

UÇUCU GAZLARI ALGILAYICI QCM SENSÖR DİZİ VERİLERİNİ BİLGİSAYAR  
ORTAMINDA TANI YÖNTEMLERİ

Mustafa İlhami Cusundaş

Dumlupınar Üniversitesi  
Fen Bilimleri Entitüsü  
Lisansüstü Yönetmeliği Uyarınca  
Elektrik-Elektronik Anabilim Dalında  
YÜKSEK LİSANS TEZİ  
Olarak Hazırlanmıştır.

Danışman : Yrd. Doç. Dr. Mehmet Ali EBEOĞLU

**Y.C. YÜKSEKÖĞRETİM KURULU  
DOKÜMANTASYON MERKEZİ**

Ağustos - 2002

121679  
121679

**KABUL VE ONAY SAYFASI**


M. İlhami Cusundaş'ın YÜKSEK LİSANS tezi olarak hazırladığı Uçucu Gazları Algılayıcı QCM Sensör Dizi Verilerini Bilgisayar Ortamında Tanı Yöntemleri başlıklı bu çalışma, jürimizce lisansüstü yönetmeliğinin ilgili maddeleri uyarınca değerlendirilerek kabul edilmiştir.

..05.10.2002

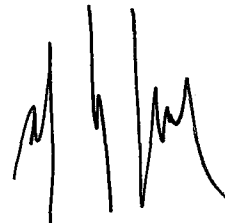
Üye: Prof. Dr. Zafer Ziya Öztürk

Üye: Yrd. Doç. Dr. M. Ali

Üye: Yrd. Doç. Dr. Ahmet ÖZMEN



Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 12.10.2002 gün ve .....14..... sayılı kararıyla onaylanmıştır.



Prof. Dr. I. Göktay EDİZ  
Fen Bilimleri Enstitüsü Müdürü

## UÇUCU GAZLARI ALGILAYICI QCM SENSÖR DİZİ VERİLERİNİ BİLGİSAYAR ORTAMINDA TANI YÖNTEMLERİ

Mustafa İlhami Cusundaş

Elektrik-Elektronik Mühendisliği Yüksek Lisans Tezi, 2002

Tez Danışmanı: Yrd. Doç. Dr. Mehmet Ali Ebeoğlu

### ÖZET

Kimyasal sensör dizileri; uçucu analitlerin analizinde önemli bir role sahip olmuştur. Değişik diziler; geniş duyarlılık profillerinin çakıştığı veriyi toplarlar. Bunların sağlam veri analizlerine ihtiyaçları vardır, adreslenen problemleri çözmek için sıklıkla numune tanıma metotlarını ilgilendirirler.

İstatistik yaklaşım; geleneksel olarak formüle edilen numune tanımalardaki sayısız yapılar arasında en yoğun çalışılan ve pratikte en çok kullanılanıdır. Daha yakınlarda istatistik öğrenme teorisinden alınan yapay sinir ağı teknikleri ve metotlarında önemli bir artış görünmüştü. Bir tanıma sisteminin tasarımında şu noktalara dikkat edilir: numune sınıflarının tanımlaması, çevresel algılama, numune tanıtılması, özellik çıkarımı ve seçimi, küme analizi, sınıflandırıcı tasarımı ve öğrenmesi, eğitime ve test örneklerinin seçimi ve performans değerlendirmesi.

Bu alanda, yaklaşık 50 yıllık araştırma ve gelişmeye rağmen, kompleks numuneleri keyfi yönlendirmeye, yerleştirme ve ölçü ile tanımlama genel problemi hala çözülememiştir.

Bu çalışmada; uçucu analitlerin algılanması ve tanınması aşamalarında kullanılan, iyi bilinen metotları gözden geçirmek ve karşılaştırmak amacımızdır.

**Anahtar Kelimeler:** Hata Tahmini, Kimyasal Sensör Dizisi, Numune Tanıma, Özellik Seçimi Sınıflandırma, Sinir Ağları.

# COMPUTERISED RECOGNITION METHODS OF DATA OF QCM SENSOR ARRAY SENSING VOLATILE GASES

Mustafa İlhami Cusundaş

Electric-Electronic Engineering Ms. thesis, 2002

Supervisor: Yrd. Doç. Dr. Mehmet Ali Ebeoğlu

## SUMMARY

Chemical sensor arrays have come to have an important role in the analysis of volatile analytes. Such arrays gather data which have broad overlapping sensitivity profiles, which require substantial data analysis, often involving pattern recognition methods to solve the problems being addressed.

Among the various frameworks in which pattern recognition has been traditionally formulated the statistical approach has been most intensively studied and used in practice. More recently, neural network techniques and methods imported from statistical learning theory have been receiving increasing attention. The design of a recognition system requires careful attention to the following issues: definition of pattern classes, sensing environment, pattern representation, feature extraction and selection, cluster analysis, classifier design and learning, selection of training and test samples, and performance evaluation.

In spite of almost 50 years of research and development in this field, the general problem of recognizing complex patterns with arbitrary orientation, location, and scale remains unsolved.

In this study; our goal is to look over and compare of the well known methods used in recognition and perceive stages of volatile analytes.

**Keywords:** Error Estimation, Chemical Sensor Array, Pattern Recognition, Feature Selection, Classification, Neural Network.

