanması olasılığı $P(1|1)$ ile gösterilmek üzere;

$$P(1|1) = P(G_1|G_1, R) = \int_{R_1} f_1(x) dx \quad dx = dx_1, dx_2, \dots, dx_p \quad [59]$$

ve gerçekte G_2 grubuna ait olan bir bireyin G_2 grubuna atanması olasılığı $P(2|2)$ ile gösterilmek üzere;

$$P(2|2) = P(G_2|G_2, R) = \int_{R_2} f_2(x) dx \quad dx = dx_1, dx_2, \dots, dx_p \quad [60]$$

olarak elde edilir. Bu ifadeler doğru sınıflandırma olasılıklarıdır. Atama yapılırken, birey ait olduğu gruptan farklı bir gruba atanabilir. Bu durumda hatalı atama yapılmış olur ve yanlış sınıflandırma olasılıkları aşağıdaki gibi elde edilir:

Gerçekte G_1 grubuna ait olan bir bireyin G_2 grubuna atanması olasılığı $P(2|1)$ ile gösterilmek üzere;

$$P(2|1) = P(G_2|G_1, R) = \int_{R_2} f_1(x) dx = 1 - \int_{R_1} f_1(x) dx \quad [61]$$

ve gerçekte G_2 grubuna ait olan bir bireyin G_1 grubuna atanması olasılığı $P(1|2)$ ile gösterilmek üzere;

$$P(1|2) = P(G_1|G_2, R) = \int_{R_1} f_2(x) dx = 1 - \int_{R_2} f_2(x) dx \quad [62]$$

olarak elde edilir. Bu durumda rasgele alınan bir birimin sırasıyla G_1 'den ve G_2 'den gelme olasılıkları olan q_1 ve q_2 önsel olasılıkları göz önüne alındığında toplam hatalı sınıflandırma olasılığı;

$$P(R, f) = P(1|2) q_1 + P(2|1) q_1 \quad [63]$$

olarak elde edilir (29, 35).

2.4.1.1. Toplam Hatalı Sınıflandırma Olasılığının Minimizasyonu Yaklaşımı

Yukarıdaki iki eşitlikten,

$$P(R, f) = q_1 \left(1 - \int_{R_1} f_1(x) dx\right) + q_2 \int_{R_1} f_2(x) dx = q_1 + \int_{R_1} [q_2 f_2(x) - q_1 f_1(x)] dx \quad [64]$$

eşitliği elde edilebilir. Toplam hatalı sınıflandırma olasılığını minimize eden optimal sınıflandırma kuralı R_{OPT} ile gösterilmek üzere, Welch tarafından aşağıdaki gibi verilmiştir:

$$R_{OPT} = \begin{cases} \frac{f_1(x)}{f_2(x)} > \frac{q_2}{q_1} & \text{ise } G_1 \text{'e} \\ \text{diğer durumda} & G_2 \text{'ye ata} \end{cases} \quad [65]$$

$$\frac{f_1(x)}{f_2(x)} = \frac{q_1}{q_2} \quad [66]$$

olması durumunda ise atama sezgiseldir (28, 29, 36).

2.4.1.2. Toplam Hatalı Sınıflandırma Maliyetinin Minimizasyonu Yaklaşımı

$C(2|1)$, G_1 grubuna ait bir bireyin G_2 grubuna hatalı sınıflandırılmasının maliyeti ve $C(1|2)$, G_2 grubuna ait bir bireyin G_1 grubuna hatalı sınıflandırılmasının maliyetini gösterebilir. Bu durumda, hatalı sınıflandırmanın toplam beklenen maliyeti,

$$C_T = C(2|1)P(2|1)q_1 + C(1|2)P(1|2)q_2 \quad [67]$$

olarak verilir. Toplam hatalı sınıflandırma maliyetini minimize eden optimal sınıflandırma kuralı,

$$R_{OPT} = \begin{cases} \frac{f_1(x)}{f_2(x)} > \frac{q_2 C(1|2)}{q_1 C(2|1)} & \text{ise } G_1 \text{'e} \\ \text{diğer durumda} & G_2 \text{'ye ata} \end{cases} \quad [68]$$

şeklinde verilmiştir (37). Maliyetler eşit alınırsa bu kural Welch'in yaklaşımıyla aynıdır.

2.4.1.3. Bayes Teoremi Yaklaşımı (Sonsal Olasılıkların Maksimizasyonu Yaklaşımı)

Bayes Teoremi yaklaşımına göre, birim en büyük sonsal olasılığa sahip olan gruba atanır.

Birimin G_1 grubundan geldiği bilindiğinde, x 'in koşullu yoğunluk fonksiyonu $f_1(x|G_1)$ olmak üzere, G_1 'in sonsal olasılığı;

$$P(G_1|x) = \frac{q_1 f_1(x)}{q_1 f_1(x) + q_2 f_2(x)} \quad [69]$$

şeklinde elde edilir. Buna göre optimal sınıflandırma kuralı;

$P(G_1|x) > P(G_2|x)$ ise, yani

$$\frac{q_1 f_1(x)}{q_1 f_1(x) + q_2 f_2(x)} > \frac{q_2 f_2(x)}{q_1 f_1(x) + q_2 f_2(x)} \quad [70]$$

ise, birim G_1 grubuna atanır ki bu da toplam hatalı sınıflandırma olasılıklarının minimizasyonu kuralına eşdeğerdir.

2.4.2. İki'den Fazla Grup Olması Durumunda Diskriminant Analizi ve Sınıflandırma

$X = (X_1, X_2, \dots, X_p)'$, birey üzerinden elde edilmiş ölçüm değerlerine karşılık gelen p boyutlu X rasgele vektörü, Π_i popülasyonundan gelen bir rasgele vektör ise olasılık yoğunluk fonksiyonu $f_i(x)$ ile gösterilsin ($i = 1, 2, \dots, g$). X vektörünün alabileceği tüm değerler p boyutlu R^p örneklem uzayında yer almak üzere, sınıflandırma problemi bu uzayı öyle $\{R_1, R_2, \dots, R_g\}$ bölgelerine ayırır ki, X vektörü R_i bölgesine düştüğünde X gözlem vektörünün elde edildiği birey G_i grubuna sınıflandırılır. Burada,

$R_1 \cup R_2 \cup \dots \cup R_g = R^p$ olup, $i = 1, 2, \dots, g$; $j = 1, 2, \dots, g$ ve $i \neq j$ için $R_i \cap R_j = \emptyset$ dir.

q_1, q_2, \dots, q_g , sırasıyla bireyin G_1, G_2, \dots, G_G gruplarına ait olma olasılıklarını gösterebilir. Bu durumda, karıştırılmış g tane grubun birleşmesinden oluşan tek bir gruptan rasgele alınan bir bireyin G_i 'den gelme olasılığı q_i 'dir. q_i 'ler önsel olasılıklardır.