## TEŐEKKÜR

Çalıőmalarım sırasında çok yakın ilgi ve yardımlarını gördüğüm; başta danışman hocam sayın Yrd. Doç. Dr. Mehmet Ali Ebeođlu'na, TÜBİTAK Marmara Araőtırma Merkezi Malzeme ve Kimya Teknolojileri Araőtırma Enstitüsü araőtırmacılarından Cihat Tasaltın'a, TÜBİTAK Marmara Araőtırma Merkezi Biliőim Teknolojileri Araőtırma Enstitüsü araőtırmacılarından Zeki Erdem'e ve hiçbir zaman maddi ve manevi desteklerini esirgemeyen sevgili aileme sonsuz teőekkür ederim.



## İÇİNDEKİLER

	<u>Sayfa</u>
ÖZET .....	iv
SUMMARY .....	v
TEŞEKKÜR .....	vi
ŞEKİLLER DİZİNİ .....	ix
ÇİZELGELER DİZİNİ .....	x
SİMGELER VE KISALTMALAR DİZİNİ .....	xi
1. GİRİŞ .....	1
2. ELEKTRONİK BURUN .....	3
3. KİMYASAL SENSÖR DİZİLERİ .....	5
3.1. Metal Oksit Yarıiletkenleri (MeOX) (MeOx) .....	5
3.2. Metal Oksit Yarıiletken Alan Etkili Transistörleri .....	5
3.3. Kuartz Kristal Mikrobalans Sensörleri (QCM) .....	6
3.4. Yüzey Akustik Dalga Cihazları (SAW) .....	6
3.5. Polimerler-Kimyasal Dirençler .....	7
3.6. Optik Sensörler .....	7
4. NUMUNE TANIMA .....	8
4.1. Düzenleme (Template Matching) .....	10
4.2. İstatistik Yaklaşım .....	10
4.3. Sentaktik Yaklaşım .....	11
4.4. Sinir Ağları .....	11
4.5. Bulanık Sinir Ağları .....	12
5. İSTATİSTİK NUMUNE TANIMA .....	13
6. BOYUTLANABİLİRLİK .....	17
6.1. Boyutlanabilirlik İndirgenmesi .....	18
6.1.1. Özellik çıkarımı .....	20
6.1.2. Özellik seçimi .....	23

## İÇİNDEKİLER (devam)

	<u>Sayfa</u>
7. SINIFLANDIRICILAR .....	27
7.1. Sınıflandırıcı Kombinasyonu .....	30
8. HATA TAHMİNİ .....	34
9. DENETLENMEMİŞ SINIFLANDIRMA .....	39
9.1. Kare Hata Kümelenendirme .....	39
9.2. Karışım Parçalaması .....	41
10. KİMYASAL SENSÖR DİZİSİ VERİ ANALİZİ METOTLARI .....	42
10.1. Lineer Ayrım Analizi (LDA) ve Uygulamaları .....	43
10.2. Baş Bileşen Analizi (PCA) ve Uygulamaları .....	47
10.3. Baş Bileşen Regresyonu (PCR) ve Uygulamaları .....	51
10.4. Parçalı En Küçük Kareler (PLS) ve uygulamaları .....	52
10.5. Küme Analizleri .....	54
10.5.1. Haritalama ve gösterme .....	54
10.5.2. Hiyerarşili kümelendirme .....	56
10.5.3. K- en yakın komşuluk sınıflandırması .....	57
10.5.4. Görülebilir olarak – etkinin deneysel bölgesi .....	58
10.6. Besleme Ön Sinir Ağı (FFNN) ve Uygulamaları .....	58
10.7. Kendiliğinden Organize Haritalar (SOM) .....	62
10.8. Öğrenme Vektör Kuantizasyonu (LVQ) .....	63
10.9. Uyumlu Rezonans Teorisi (ART) .....	65
11. SONUÇLAR .....	66
KAYNAKLAR DİZİNİ .....	68
EKLER	
1. Cyranose 320'ye genel bir bakış	
2. Cyranose 320'nin özellikleri	
3. Cyranose 320'nin görünümü	
4. Kompozit malzemenin analite maruz kalma durumunda cevabı	
5. Dizideki 32 adet sensörün cevabı	
6. Cyranose 320'de işlem süreci	
7. Kullanım alanlarının görüntüsü	

**ŞEKİLLER DİZİNİ**

<b><u>Sekil</u></b>	<b><u>Sayfa</u></b>
5.1. İstatistik numune tanıma modeli .....	13
5.2. İstatistik numune tanımda değişik yaklaşımlar .....	16
6.1. Chernoff'un orta vektörlerine bağlı yüzleri .....	19
6.2. Üç boyutlu alt uzayın bulunması için otomatik çağrışımlı ağlar .....	22
8.1. Dijit veri seti üzerinde eğitime ve test setleri için eğitime örneklerinin bir fonksiyonu Bayes plug-in lineer sınıflandırıcısının sınıflama hatası .....	35
8.2. Dijit veri seti için Bayes plug-in lineer sınıflandırıcısının ROC eğrisi .....	38
8.3. Dijit veri seti için Bayes plug-in lineer sınıflandırıcısının hata ret eğrisi .....	38
9.1. Karışım parçalama örneği .....	41
10.1. Sensör dizisi veri analizi metotlarının gruplaşma yolları .....	42
10.2. Bir Perceptron'un katman yapısı .....	59
10.3. Genel bir işlem elemanın yapısı .....	60