Gerçekte G_i grubuna ait olan bir bireyin yine G_i grubuna atanması olasılığı $P(i|i)$ ile gösterilmek üzere;

$$P(i|i) = P(G_i|G_i, R) = \int_{R_i} f_i(x) dx \quad dx = dx_1, dx_2, \dots, dx_p \quad [71]$$

ve gerçekte G_j grubuna ait olan bir bireyin G_j grubuna atanması olasılığı $P(j|j)$ ile gösterilmek üzere;

$$P(j|j) = P(G_j|G_j, R) = \int_{R_j} f_j(x) dx \quad dx = dx_1, dx_2, \dots, dx_p \quad [72]$$

olarak elde edilir. Bu ifadeler doğru sınıflandırma olasılıklarıdır. Atama yapılırken, birey ait olduğu gruptan farklı bir gruba atanabilir. Bu durumda hatalı atama yapılmış olur ve yanlış sınıflandırma olasılıkları aşağıdaki gibi elde edilir:

Gerçekte G_i grubuna ait olan bir bireyin G_j grubuna atanması olasılığı $P(j|i)$ ile gösterilmek üzere;

$$P(j|i) = P(G_j|G_i, R) = \int_{R_j} f_j(x) dx = 1 - \int_{R_i} f_i(x) dx \quad dx = dx_1, dx_2, \dots, dx_p \quad [73]$$

ve gerçekte G_j grubuna ait olan bir bireyin G_i grubuna atanması olasılığı $P(i|j)$ ile gösterilmek üzere;

$$P(i|j) = P(G_i|G_j, R) = \int_{R_i} f_i(x) dx = 1 - \int_{R_j} f_j(x) dx \quad dx = dx_1, dx_2, \dots, dx_p \quad [74]$$

olarak elde edilir.

Ayrıca G_i grubunun bir üyesinin yanlış sınıflandırma olasılığı $P(i)$ ile gösterilmek üzere;

$$P(i) = \sum_{j=1, j \neq i}^g P(j|i) = 1 - P(i|i) \quad [75]$$

olarak elde edilir. Bu durumda q_i önsel olasılıkları da göz önüne alındığında toplam hatalı sınıflandırma olasılığı;

$$P(R, f) = \sum_{i=1}^g q_i P(i) = 1 - \sum q_i P(i|i) \quad [76]$$

şeklinde verilir (29).

2.4.2.1. Toplam Hatalı Sınıflandırma Olasılığının Minimizasyonu Yaklaşımı

G grup bulunması durumunda toplam hatalı sınıflandırma olasılığını minimize eden optimal sınıflandırma kuralı, 2 grup bulunması durumundakinin genelleştirilmiş şeklidir ve aşağıdaki gibi verilmiştir:

$$\frac{f_i(x)}{f_j(x)} > \frac{q_j}{q_i} \quad (j = 1, 2, \dots, g) \text{ ise birim } G_i \text{ grubuna, aksi taktirde } G_j \text{ grubuna atanır.}$$

Böylece, $q_i f_i(x)$ maksimum değerine ulaşır. Bütün q_i olasılıklarının eşit olması durumunda bu kural maksimum olabilirlik kuralı olarak adlandırılır (29, 36).

2.4.2.2. Toplam Hatalı Sınıflandırma Maliyetinin Minimizasyonu Yaklaşımı

$C(j|i)$, G_i grubuna ait bir bireyin G_j grubuna hatalı sınıflandırılmasının maliyeti ve $C(i|j)$, G_j grubuna ait bir bireyin G_i grubuna hatalı sınıflandırılmasının maliyetini göstermek üzere, hatalı sınıflandırmanın toplam beklenen maliyeti,

$$C_T = \sum_{i=1}^g q_i \left[\sum_{j=1, j \neq i}^g C(j|i) P(j|i) \right] \quad [77]$$

olarak verilir. Toplam hatalı sınıflandırma maliyetini minimize eden optimal sınıflandırma kuralı,

$$R_{OPT} = \begin{cases} \frac{f_i(x)}{f_j(x)} > \frac{q_j C(i|j)}{q_i C(j|i)} \text{ ise } G_i \text{ 'e} \\ \text{diger durumda } G_j \text{ 'ye ata} \end{cases} \quad [78]$$

şeklinde verilmiştir (37).

2.4.2.3. Bayes Teoremi Yaklaşımı (Sonsal Olasılıkların Maksimizasyonu Yaklaşımı)

Bayes Teoremi yaklaşımına göre, birim en büyük sonsal olasılığa sahip olan gruba atanır.

Birimin G_i grubundan geldiği bilindiğinde, x 'in koşullu yoğunluk fonksiyonu $f_i(x|G_i)$ olmak üzere, G_i 'in sonsal olasılığı;

$$P(G_i|x) = \frac{q_i f_i(x)}{\sum_{m=1}^g q_m f_m(x)} \quad [79]$$

şeklinde elde edilir. Buna göre optimal sınıflandırma kuralı;

$P(G_i|x) > P(G_j|x)$ ise, yani

$$\frac{q_i f_i(x)}{\sum_{m=1}^g q_m f_m(x)} > \frac{q_j f_j(x)}{\sum_{m=1}^g q_m f_m(x)} \quad [80]$$

ise, birim G_i grubuna atanır ki bu da toplam hatalı sınıflandırma olasılıklarının minimizasyonu kuralına eşdeğerdir (38).

2.5. Karesel Diskriminant Analizi

Diskriminant analizinin varsayımlarından bir tanesi olan grup varyans-kovaryans matrislerinin homojen olması özelliği gerçekleşmediğinde, doğrusal diskriminant analizi yerine karesel diskriminant analizi (KDA) yöntemiyle ayırma fonksiyonları hesaplamalıdır.

KDA'nde iki ayırım bölgesi arasındaki sınır karesel bir yüzeydir. Kovaryans matrisleri arasındaki fark çok büyük değilse ve özellikle veri setindeki birim sayısı az ya da orta düzeyde ise doğrusal diskriminant analizi daha etkili sonuçlar vermektedir (39).

KDA'nde karşılaşılabilecek temel problem, tahmin edilmesi gereken parametre sayısının büyük olması sonucunda, kovaryans matrisinden elde edilecek özdeğerlerin 0 ya da 0'a eşit olmasıdır. Bu problemi ortadan kaldırmak için kullanılan yöntemler regülerizasyon yöntemleri olarak bilinmekte olup, en yaygın olarak kullanılanı Friedman tarafından öne sürülmüştür. Friedman tahmin edicilerin iki parametreliliğiyle, karesel ve doğrusal diskriminant yöntemleri arasında bir ilişki kurmuştur.