**ÇİZELGELER DİZİNİ**

<b><u>Çizelge</u></b>	<b><u>Sayfa</u></b>
4.1. Numune tanıma uygulamalarının örnekleri .....	8
4.2. Numune tanıma modelleri .....	10
4.3. İstatistik ve sinir ağları metotları arasındaki ilişkiler .....	12
5.1. Bazı veri ön işleme ve normalizasyon tekniklerinin tanımları .....	14
6.1. Özellik çıkarma ve yansıtma metotları .....	23
6.2. Özellik seçme metotları .....	25
7.1. Sınıflandırma metotları .....	29
7.2. Sınıflandırıcı kombinasyon şeması .....	33
8.1. Hata tahmin metotları .....	36



## SİMGELER VE KISALTMALAR DİZİNİ

<u>Simgeler</u>	<u>Açıklama</u>
$\Delta f$	Frekans değişimi
$f_0$	Piezoelektrik kristalin temel rezonans frekansı
$P_e( )$	Hata olasılığı
$X$	$n \times d$ numune matrisi
$Y$	$n \times m$ numune matrisi
$H$	$d \times m$ transformasyon matrisi
$K(X, X)$	Kernel matrisi
$J( )$	Özellik seçimi kriter fonksiyonu
$\tau$	Yanlış sınıflandırılmış test numune sayısı
$\hat{P}_e$	Maksimum olasılık tahmini
$E_K^2$	Kare hatası
$X_A^*$	Orta vektör
$S_A$	Kovaryans matrisi
$S_p$	Birleştirilmiş kovaryans matrisi
$U$	$m \times n$ ortonormal matrisi
$V$	$n \times n$ ortonormal matrisi
$\lambda_i$	Öz değer
$\text{var}_i$	$i$ . öz değere katkıda bulunan varyans yüzdesi
$d_y$	İki numune arasındaki mesafe
$P$	Ayarlanabilir parametre sayısı
$I$	Giriş katmanındaki ünite sayısı
$H$	Saklı katmandaki nöron sayısı
$O$	Çıkış katmanındaki ünite sayısı
$\eta$	Öğrenme oranı

## SİMGELER VE KISALTMALAR DİZİNİ (devam)

<u>Kısaltmalar</u>	<u>Açıklama</u>
Q.C.M	Quartz Crystal Microbalance (Kuartz Kristal Mikrodengeleyici)
S.A.W	Surface Acoustic Wave (Yüzey Akustik Dalga)
N.N	Neural Network (Sinir Ağı)
F.N.N	Fuzzy Neural Network (Bulanık Sinir Ağı)
M.D.S	Multi Dimensional Scaling (Çok Boyutlu Ölçekleme)
S.F.S	Sequential Forward Selection (Sıralı İleri Seçim)
S.B.S	Sequential Backward Selection (Sıralı Geriye Seçim)
S.F.F.S	Sequential Forward Floating Search (Sıralı İleri Yüzdürme Araştırması)
S.B.F.S	Sequential Backward Floating Search (Sıralı Geri Yüzdürme Araştırması)
F.F.N.N	Feed Forward Neural Network (İleri Beslemeli Sinir Ağı)
F.A.R	False Acceptance Rate (Hata Kabul Oranı)
F.R.R	False Reject Rate (Hata Ret Oranı)
P.C.R	Principal Component Regressions (Baş Bileşen Regresyonu)
P.L.S	Partial Least Squares (Parçalı En Küçük Kareler)
L.D.A	Linear Discriminant Analysis (Lineer Ayrım Analizi)
P.C.A	Principal Component Analysis (Baş Bileşen Analizi)
B.P	Back Propagation (Geri Yansıma)
L.V.Q	Learning Vector Quantization (Öğrenme Vektör Kuantizasyonu)
A.R.T	Adaptive Resonance Theory (Uyumlu Rezonans Teorisi)
S.O.M	Self Organizing Maps (Kendinden Organizeli Haritalar)
C.V	Cross Validation (Çapraz Geçerlilik)

## 1. GİRİŞ

Bu çalışmanın amacı; uçucu organik bileşiklerin algılanması ve tanınması aşamalarında kullanılan, iyi bilinen metotların bazılarını özetlemek ve karşılaştırmaktır.

Numune tanımlama ve numuneleri gruplandırma; biyoloji, psikoloji, tıp, yapay zeka ve uzaktan kontrol gibi mühendislik ve bilimsel disiplin türleri arasında en önemli problemlerdendir. Sözü edilen numune; koku, parmak izi görüntüsü, el yazısı bir kelime, bir insan yüzü veya konuşma sinyali olabilir.

Numune tanımanın amacı denetlenmiş veya denetlenmemiş sınıflandırmadır. Denetlenmiş sınıflandırma; giriş numunesinde önceden tanımlanmış bir sınıfın üyesi olarak tanımlanır. Denetlenmemiş sınıflandırmada; numune bilinmeyen bir sınıfa atanır.

Bu çalışmada söz konusu numuneler genellikle uçucu organik bileşiklerdir. Bu yüzden, uçucu organik bileşiklerin analizinde önemli bir role sahip olan kimyasal sensör dizilerinden bahsedilmiştir. Değişik kimyasal sensör dizileri; uçucu analitlerle karşı karşıya iken duyarlılık profillerinin çıktığı veriyi toplarlar. Toplanan kimyasal sensör dizisinin cevap numunesi, analitik problemi çözmek için hesaplama metotları ile birlikte kullanılır. Sensör deneyleri ve deneyleri desteklemek amacıyla kullanılan hesaplama metotları, nitelik veya nicelik analizleri için kullanılabilir. Nitelik analizi; analitin varlığını veya yokluğunu veya önceden belirlenmiş eşığı aşan analitlerin varlığını araştırma anlamına gelir. Bu aynı zamanda, hangi veya kaç tane sensörün kullanılacağını belirlemede de müracaat yeridir. Niceliksel analizler; konsantrasyonların, aktivitelerin veya analitlerin kısmi basınçlarının ölçülmesi, fermantasyon zamanı gibi parametrelerin tahmini anlamına gelir.

Bünyesinde numune tanıma yazılımı ve kimyasal sensör dizisi bulduran elektronik burunlar, insan beynindeki algılama yöntemine benzer bir yöntem kullanarak koku örneklerini hisseder ve onları ayırır. Elektronik burunlar kimyasal sensör dizilerini kapsadığı için kokuya tepki olarak zamana bağlı veri sinyallerini üretirler. Dinamik sensör sinyali sensör dizisine kaynaktan kokuyu taşıyan akışın hızı, kokunun kimyasal doğası, aktif sensör maddeleriyle kokunun reaksiyonu ve difüzyonu ve basınç, ısı, nem gibi bir çok fiziksel parametreye bağlıdır. Bu tür kompleks bilgilerin işlenmesi için sofistike bir yazılım gereklidir. Genelde bu tür yazılımlar numune tanımlaması için güçlü bir algoritmaya dayanır. Sıkça kullanılan algoritmalar; baş bileşen analizi, küme analizi ve sinir ağlarına dayanan çok katmanlı perceptronları içerirler.

Bu alıřmanın birinci blmnde, elektronik burun tanıtılmıřtır. İkinci blmde, hesaplamalı metotların ve bunların uygulamalarının tartıřılmasında hazırlık saęlayan bakıřı vermeyi amalayan sensr tipleri incelenmiřtir. Daha sonraki blmlerde ise sırasıyla; numune tanıma yaklařımları, istatistik numune tanıma, boyutlanabilirlik, sınıflandırıcılar, hata tahmini ve denetlenmemiř sınıflandırma konuları incelenerek, kimyasal sensr dizilerinin kullanıldıęı, ana veri analizi yaklařımları sistematik tanımı, her bir metot iin; yaklařımın, uyumluluęunun, bugne dek grlen uygulamaların temeli sunulmuřtur. Son olarak da 11. blmde kimyasal sensr dizisi veri analizi metotlarının genel bir karřılařtırılması yapılmıřtır.



## 2. ELEKTRONİK BURUN

Gıda, ilaç ve kozmetik endüstrisi için, üretilecek olan ürünün kokusu büyük önem taşır. Bu nedenle yeni ürünlerin geliştirilmesinde ve üretimde uzman kişilerce yürütülen koku analizleri yapılmaktadır. Uygulamada bu tür analizler; şayet üretim kontrol amacıyla kullanılırsa pahalı ve problemlidir. Bu yüzden elektronik burunlar geliştirilmektedir. Elektronik burun tasarımları güçlü, duyarlı sensör kombinasyonları ve sofistike yazılımlar içerir. Bunların işleyişleri insanların koku tanımlamasına benzerdir. Bu tür elektronik burunlar sayesinde güvenilir koku ölçüleri yapılabilmektedir. Singh ve arkadaşlarının çalışmaları [1] buna iyi bir örnek teşkil eder.

Günümüzde koku analizleri, gaz algılayıcı sensör dizileriyle yapılabilmekte, koku analizleri basit, hızlı ve objektif olmaları sebebiyle birçok yiyecek, kozmetik ve ambalaj sektöründe kalite kontrol, benzer ürün sağlama, bozulmanın belirlenmesi, katkı miktarının belirlenmesi gibi farklı birçok uygulamada kapsamlı bir şekilde kullanılmaktadır. Ön bir işleme gerek duymadan elektronik burunlar kompleks bir uçucunun içeriğini belirleyebilirler. Bunun için yarı spesifik sensörler kullanılır. Elektronik burun koku bileşenlerini genel olarak eğiterek tanımlar. Eğitim ve tanıma koku izleri veri tabanında ayrışık olarak depolanır ve bu bilgiler insan beynindeki algılama yöntemine benzer bir yöntem kullanılarak işlenir.

Elektronik burun dört temel birimden oluşur.

- a) Bir örnekleme sistemi
- b) Koku sensör düzeneği
- c) Elektronik veri elde etme kontrol sistemi
- d) Numune tanımlama yazılımı.

Örnekleme sistemi kokuyu yada örneği sensör dizisine tanıtır ve bunu tekrar, tekrar yapabilecek şekilde tasarlanmıştır. Diğer birimler ileri ki başlıklarda verilmiştir.

Kalite kontrol, benzer ürün de dahil olmak üzere elektronik burun birçok alanda kullanılmaktadır [2]. Elektronik burunlar canlı koku denetleyicileri gibi hem nicelik hem de nitelik ölçüsü olarak kullanılabilir. Niceliksel uygulamalar sensör değer bağıntısını, yiyecek ve kozmetik ürünlerinde kullanılan maddelerin incelenmesi ve kokularının belirlenmesi gibi işlemleri de içermekte, nitelik uygulamasını inceleyen sensör uygulamaları da kaynağın belirlenmesi, hammadde kalitesinin belirlenmesi, ürünün tutarlılığı ve kusursuz ürünlerin

oluřturulması gibi iřlemleri de kapsamaktadır. Bunlara ilave olarak raf mr, paketleme ve yařlanma sonrası rnlerdeki benzerlik gibi konularda da kullanılmaktadır [2,3].