KDA'nde katsayıların hesaplanmasında ortak kovaryans matrisi S_{pooled} yerine grupların kovaryans matrislerinin farkı alınır. İki grup olması durumunda KDA için doğrusal bileşenler [82] nolu eşitlikteki gibi elde edilir.

$$V = (S_1^{-1} - S_2^{-1})(\bar{x}_1 - \bar{x}_2)' \quad [82]$$

Ayırma fonksiyonu ise, [83] nolu eşitlikteki gibi elde edilir.

$$Y = (\bar{x}_1 - \bar{x}_2)'(S_1^{-1} - S_2^{-1})X \quad [83]$$

x_0 gözlem vektörünün sınıflandırma kriteri aşağıdaki gibi tanımlanır:

$$-\frac{1}{2}x_0'(S_1^{-1} - S_2^{-1})x_0 + (\bar{x}_1'S_1^{-1} - \bar{x}_2'S_2^{-1})x_0 - \frac{1}{2}\ln\left(\frac{|S_1|}{|S_2|}\right) + \frac{1}{2}(\bar{x}_1'S_1^{-1} - \bar{x}_2'S_2^{-1}) \geq \left(\frac{p_2}{p_1}\right) \quad [84]$$

ise x_0 G_1 'e, aksi halde G_2 'ye atanır (37, 40).

Çoklu karesel diskriminant analizinde, bu işlemler geliştirilerek ayırma fonksiyonları ve sınıflandırma kuralı elde edilebilir.

Marks ve Dunn, normal dağılıma sahip veri seti için, küçük örneklem genişliklerinde, KDA ve DDA'nin performanslarını karşılaştırmıştır. Bu çalışma sonucunda, küçük örneklem genişlikleri için (yaklaşık olarak $n_1, n_2 < 25$), grupların kovaryans matrisleri yaklaşık eşit olduğunda ve bağımsız değişken sayısı yeterli olduğunda (> 6), KDA'nin DDA'nden daha kötü bir performansa sahip olduğu gösterilmiştir. Ancak kovaryans

matrisleri arasındaki fark arttıkça KDA'nin performansı DDA'ninkine yaklaşmaktadır. Wahk ve Kronmal örneklem genişlikleri yeteri kadar büyük olduğunda, kovaryans matrisleri arasındaki fark büyük ise, KDA'nin daha iyi sonuç verdiğini göstermiştir. Grupların kovaryans matrisleri eşit olmadığında, örneklem genişliğinin küçük olması durumunda KDA'nin zayıf sonuçlar verdiğini, Aitchison ve arkadaşları ve Remme ve arkadaşları tarafından yapılan simülasyon çalışmalarıyla da desteklenmiştir (29).

GEREÇ VE YÖNTEM

01.05.2003 ve 01.05.2005 tarihleri arasında, Uludağ Üniversitesi Tıp Fakültesi Uygulama ve Araştırma Hastanesi Gastroenteroloji servisine başvuran ve yapılan tetkik-incelemeler ile kronik karaciğer parenkim hastalığı düşünülerek karaciğer biyopsisi uygulanması sonucunda Hepatit B ve C virüslerine bağlı kronik hepatit veya siroz tanısı alan hastalar çalışmaya dahil edildi. Bu hastalara ilişkin geriye dönük olarak toplanmış olan verilere diskriminant analizi ve yapay sinir ağları uygulanarak, iki yöntemin elde edilmiş olan veri seti üzerinde kronik hepatit ve siroz hastalarını sınıflandırma performanslarının incelenmesi amaçlandı.

Bu doğrultuda birimlerin sınıflandırılmasında kullanılabilecek olan değişkenler Uludağ Üniversitesi Gastroenteroloji Bilim Dalı uzmanları ile birlikte değerlendirilerek belirlendi. 250 hastaya ilişkin veriler retrospektif olarak dosyalardan toplandı. Toplanan verilerin, hastaya biyopsi uygulanma tarihiyle mümkün olduğu kadar yakın tarihte elde edilmiş olmasına dikkat edildi. Çalışmaya sadece, biyopsi uygulamasından önce ve sonraki bir aylık dönemde veri elde edilebilen 164 birim dahil edildi.

İncelenen değişkenler aşağıdaki gibidir:

- Yaş (yıl)
- Cinsiyet (1: Erkek, 2: Kadın),
- Lökosit sayısı ($\times 1000/\text{ml}$)
- Platelet sayısı ($\times 1000/\text{ml}$)
- Sodyum (mEq/L)
- Albumin/Globulin oranı (g/dl)
- Total bilirubin (mg/dl)
- Aspartat aminaz transferaz /Alanin amino transferaz oranı (U/L) (AST/ALT)
- Laktat dehidrogenaz (U/L)
- Protrombin zamamı (saniye)
- INR (International normalized ratio)
- Serum protein elektroforezinde gamaglobulin fraksiyonu (%).

Birimlerin ait oldukları grubu göstermek üzere kodlanmış olan, diskriminant analizinde bağımsız değişken ve yapay sinir ağlarında ise çıktı birimi, “durum” olarak adlandırılmıştır.

- durum (1: siroz, 0: kronik hepatit)

Elde edilen veri setine diskriminant analizi uygulanmadan önce, diskriminant analizinin temel varsayımlarının sağlanıp sağlanmadığı araştırıldı. Örneklemelerin çok değişkenli normal dağılımdan geldikleri varsayımının sağlanıp sağlanmadığı PAST 1.45 paket programında, Mardia'nın Çok Değişkenli Normallik testi ile araştırıldı (41). Gruplara ilişkin kovaryans matrislerinin homojen olması varsayımının sağlanması ancak çok değişkenli normallik varsayımının sağlanmaması üzerine, değişkenler üzerinde gerekli dönüşümler uygulandı. Temel varsayımlar tekrar sınılandıktan sonra diskriminant analizi, SPSS 13.00 istatistiksel analiz paket programı kullanılarak yapıldı. Yapay sinir ağları uygulaması için ise STATISTICA 6.0 istatistiksel analiz paket programı kullanıldı.

BULGULAR

Analize dahil edilen 164 hastaya uygulanmış olan biyopsi sonuçlarına göre 79 (%48,2) hasta siroz, 85 (%51,8) hasta ise kronik hepatit tanısı almıştır. 164 hastadan elde edilen verilere ait tanımlayıcı istatistikler Tablo-1 ve Tablo-2’de verilmiştir.

Tablo-1: Çalışmada kullanılan nümerik değişkenlere ait tanımlayıcı istatistikler.

Değişkenler	n	Ortalama	Standart sapma	Standart hata	Minimum	Maksimum
Kronik hepatit						
Yaş	85	44,42	13,65	1,48	16,00	76,00
Lökosit sayısı	85	6,70	1,76	0,19	3,20	12,20
Platelet sayısı	85	194,88	55,17	5,98	79,00	312,00
Sodyum	85	142,72	6,62	0,72	131,00	192,00
Albumin/Globulin	85	1,44	0,29	0,03	0,79	2,25
Total bilirubin	85	0,81	0,46	0,05	0,20	3,70
AST/ALT	85	0,85	0,46	0,05	0,25	2,51
Laktat dehidrogenaz	85	293,75	113,83	12,35	109	939
Protrombin zamanı	85	13,10	1,79	0,19	11,00	21,10
INR	85	1,13	0,18	0,02	0,79	2,01
Gamaglobulin fraksiyonu	85	21,18	4,39	0,48	12,20	41,00
Siroz	n	Ortalama	Standart sapma	Standart hata	Minimum	Maksimum
Yaş	79	59,18	10,98	1,23	23,00	85,00
Lökosit sayısı	79	4,78	1,89	0,21	1,03	9,80
Platelet sayısı	79	133,50	85,14	9,58	22,30	431,00
Sodyum	79	135,83	6,55	0,74	114,00	149,00
Albumin/Globulin	79	0,88	0,29	0,03	0,39	2,07
Total bilirubin	79	3,79	3,80	0,43	0,30	16,80
AST/ALT	79	1,71	1,04	0,12	0,50	5,29
Laktat dehidrogenaz	79	376,32	131,02	14,74	109	755
Protrombin zamanı	79	18,54	6,26	0,70	11,60	48,70
INR	79	1,66	0,56	0,06	0,79	4,90
Gamaglobulin fraksiyonu	79	30,80	6,76	0,76	14,00	49,90

Tablo-2: Çalışmada kullanılan kategorik değişkene ait tanımlayıcı istatistikler.

Değişken	Kronik hepatit		Siroz	
	Sayı	Yüzde (%)	Sayı	Yüzde (%)
Erkek	48	56,47	53	67,09
Kadın	37	43,53	26	32,91

Cinsiyet bakımından kronik hepatit ve siroz grupları arasında anlamlı fark olup olmadığını ortaya koymak için Ki kare testi uygulanmıştır. Buna göre Pearson Ki kare test istatistiğinin değeri 1,95 ve p değeri 0,16 olarak elde edilmiştir. Bunun sonucunda, iki grup arasında cinsiyet bakımından istatistiksel olarak anlamlı bir fark olmadığı sonucuna varılmıştır.

1. Diskriminant Analizi Sonuçları

Diskriminant analizine bağımlı değişken olarak, hastaların siroz ya da kronik hepatit olma durumunu gösteren ‘durum’ değişkeni, bağımsız değişkenler olarak ise yaş, lökosit sayısı, platelet sayısı, sodyum, albumin/globulin oranı, total bilirubin, laktat dehidrogenaz, AST/ALT oranı, protrombin zamanı, INR, serum protein elektroforezinde gamaglobulin fraksiyonu değişkenleri dahil edilmiştir.

Elde edilen veri setine diskriminant analizi uygulanmadan önce, temel varsayımların sağlanıp sağlanmadığı araştırıldı. Her bir gruba ilişkin kovaryans matrislerinin homojen olması varsayımının test edilmesi için Box’s M testi uygulandı ve test sonucunda $p > 0,05$ olarak bulundu. Bu durumda “gruplara ait kovaryans matrisleri homojendir” yokluk hipotezi kabul edildi.

Örneklemelerin çok değişkenli normal dağılımdan geldikleri varsayımının sağlanıp sağlanmadığının test edilmesi için Mardia’nın Çok Değişkenli Normallik testi uygulandı. Test sonucunda p değeri 0,05’den küçük olarak elde edildiğinden örneklemelerin çok değişkenli normal dağılımdan geldikleri yönünde kurulan H_0 hipotezi reddedildi.

Bu durumda değişkenlere ayrı ayrı, Kolmogorov Smirnov testi uygulanarak tek değişkenli normal dağılıma uymayan değişkenler belirlendi. Platelet sayısı, sodyum, albumin/globulin oranı, total bilirubin, laktat dehidrogenaz, AST/ALT oranı, protrombin zamanı, INR, serum protein elektroforezinde gamaglobulin fraksiyonu değişkenleri için $p < 0,05$ olarak elde edildi. Bu değişkenlere normal dağılıma uymadıklarından dolayı, logaritmik dönüşüm uygulandı ve çok değişkenli normallik varsayımının sağlanıp

sağlanmadığı Mardia'nın Çok Değişkeni Normallik testi ile tekrar test edildi. Sonuçlar Tablo-3'deki gibi elde edilmiştir.

Tablo-3: Mardia'nın Çok Değişkeni Normallik testi sonuçları

	Katsayı	Test istatistiği değeri	Serbestlik derecesi	p değeri
Çarpıklık	8,12	221,90	220	0,45
Basıklık	121,80	0,73	220	0,47

p değeri çarpıklık için 0,45, basıklık için ise 0,47 olarak elde edildi. Dolayısıyla veri setinin çok değişkenli normal dağılıma uygun olduğu hipotezi kabul edildi.

Uygun dönüşümler sonucu elde edilmiş yeni değişkenler diskriminant analizine dahil edilmiştir.

Bağımsız değişkenler arasında çoklu bağlantı olup olmadığının araştırılması amacıyla, korelasyon matrisi oluşturularak bağımsız değişkenler arasındaki korelasyonlar değerlendirilmiştir. Herhangi iki değişken arasında, korelasyon matrisinin elemanları arasında 0,5'den büyük korelasyon katsayısı elde edilmemiştir.

Grupların kovaryans matrislerinin eşitliğine ilişkin Box's M testi sonuçları Tablo-4'de verilmiştir.

Tablo-4: Grupların kovaryans matrislerinin eşitliğine ilişkin Box's M testi sonuçları.

Box's M test istatistiğinin değeri	81,33
F değeri	1,14
Serbestlik derecesi 1	66
Serbestlik derecesi 2	82760,02
p değeri	0,20

Bu sonuçlara göre, p değeri 0,05'den büyük olduğu için, H_0 hipotezi kabul edilmiştir. Dolayısı ile grupların kovaryans matrislerinin homojen olduğu söylenebilir. Bu sonuçlara göre veri setine DDA uygulanmasına karar verildi.

DDA'ne ilişkin sonuçlar Tablo-5 ve Tablo-6'da verilmiştir.

Tablo-5: DDA sonucunda elde edilen özdeğer ve kanonik korelasyon.

Fonksiyon	Özdeğer	Toplam varyansı açıklama oranı (%)	Kanonik Korelasyon
1	3,203	100,00	0,873

Tablo-6: Fonksiyonların Wilks' Lambda test sonuçları.

Wilks' Lambda Değeri	Ki kare değeri	Serbestlik derecesi	p değeri
0,238	224,716	11	< 0,001

Buna göre DDA sonucunda Wilks' lambda değeri 0,238 olarak, Ki kare değeri 224,716 olarak elde edilmiş ve modelin istatistiksel olarak anlamlı olduğu sonucu elde edilmiştir ($p < 0,05$).