Elektronik burunların; hızlı ve objektif sensr lmleriyle, gıda, ila ve kozmetik rnlerinde uygunluęunu kanıtlamıř olup, niteliksel veya niceliksel arařtırmalarda geniř aplı uygulama saharına sahip oldukları grlmektedir.



### 3. SENSÖR DİZİLERİ

Sensör dizisi birden fazla sensörden oluşmuş n tane sensör grubudur. Uçucu bir kimyasal bileşen sensör dizisiyle temasa geçince elektriksel niceliği değişmektedir. Bir sensör dizisi, farklı türlerde yada aynı türde olsun, farklı birçok kimyasal ve kokuyu elektriksel nicelikleri değiştirerek tanımlayabilir. Kullanılan sensör dizisi; farklı örnekler için fark edilebilir şekilde farklı olan cevap üretmelidir. Bu nedenle dizi seçimi bireysel sensörler yerine dizinin geniş bir limitine cevap verebilen, cevaplarında özel olmamalıdır. Sensör seçimi uygulamaya göre yapılır. Önce birçok sensörün kullanıldığı elektronik burunlarla bir araştırma yürütülür ve daha sonra gerekli metodun geçerliliği ve uygulanabilirliği saptanır. Farklı sensörlerin verdiği tepkilere bağlı olarak uygun olan sensör bulunmalıdır [1].

Kullanılan sensör dizilerinde farklı yapılar:

#### 3.1. Metal Oksit Yarıiletkenleri (Metal Oxide Semiconductors) (MeOX) (MeOx)

Metal oksit yarıiletken sensörler, yüzeydeki gaz adsorbsiyonu ve sıra yüzey reaksiyonlarının sebep olduğu iletim değişimidir. Sensör; elektrikle ısıtılmış seramik topak üzerine metal oksit malzemesinin ince, gözenekli bir film tortusu bırakması ve yüksek sıcaklıkta tavllanmasıyla yapılır. Bu cihazlar  $\text{SnO}_2$ ,  $\text{ZnO}$ ,  $\text{In}_2\text{O}_3$ ,  $\text{WO}_3$ ,  $\text{Fe}_2\text{O}_3$ ,  $\text{Ga}_2\text{O}_3$ ,  $\text{TiO}_2$  ve platin veya palladium gibi katalitik metallerin katılanmasıyla üretilirler. Sistemde direnci artıran, yarıiletken malzemenin iletim bandındaki elektronların yoğunluğunu azaltarak, sensör yüzeyi oksijen adsorbe ederek, algılayıcı malzemenin elektriksel iletimini azaltır. Sistemde 100 ile 600 °C arasında artırılan sıcaklık, yüzey reaksiyonlarının hızlanmasından ve kimyasal sistemdeki adsorbe edilmiş suyun çıkış bilgisine girişmesinden emin olmak için kullanılır. Çok yaygın kullanılan bir sensör ( $\text{SnO}_2$ ); elektrik bağlantılarına da sahip ısıtılmış bir seramik taban üstünde sinterlenmiş kalay oksit yapısıdır. Bu basit sensör dizileri, çapraz-reaktif sensör dizisi (cross-reactive sensor array) olarak kullanılabilir. Metal oksitlerin avantajı, gazlara karşı duyarlılıklarından dolayıdır.

#### 3.2. Metal Oksit Yarıiletken Alan Etkili Transistörleri (Metal Oxide Semiconductor Field Effect Transistors) (MOSFET)

MOSFET sensörü katılanmış bir yarıiletkenle katalitik metal kaplı bir izolatörden oluşur. Gaz, molekülleri katalitik yüzeyde tepki verirken, cihazın çıkış sinyali elektrik polarizasyonu sebebiyle sensördeki potansiyel değişimden FET'in  $I_d$  akım değişimleridir. Bu sensörler 100 ile 200 °C arasındaki sıcaklıklarda çalışırlar.



### 3.3. Kuartz Kristal Mikrobalans Sensörleri (Quartz Crystal Microbalance Devices) (QCM)

QCM sensörleri adsorbe edilen moleküllerin kütesini ölçer. Aynı zamanda hacim akustik dalga sensörleri (BAW) ve kalınlık kayma mod sensörleri olarak da bilinirler. Aktif element, temel rezonans frekanslı bir piezoelektrik kristalidir. Kütle ilave edildiğinde yayının rezonans frekansı azalır. Frekans değişimi Sauerbrey eşitliği ile;

$$\Delta f = - (1/\rho_m k_f) f_0^2 (\Delta m / A)$$

$$\Delta f = - 2.3 \times 10^6 f_0^2 (\Delta m / A) \quad (1)$$

$\Delta f$ ; adsorbe edilen kütleden dolayı frekanstaki değişim,  $f_0$  ise piezoelektrik kristalinin temel rezonans frekansıdır.  $\Delta m$ ; yüzey üzerinde biriken kütledir.  $A$  ise sensör alanıdır. Bu yüzden ölçülmüş frekansta ve cihazın kütle yükünde tahmin edilen lineer bir değişme mevcuttur.

Kuartz kristal osilatör yüzeyi; algılayıcı kimyasal, absorbe edici maddenin ince bir filmiyle kaplıdır. Kullanılan malzemeler; ince polimer filmler, tek tabaka filmler ve yüzeye bağlı moleküllerdir. Moleküller absorbe edilirken, rezonans frekansı artan kütle yüzünden azalır. Frekans değişimi, sistemin çıkış sinyali olarak kullanılabilir.

QCM cihazının zaman sabitleri, saniyelerin veya daha uzun sürelerin onlu sırasıyla kayıt edilmiştir. Sistemin diğer karakteristikleri; yüksek duyarlılık, analit konsantrasyonunun bir fonksiyonu olarak cevabın lineerliği, yeniden üretilebilirlik, kararlılık, küçültülebilme, oda sıcaklığı operasyonu ve düşük güç tüketimidir.

### 3.4. Yüzey Akustik Dalga Cihazları (Surface Acoustic Wave Devices) (SAW)

SAW cihazları; kuartz gibi piezoelektrik maddeler üzerinde dijite edilmiş elektrotların şekillendirilmesiyle üretilir. Seçici absorbe edici maddenin ince film kaplaması yüzey üzerinde biriktirilir. Rayleigh yüzey akustik dalgasını üreten bir radyo frekansı uygulanır. Moleküllerin yüzey üzerinde absorbe edilmesi, kütleyi artırır, dalga frekansını değiştirir. Sistemi belirleyici eşitlik, QCM için olan akustik ince katı filmleriyle aynı Sauerbrey eşitliği ile tanımlanır. Kütle birikmesinde sıcaklık kontrolü önemlidir. QCM ile birlikte cihazın cevabı, farklı zar özelliklerini kullanarak değiştirilebilir. Kullanılan kaplamalar; polimerler, lipidler ve kendiliğinden kurulan tek tabakalardır (self-assembled mono-layers). SAW cihazları birkaç yüz MHz frekansta çalıştırılır ve çok küçük yüzeylere sahiptirler. Her ikisi birden QCM cihazlarına kıyasla SAW için daha yüksek hassasiyet ve daha hızlı cevap özelliklerine sahiptir.

SAW cihazlarının QCM'e ek avantajları da vardır. Yeniden üretilebilir karakteristiklerinden başka düşük maliyet ve imalatta fotolitografi metodu kullanılarak minyatürize edilebilirlik. Fakat neme karşı duyarlılıklarıdır.

### 3.5. Polimerler-Kimyasal Dirençler (Conducting Polymers- Chemiresistors)

Direnç ölçümü üzerine kurulu gaz sensörleri, yaygın olarak kullanılmakta olan ince film polimerlerdir. Polipirol, bu sensörlerde en yaygın olarak kullanılan malzemedir. Polimerler; polimerin yapısı boyunca genişleyen, konjuge edilmiş  $\pi$  elektron sistemine sahiptirler. Buharların emilmesi, iç zincir sekmesini, ana zincir iletimini, sayıcı elektrostatikleri ve bunların her birinin iletimini değiştirir. Monomerik değişimlerin farklı sınıfları üzerine kurulmuş polimerler bu sensörlerde kullanılmaktadır. Ölçülen genel özellik; uçucu moleküllerin polimerle ilişki içindeyken polimer direncindeki değişimdir. Değişen yapılar veya fonksiyonel gruplar polimerle birleştirilmiş ve farklı doping yapıcı iyonlar kullanılarak elde edilmektedir. Bunlar seçim ve hassasiyetteki değişimleri belirlemektedir.

İletici polimer sensörler çevre sıcaklığında çalışabilirler. İmalatları kolay olup güç tüketimini de düşürürler. İletici polimer sensörlerin cevap süreleri uzundur. Avantajları; küçük ebat, oda sıcaklığında çalışma ve polar (kutupsal) analitlere hassasiyetidir. Limitleri; sensör fabrikasyon yeniden üretilebilirliği, nem hassasiyeti, zaman üzerine temel çizgi iletiminin yığılması (drift off baseline conductivity over time), cevabın yeniden üretilebilirliğinin noksanlığı, dengeleme duyarlılığı ve ışık hassasiyetidir.

Alternatif polimer kaplı kimyasal rezistör sensör dizisi cihazı, siyah karbon organik polimer kompoziti üzerine kurularak Lewis ve yardımcıları tarafından geliştirilmiştir [4]. İletken bir malzeme olan karbon farklı polimerlerle birleştirilerek yapılabilir. Polimer kimyasal sensörler duyarlı oldukları bir gaz ortamda iken iletkenlikleri ve dielektrik özelliklerinde değişimler oluşmaktadır. Bunun sonucu olarak da sensörün direnci ve kapasitesi algılayıcı gaz nedeniyle büyümektedir.

### 3.6. Optik Sensörler (Optical Sensors)

Optik gaz sensörleri yapay burunlarda da sıkça kullanılmaktadır. Burada ışık kaynağı hem algılanan gazı hem de referans molekülü tahrik eder. Ölçülen sinyal; absorbe, yansıtma, floresanslık veya chemiluminescence özelliklerindeki optik değişimlerdir.

#### 4. NUMUNE TANIMA

Numune tanıma; sensörlerin çevreden nasıl etkilendikleri, girişim faktörlerinden çıkış sinyalinin korunarak numuneleri ayırt etmeyi nasıl öğreneceği ve bunu en doğru yapabilmesidir. Numune sınıflandırma hakkında mantıklı kararlar vereceğinin algoritmasıdır. Fakat uzun yıllar boyunca bu konuda çalışılmasına rağmen elektronik burun numune tanıma uygulamalarında başarılması güç bir hedef olarak kalmıştır.