Diskriminant fonksiyonu katsayıları, standartlaştırılmış kanonik diskriminant fonksiyonu katsayıları, bağımsız değişkenler ve standartlaştırılmış kanonik diskriminant fonksiyonu katsayıları arasındaki grup içi korelasyonlar ve fonksiyondaki korelasyon büyüklüklerine göre değişkenlerin sırası Tablo-7'de verilmiştir.

Tablo-7: Diskriminant fonksiyonu katsayıları, standartlaştırılmış kanonik diskriminant fonksiyonu katsayıları, bağımsız değişkenler ve standartlaştırılmış kanonik diskriminant fonksiyonu katsayıları arasındaki grup içi korelasyonlar ve fonksiyondaki korelasyon büyüklüklerine göre değişkenlerin sırası.

Değişkenler	Diskriminant Fonksiyonu Katsayıları		Standartlaştırılmış Kanonik Diskriminant Fonksiyonu Katsayıları	Korelasyon	Korelasyon Sırası
	Kronik hepatit	Siroz			
Yaş	0,059	0,136	0,265	0,342	8
Lökosit sayısı	2,774	2,449	0,167	0,295	10
Platelet sayısı	0,112	0,098	0,206	0,411	5
Sodyum	5,812	5,685	0,176	0,337	9
Albumin/Globulin	21,254	17,871	0,277	0,548	2
Total bilirubin	18,731	20,892	0,195	0,428	4
AST/ALT	1,799	2,049	0,035	0,350	7
Laktat dehidrogenaz	19,768	20,580	0,086	0,186	11
Protrombin zamanı	36,342	33,976	0,156	0,411	6
INR	31,813	36,753	0,276	0,482	3
Gamaglobulin fraksiyonu	248,231	263,426	0,439	0,638	1
Sabit	797,150	813,054	-	-	-

Yapılan sınıflandırma işlemi sonucu, atama değerleri ile gözlem değerleri karşılaştırılarak doğru sınıflandırma yüzdeleri hesaplanmıştır. Elde edilen sonuçlar Tablo-8'de verilmiştir.

Tablo-8: Diskriminant analizi sınıflandırma sonuçları

	Durum	Tahmin Edilen Grup Üyeliği		
		Kronik hepatit	Siroz	Toplam
Gerçek grup üyeliği	Kronik hepatit	79	6	85
	Siroz	4	75	79
Toplam		83	71	164
Doğruluk		% 92,94	% 94,94	% 93,94

Tablo 8'de görüldüğü gibi, uygulamadaki 14 hastadan gerçekte 85 tanesi kronik hepatit, geriye kalan 79 tanesi ise sirozdur. Kronik hepatit olan 85 hastanın 79 tanesi DDA

ile yapılan sınıflandırma işleminde doğru, 4 tanesi ise hatalı olmak üzere % 92,94'lük bir doğruluk yüzdesi ile sınıflandırılmıştır. Siroz olan 79 hastanın ise 75 tanesi DDA ile yapılan sınıflandırma işleminde doğru, 6 tanesi ise hatalı olmak üzere % 94,94'lük bir doğruluk yüzdesi ile sınıflandırılmıştır. Buna göre modelin genel doğruluk yüzdesi ise % 93,94'dür.

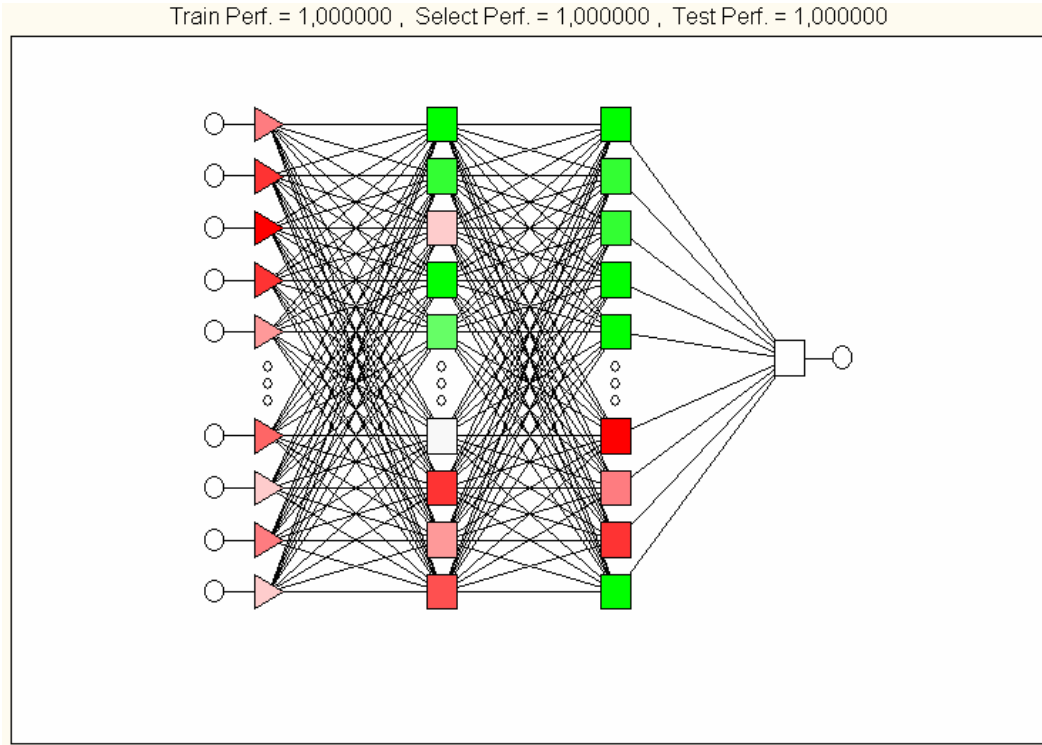
2. Yapay Sinir Ağı Modeli ile Sınıflandırma Sonuçları

YSA uygulamasında, diskriminant analizine bağımlı değişken olarak alınmış olan hastaların siroz ya da kronik hepatit olma durumunu gösteren 'durum' değişkeni çıktı birimi olarak alınmıştır. Diskriminant analizine dahil edilmiş olan değişkenlerin tümü YSA'ında girdi birimleri olarak alınmıştır.

YSA uygulamasına geçmeden önce, kullanılacak olan veri seti %30'u eğitim, %30'u geçerlilik ve %40'ı test seti olmak üzere üçe ayrılmıştır. En yüksek sınıflandırma performansını elde edebilmek için simülasyon yapılarak 1000 tane farklı yapay sinir ağı modeli denenmiştir. En yüksek sınıflandırma performansına sahip olan model seçilmiştir.