Numune tanıyıcılar arasında en iyi örnek insanoğludur ve insanların numuneleri nasıl ayırt ettiği henüz çözülememiştir. Bilgisayarlara numune tanıma öğretiler. Bu konuyla ilgili başarılı bilgisayar programları şu işleri yapabilir: Kredi müracaatları için bankalara, teşhis konusunda doktorlara yardımcı olur ve numune tanımasına bağımlı olarak yol izleyen uçaklarda pilotlara yardım eder.

Kimyasal sensör dizileri; numune tanımada özellikle uçucu analitleri analiz etme konusunda önemli bir role sahip olmuştur. Değişik diziler; geniş duyarlılık profillerinin çakıştığı veriyi toplarlar. Bunların sağlam veri analizlerine ihtiyaçları vardır. Bu analizleri genellikle numune tanıma metotları ile gerçekleştiririz [5].

Çizelge 4.1.'de değişik numune tanıma uygulamalarından bazıları verilmiştir.

**Çizelge 4.1.** Numune tanıma uygulamalarının örnekleri

Çalışma Alanı	Uygulama	Giriş Numunesi	Numune Sınıfları
Bioinformatik	Sıra Analizi	DNA/Protein Sırası	Bilgilerin Bilinen Tipleri/Numuneler
Biometrik Tanıma	Kişisel Tanımlama	Yüz, İris, Parmak İzi	Giriş Kontrolü İçin Müsaadeli Kullanıcılar
Konuşmayı Tanıma	Operatör Yardımı Olmadan Telefon Rehberi Soruşturma	Konuşma Dalga Formu	Konuşma Kelimeleri
Doküman Sınıflandırması	İnternet Aramaları	Text Doküman	Semantik Kategoriler(iş,spor gibi)
Multimedya Veri Tabanının Elde Edilmesi	İnternet Aramaları	Video Klip	Video Türü(hareket, diyalog gibi)
Doküman Görüntü Analizi	Körler İçin Okuma Makinası	Doküman Görüntüsü	Alfa nümerik Karakterler,Kelimeler
Endüstriyel Otomasyon	Devrenin Bord'a Basımının Denetimi, CNC	Yoğunluk veya Limit Görüntüsü	Kusurlu veya Kusursuz Ürünün Tabiatı
Uzaktan Algılama	Üretim, Ürün Tahmini	Multispektral Görüntü	Toprak Kullanımı Kategorileri, Ürünlerin Büyüme Numunesi
Data Mining	Anlamlı Numunelerin Aranması	Çok Boyutlu Uzaydaki Noktalar	Sağlam ve İyi Ayrılmış Kümeler

Otomasyon tanıma sistemleri üzerine olan talepler büyük veri tabanlarının ve sıkı performans gereksinimlerinin (hız, doğruluk, maliyet) kullanılabilirliği yüzünden büyük bir ölçek de artmaktadır. Uygulamaların pek çoğunda açıkça görülen; sınıflandırma için tek bir yaklaşım optimal değildir birçok metot ve yaklaşım aynı anda kullanılmak zorundadır. Cyrano Sciences şirketinin geliştirdiği Cyranose 320 isimli elektronik burnun özellikleri ve görünüşü ekte verilmiştir.

Bir numune tanıma sisteminin tasarımında

1-Veriyi elde etme ve ön işleme (data acquisition and preprocessing)

2-Veri temsili (data representation)

3-Karar verme (decision making)

aşamaları bulunmaktadır.

Sensörlerin seçimine, ön işleme tekniğine, temsil şemasına ve karar verme modeline çalışma alanı karar verir. İyi tanımlı ve yeterli sınırlandırılmış numune tanıma uygulaması doğru numune temsilini ve basit karar verme stratejisini yönetmelidir.

Bir numunenin tanınması veya sınıflandırılması aşağıdaki iki modelden birini içine alabilir:

1-Denetlenmiş Sınıflandırma: Giriş numunesinde önceden tanımlanmış bir sınıfın üyesi olarak tanımlanır.

2-Denetlenmemiş Sınıflandırma: Numune bilinmeyen sınıfa atanır.

Buradaki tanıma problemine sınıflandırma veya kategorizasyon görevi verilir. Sınıflandırma gerek sistem tasarımcısı (denetlenmiş sınıflandırma) olarak tanımlanır gerekse numune benzerliği üzerine kurularak öğrenilir(denetlenmemiş sınıflandırma).

Numune tanıma modellerinin kısa bir özeti çizelge 4.2.'de verilmiştir. Bu tabloda sinir ağları (neural network) grubu başka bir yaklaşım olan bulanık sinir ağları (Fuzzy neural network) yaklaşımını da içerir.