Uygulamada kullanılan veri setinin sınıflandırılması için girdi katmanında 11 birim, çıktı katmanında 1 birim olan ve ilk gizli katmanında 10, ikinci gizli katmanında 9 birim olmak üzere iki gizli katmana sahip olan üç katmanlı bir yapay sinir ağı modeli kurulmuştur.

Kurulan model Şekil-1'de gösterilmiştir.



Şekil 1: YSA modeli

Bu şekil Statistica 6.0 istatistiksel paket programında, YSA uygulaması sonucunda elde edilmiştir. Train Perf.: Eğitim setinin performansı, Select Perf.: Geçerlik setinin performansı, Test Perf.:Test setinin performansı.

Şekil-1’de gösterilen yapay sinir ağında girdi katmanında 11 adet, gizli katmanlarından birincisinde 10, ikincisinde 9 adet ve çıktı katmanında 1 adet nöron bulunmaktadır. Gizli ve çıktı katmanlarında kullanılmış olan sinaptik fonksiyonların tümü doğrusaldır. Ayrıca birinci ve ikinci gizli katmanların her ikisinde de hiperbolik tanjant aktivasyon fonksiyonları, çıktı katmanında ise lojistik aktivasyon fonksiyonu kullanılmıştır.

Şekil-1’de gösterilen yapay sinir ağı, danışmanlı öğrenme yöntemlerinden olan geriye yayılım algoritması ile eğitilmiştir. Eğitim esnasında eşlenik eğitim azalış eğitim algoritması yöntemi kullanılmıştır. Kurulan model yardımıyla eğitim, geçerlilik ve test setinden elde edilen doğru sınıflandırma oranları Tablo-9’da verilmiştir.

Tablo-9: Eğitim, geçerlilik ve test setinden elde edilen doğru sınıflandırma oranları

	Eğitim Seti	Geçerlilik Seti	Test Seti
Sınıflandırma Oranı (%)	100	100	100

Ağın eğitimi için 1000 iterasyon gerçekleştirilmiştir. Yapay sinir ağları analizi sonucunda doğru sınıflandırma oranı % 100 olarak bulunmuştur. Sınıflandırma tablosu Tablo-10'da gösterilmiştir.

Tablo-10: Yapay sinir ağları analizi sonucunda elde edilen doğru sınıflandırma oranları.

			Tahmin Değerleri			
			Durum			
			Kronik hepatit	Siroz	Toplam	Doğruluk
Gerçek Değerler	Durum	Siroz	85	0	85	%100
		Kronik hepatit	0	79	79	%100
Toplam			85	79	164	%100

Tablo-10'da görüldüğü gibi, uygulamadaki 164 birimden gerçekte 85 tanesi kronik hepatit, geriye kalan 79 hasta ise sirozdur. Kronik hepatit olan 85 hastanın 85 tanesi yani tümü yapay sinir ağları ile yapılan sınıflandırma işleminde doğru olmak üzere % 100'lük bir doğruluk yüzdesi ile sınıflandırılmıştır. Siroz olan 79 hastanın ise 79 tanesi yapay sinir ağları ile yapılan sınıflandırma işleminde doğru olmak üzere % 100'lük bir doğruluk yüzdesi ile sınıflandırılmıştır.

Yapay sinir ağları ile yapılan sınıflandırma işleminde genel doğruluk değeri 164 hastanın 164 tanesi doğru sınıflandırılarak % 100 olarak hesaplanmıştır.

TARTIŞMA VE SONUÇ

Diskriminant analizi, arařtırmacılar ve uygulamacılar tarafından sınıflandırma modelleri geliřtirmede kullanılan ilk çok deęişkenli istatistiksel sınıflandırma yöntemi olmuştur. Yapay sinir aęı modelleri ise son yıllarda farklı alanlarda yapılan arařtırmalarda sıklıkla kullanılmış ve karmaşık problemlerin çözümündeki yetenekleri ortaya koyulmuştur.

Bu çalışmada yapay sinir aęı modellerine deęinilmiş, bazı istatistiksel tekniklerle bazı yapay sinir aęı modelleri arasındaki iliřkiler incelenmiş ve saęlık alanından elde edilmiş gerçek bir veri seti üzerinde uygulama yapılmıştır. Bu veri seti, önce DDA, daha sonra ise yapay sinir aęı modelleriyle çözümlenmiştir. Çözümleme sonrasında yapay sinir aęı modellerinin sınıflandırma problemlerindeki etkinlięi arařtırılmıştır.

Siroz ve kronik hepatit tanısı almış 164 hastadan elde edilen veri setine DDA uygulanırken baęımlı deęişken olarak, hastaların siroz ya da kronik hepatit olma durumunu gösteren ‘durum’ deęişkeni, baęımsız deęişkenler olarak ise yaşı, lökosit sayısı, platelet sayısı, sodyum, albumin/globulin oranı, total bilirubin, laktat dehidrogenaz, AST/ALT oranı, protrombin zamanı, INR, serum protein elektroforezinde gamaglobulin fraksiyonu deęişkenleri dahil edilmiştir.

DDA sonucunda Wilks’ lambda deęeri 0,238 olarak, Ki kare deęeri 224, 716 olarak elde edilmiş ve modelin istatistiksel olarak anlamlı olduęu görülmüştür ($p < 0,05$). Kronik hepatit olan 85 hastanın 79 tanesi DDA ile yapılan sınıflandırma işleminde doęru, 4 tanesi ise hatalı olmak üzere % 92,94’lük bir doęruluk yüzdesi ile sınıflandırılmıştır. Siroz olan 79 hastanın ise 75 tanesi DDA ile yapılan sınıflandırma işleminde doęru, 6 tanesi ise hatalı olmak üzere % 94,94’lük bir doęruluk yüzdesi ile sınıflandırılmıştır. Elde edilen modelin genel doęru sınıflandırma yüzdesinin % 73,3 olduęu görülmüştür.

DDA’ inde kullanılan baęımsız deęişkenlerin aynısı YSA’ ında girdi birimi olarak, baęımlı deęişken ise çıktı birimi olarak kullanılmıştır. YSA’larından girdi katmanında 11 birim, ilk gizli katmanında 7, ikinci gizli katmanında 13 birim olmak üzere iki adet gizli katmana sahip, çıktı katmanında 1 birim bulunan çok katmanlı perseptron kullanılmıştır. Simülasyon yapılarak elde edilen 1000 adet yapay sinir aęı modelinden en iyisi seçilmiştir.

Birinci ve ikinci gizli katmanların her ikisinde de hiperbolik tanjant, hiperbolik tanjant aktivasyon fonksiyonları, çıktı katmanında ise lojistik aktivasyon fonksiyonu kullanılan ve danışmanlı öğrenme yöntemlerinden olan geriye yayılım algoritmalarından eşlenik eğim azalış algoritması kullanılarak eğitilen sinir aęı sonucunda elde edilen

modelde kronik hepatit olan 85 hastanın 85 tanesi ve siroz olan 79 hastanın 79 tanesi doğru sınıflandırılmıştır. Modelin genel doğru sınıflandırma oranı %100 olarak elde edilmiştir.

Bu sonuçlara göre, bu veri seti için YSA modelleri ile elde edilen sınıflandırmanın DDA ile elde edilenden daha iyi sonuçlar verdiği görülmektedir. Elde edilen sonuçlar daha önce yapılan bu konudaki çalışmaları destekler niteliktedir.

Gastroenteroloji alanında Hamamoto ve arkadaşları, 54 hastadan elde ettikleri klinik bilgileri kullanarak YSA modeli oluşturmuşlar ve test setinin hepatit hastalarını %100 doğrulukla sınıflandırdığı sonucunu elde etmişlerdir (42). Kardiyolojide akut miyokard enfarktüs tanısı için yapay sinir ağlarında serum enzim düzeyleri giriş birimleri olarak kullanılmış ve tanısal doğruluğun, %8 yanlış pozitiflik oranıyla %100 olarak elde edildiği görülmüştür (43).

Youngoch ve arkadaşları, hisse sentlerinin performansını tahmin etmek için YSA ve diskriminant analizin uygulamışlar ve YSA'nın diskriminant analizinden daha yüksek doğrulukta bir sınıflandırma gerçekleştirdiğini ortaya koymuşlardır (17). Keogan ve arkadaşları, akut pankreatit hastalarından elde edilen veri seti üzerinde diskriminant analizi ve YSA uygulamışlar ve YSA'nın diskriminant analizinden az bir farkla daha iyi performans gösterdiğini ortaya koymuşlardır (44).

Yapay sinir ağları farklı istatistiksel işlemlerin yerine getirilmesinde önemli bir rol oynamaktadır. Yapay sinir ağı yöntemlerinin genellikle doğrusal olmaması, istatistiksel modellerde, tahmin analizlerinde, sınıflandırma problemlerinde önemini yükseltmektedir.

YSA'larında verilerin eğitimde kullanılması için gerekli olan bir varsayıma rastlanmamıştır. Bir başka deyişle her tür veri, sayılarla kodlanması şartıyla eğitim için kullanılabilir. Bu yapay sinir ağı teknolojisinin bazı varsayımlara dayanan çok boyutlu istatistiksel yöntemlere kıyasla sahip olduğu önemli bir avantajdır.

Bununla birlikte, diskriminant analizinde veri setini ikiye ayırma zorunluluğu yoktur. Ancak yapay sinir ağlarında, aşırı uyumdan kaçınmak için durdurma kriteri olarak verilerin en az % 10'unun test setine ayrılması gerekir. Bu durumda veri setinin hacminin azalmasına neden olur.

Ayrıca diskriminant analizinin uygulanması bilgisayarda kısa bir sürede yapılabilirken, YSA'da veriye en uygun modelin karar verilmesi süreci daha çok zaman gerektirir.

YSA'ları, modele anlamlı etkisi olmayan değişkenlerin elenebilmesi özelliğine sahip olmadıkları gibi, kullanılacak gizli katman sayısı da net olmaktan uzaktır (45).

İstatistiksel yöntemler beraberinde sorun alanına ilişkin anlaşılabilir ve yorumlamaya olanak veren sonuçlar üretmesine rağmen, yapay sinir ağlarının ağırlıklarını yorumlama imkanı henüz oldukça zordur. Bu nedenle yapay sinir ağıyla ulaşılan sonuçlarda model kapalı bir kutu olarak kalmaktadır. Yapay sinir ağları modelin gizli tabaka ve gizli nöron sayısı fazla olduğunda parametre tahminlerinin yorumlanması zorlaşır.

Karmaşık yapısı nedeniyle YSA'nın oluşturulma ve sonuçlarının yorumlanması aşamaları, yeterli bilgi ve deneyime sahip uzman kişiler tarafından yapılmalıdır.

YSA'ında bulunan gizli katmandaki birimlerin, kompleks problemleri çözümedeki katkıları ve önemleri aşikar olmasına rağmen, bunların YSA modelinin performansına olan etkilerinin matematiksel olarak analiz edilmesi için çok az çalışma gerçekleştirilmiştir. İleriki çalışmalarda, YSA'larının kullanımının genişletilmesi için gizli katman ve gizli katmandaki birimlerin optimal sayısını belirleyebilecek yöntemlerin geliştirilmesi gerekmektedir. Ayrıca gelecekteki çalışmalarda, YSA modellerinin doğrusal olmayan, modern çok değişkenli istatistiksel modellerle farklı koşullar altında karşılaştırılması yararlı olacaktır.

KAYNAKLAR

1. MICHIE D, SPIEGELHALTER DJ, TAYLOR CC. Machine learning, neural and statistical classification, Ellis Horwood, New York, page 2-3, 1994.
2. FIX E, HODGES JL. Discriminatory analysis, nonparametric estimation: consistency properties. Report no. 4, project no. 21-49-004, UASF School of Aviation Medicine, Randolph Field, Texas, 1951.
3. BİÇEN P. Veri madenciliği: Sınıflandırma ve tahmin yöntemlerini kullanarak bir uygulama, Yayınlanmamış Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul, sayfa 4-5, 2002.
4. TATLIDİL H. Uygulamalı çok değişkenli istatistiksel analiz, Engin Yayınları, Ankara, sayfa 10-12, 1992.
5. DOUMPOS M. Multicriteria decision aid classification methods, Kluwer Academic Publications, page 15-16, 2002.
6. FISHER RA. The use of multiple measurement in taxonomic problems. Annals of Eugenics, 7: 179-188, 1936.
7. SMITH CAB. Some examples of discrimination. Annals of Eugenics, 13: 272-282, 1947.
8. EFE MÖ. KAYNAK O. Yapay sinir ağları ve uygulamaları, Boğaziçi Üniversitesi Yayınevi, İstanbul, sayfa 1-6, 2000.
9. ABDI H, VALENTIN D, EDELMAN B. Neural networks, Sage Publications, California, page 1-9, 1999.
10. SWINGLER K. Applying neural networks a practical guide, Centre for Cognitive and Computational Neuroscience, University of Stirling, Scotland, page 221-261, 1996.
11. FAUSETT LV. Fundamentals of neural networks, Printice-Hall Inc., New Jersey, 1994.
12. MEMMEDOV M, ERYILMAZ H. Yapay sinir ağları ile bazı istatistiksel modeller arasındaki ilişki. International XII. Turkish Symposium on Artificial Intelligence and Neural Network, 2003.
13. BISHOP M. Neural networks for pattern recognition, Clarendon Press, Oxford, page 77-145, 1995.
14. KEMP RA, MACAULAY C, GARNER D, PALCIC B. Opening the black box: The relationship between neural networks and linear discriminant functions. Analytical Cellular Pathology, 14(1): 19-30, 1997.

15. MORRIS RGM. D.O. Hebb: The organization of behavior, Wiley: New York; 1949.
- Brain Research Bulletin, 50: 437-438, 1999.
16. ÖZTEMEL E. Yapay Sinir Ağları, Papatya Yayıncılık, İstanbul, sayfa 23-47, 2003.
17. YOON Y, SWALES G, MARGAVIOTH. Comparison of discriminant analysis versus neural networks. Journal Operational Research, 44(1): 51-60, 1993.
18. STACEY D. Intelligent systems architecture design techniques, artificial neural networks for intelligent manufacturing, Chapman&Hall, London, page 17-38, 1994.
19. GALLANT SI, Neural network learning and expert systems, MIT Press, London, page 1-14, 1993.
20. HERTZ J, KROGH A, PALMER RG. Introduction to the theory of neural computation, Addison-Wesley Publishing Co., New York, 1993.
21. RUMELHART DE, HINTON GE, WILLIAMS R. Learning internal representations by error propagation, paralel distributed processing, volume 1, MIT Press, Cambridge, 1986.
22. WANG S. An adaptive approach to market development forecasting. Neural Computing & Applications, 8: 3-8, 1999.
23. ANDERS U. Statistical model building for neural networks. 6th AFIR Colloquium, 963-979,1996.
24. WARREN SS. Neural networks and statistical models. Proceeding of the Nineteenth Annual SAS Users Group International Conference, Sas Institute, 1538-1550, 1994.
25. EDİZ B. Lojistik Regresyon-Ayrırma Analizi, Ayrımsama Sorunu ve Kalp Hastalarında Lojistik Model Yardımıyla Risk Ölçütlerinin Belirlenmesi, Yayınlanmamış Doktora Tezi, Uludağ Üniversitesi Sağlık Bilimleri Enstitüsü, Bursa, sayfa 5-42, 1997.
26. KLECKA WR. Discriminant analysis, Sage Publications, London, page 7-56, 1987.
27. FISHER RA. The use of multiple measures in taxonomic problems. Annals of Eugenics, 7: 105-109, 1936.
28. WELCH BL. Note on discriminant functions, Biometrika, 31: 218-220, 1939.
29. SEBER GAF. Multivariate observations, John Wiley and Sons, New York, page 279-342, 1984.
30. NORUSIS MJ. SPSS 13.0 Advanced statistical procedures companion, Prentice Hall, 2006.
31. POUISEN J, FRENCH A. Discriminant function analysis, Statsoft Electronic textbook, StatSoft, 2003.
32. <http://www2.chass.ncsu.edu/garson/pa765/discrim.htm>
33. WILKS SS. Biometrika, 24: 471-494, 1932

34. GAU GV. A note on the assesment of the results in a discriminant analysis, *Decisions Sciences*, 9: 341-342, 1978.
35. LACHENBRUCH PA, GOLDSTEIN M. Discriminant analysis. *Biometrics*, 35(1): 69-85, 1979.
36. TIMM NH. *Applied multivariate analysis*, Springer-Verlag, New York, page 424-426, 2002.
37. ANDERSON TW. *An introduction to multivariate statistical analysis*, John Wiley and Sons, New York, page 2-69, 1984.
38. ATAKAN C. Diskriminasyon ve hata oranları tahmini, *Yayınlanmamış Doktora tezi*, Ankara Üniversitesi Fen Bilimleri Enstitüsü, sayfa 9, 1996, Ankara.
39. GILBERT ES. The effect of unequal variance-covariance matrices on Fisher's linear discriminant function. *Biometrics*, 25(3): 505-515, 1969.
40. FIELDMAN JH. Regularized discriminant analysis. *Journal of the American Statistical Association*, 84: 165-175, 1989.
41. MARDIA KV. Measures of multivariate skewness and kurtosis with Applications. *Biometrika*, 57: 519-530, 1970.
42. HAMAMOTO I, OKADA S, HASHIMO T, WAKABAYASHI H, MAEBA T, MAETA H: Prediction of the early prognosis of the hepatectomized patient with hepatocellular carcinoma with a neural network. *Computational Biological Medicine*, 25(1): 49-59, 1995.
43. BAXT WG. Use of an artificial neural network for the diagnosis of myocardial infarction. *Annals of Internal Medicine*, 115(11): 843-8, 1991.
44. KEOGAN MT, LO JY, FREED KS, RAPTOPOULOS V, BLAKE S, KAMEL IR, WEISINGER K, ROSEN MP, NELSON RC. Outcome analysis of patients with acute pancreatitis by using an artificial neural network. *Academic Radiology*, 9: 410-419, 2002.
45. Statsoft Electronic textbook, StatSoft, 2003.
<http://www.statsoft.com/textbook/stathome.html>

TEŐEKKÜR

Yüksek lisans tez konusunun seçiminden tezin tamamlanma aşamasına kadar olan yardımlarından ve desteklerinden dolayı danışman hocam Yrd. Doç. Dr. Bülent EDİZ'e, tezimin başından bu yana bana verdiği desteklerinden dolayı Anabilim Dalı Başkanımız Prof. Dr. İsmet KAN'a, tezimin gerçekleştirilmesindeki desteklerinden dolayı Uzm. Dr. İlker ERCAN'a ve çalışma arkadaşlarıma, verilerini bize açan ve desteklerini esirgemeyen Uludağ Üniversitesi Tıp Fakültesi Gastroenteroloji Bilim Dalı öğretim üyesi Prof. Dr. Selim Giray NAK ve Yrd. Doç. Dr. Murat KIYICI'ya, verdiği teknik ve manevi desteğinden dolayı Anadolu Üniversitesi İstatistik Bölümü'nden değerli meslektaşım Arş. Gör. Şenay YOLAÇAN'a ve bugüne kadar desteklerini her zaman yanımda hissettiğim sevgili aileme teşekkürlerimi bir borç bilirim.

ÖZGEÇMİŞ

1981 yılında Bursa'da doğdum. İlkokulu 1993 yılında Yenişehir Aydoğdubey İlkokulu'nda bitirdim ve aynı yıl İnegöl Turgutalp Anadolu Lisesi'ne girdim. Orta öğrenimimi burada tamamladım. Lise öğrenimime, yine aynı okulda başlayıp, 1999 yılında Yenişehir Osmangazi Lisesi'nde tamamladım. 1999 yılında Kırıkkale Üniversitesi Fen-Edebiyat Fakültesi İstatistik Bölümüne girdim. Yatay geçiş yaptığım Eskişehir Anadolu Üniversitesi Fen Fakültesi İstatistik Bölümünden 2003 yılında mezun oldum. 2003 yılı Eylül ayında Uludağ Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı'nda yüksek lisansa başladım. 2004 yılı Ocak ayında Uludağ Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı'nda araştırma görevlisi olarak çalışmaya başladım.