**Çizelge 4.2.** Numune tanıma modelleri

Yaklaşım	Temsil	Tanıma Fonksiyonu	Tipik Kriter
Düzenleme (Template matching)	Örnekler, Pkseller, Eğriler	Korelasyon, Mesafe Ölçümü	Sınıflandırma Hatası
İstatistiksel (Statistical)	Özellikler	Ayrım Fonksiyonu	Sınıflandırma Hatası
Sentaktik veya Yapısal (Syntactic or Structural)	Primitives (ilkeller)	Kurallar, gramer	Kabul Hatası
Sinir Ağları(Neural Network)	Örnekler,Pkseller, Özellikler	Ağ Fonksiyonu	Orta Kare Hatası

#### 4.1. Düzenleme (Template Matching)

Numune tanımda en eski yaklaşımlardan biridir. Düzenleme numune tanımda genel bir işlem olup, aynı tipteki iki varlık arasındaki benzerliği belirlemek amacıyla kullanılır. Düzenlemede bir şablon (iki boyutlu bir matris) veya numunenin tanınabileceği bir prototip kullanılabilir. Müsaade edilen tüm poz (çevrim ve dönme) ve ölçü değişimleri hesaba katılırken, tanınacak numune biriktirilen şablonla (matris) karşılaştırılır. Benzerlik ölçütü, genellikle korelasyon, kullanılan eğitime seti üzerine kurularak optimize edilebilir. Genellikle şablonun kendisi eğitime setinden öğrenilir. Düzenleme bazı uygulama bölgelerinde etkinken, bazı dezavantajlara da sahiptir. Eğer numune görüntü işlemi, bakış noktası değişikliği veya numuneler arasındaki büyük sınıf varyasyonları sebebiyle çarpıtılmışsa başarısız olacaktır.

#### 4.2. İstatistik Yaklaşım

İstatistik yaklaşımda her bir numune d özellik veya ölçüm terimi ile temsil edilir ve d boyutlu bir uzayda nokta olarak görülür. Amaç bu özellikleri seçmektir. Bu özellikler d boyutlu özellik uzayında sıkı ve ayrık bölgeleri kullanan farklı kategorilere ait numune vektörlerine müsaade eder. Temsil uzayının (özellik seti) etkinliği, farklı sınıflardan numunelerin nasıl ayrılabilceği ile belirlenir. Her bir sınıftan verilen eğitime numunelerinin setinden amaç; özellik uzayında karar sınırlarını kurmaktır. Bu uzay farklı sınıflara ait numuneleri ayırır. İstatistik kararda teorik yaklaşım ve karar sınırları, her bir sınıfa ait numune olasılık dağılımları ile belirlenir.

Sınıflandırma yaklaşımı üzerine kurulu parçalı analizde kullanılabilir. Önce karar sınırının parametrik formu (lineer veya kuadratik) belirtilir sonra belirtilen formun en iyi karar

sınırı eğitime numunelerinin sınıflandırması üzerine kurularak bulunur. Bazı sınırlar, mesela orta karelendirilmiş hata (means square error) kriteri kullanılarak inşa edilebilir.

#### 4.3. Sentaktik Yaklaşım

Kompleks numunelere ait pek çok numune tanıma uygulamasında problemi hiyerarşik görünüme adapte etmek daha uygundur. Burada bir numune daha basit alt numunelerin bir toplamı olarak görülebilir. Bunlar daha basit olan alt numuneler tarafından belirlenir. Tanınacak en basit alt numuneye ilkel denir ve verilen kompleks numune bu ilkeller arasındaki iç ilişkilerle temsil edilir. Sentaktik numune tanımada formal benzerlik (analogy), numune yapısı ile dil sentaksı arasında belirlenir. Numuneler bir dile ait cümleler olarak, ilkeller bir dilin alfabeti olarak görülür ve cümleler bir gramere göre üretilir. Bu yüzden kompleks numunelerin büyük bir koleksiyonu, ilkellerin ve gramatik kuralların küçük bir sayısı ile tasvir edilebilir. Her bir numunenin gramerine kullanılabilir eğitime örneklerinden varılabilmelidir.

Yapısal numune tanımlamada sınıflandırmaya ilaveten, verilen numunenin ilkellerden nasıl yapılabildiğine dair bir tarif sunar. Bunun kullanıldığı durumlar şöyledir. EKG dalga formları, dokunmuş görüntü ve konturların biçim analizleri gibi kural setleriyle yakalanabilen kesin yapıya sahip numunelerdir.

#### 4.4. Sinir Ağları (Neural Network)

Sinir ağları; pek çok iç içe bağlı basit işlemcilerin çok büyük sayısını içeren, çoğunlukla paralel hesaplama sistemleri olarak tanımlanır. Sinir ağı modelleri, ağırlıklı yönetilmiş grafik ağında bazı organizasyon kurallarını (öğrenme, genelleştirme, adapte edilebilirlik, hata toleransı, dağıtımli temsil ve hesaplama) kullanmaya teşebbüs ederler. Bu ağlardaki nodlar yapay sinirlerdir. Bu nodlar sinir girişi, sinir çıkışı ve ikisi arasındaki bağlantılardan oluşan ağırlıklı saklı katmandır (hidden layer). Bir yapay sinir ağı sistemi bu katmanlarla yönetilir. Sinir ağlarının temel karakteristiği; kompleks non-lineer giriş-çıkış ilişkilerini öğrenme yeteneğine sahiptir, sıralı eğitime işlemlerini kullanır ve kendilerini veriyeye adapte ederler.

Numune sınıflandırma için en çok kullanılan sinir ağı ailesi, besleme ön ağıdır ve çok katmanlı Perceptron ve Radyal esaslı fonksiyon ağlarını içerir. Bu ağlar katmanlar arasında doğrusal olmayan bağlantılarla gerçekleştirilir. Diğer popüler ağ, kendiliğinden organize olan harita (self organizing map) (SOM) veya Kohonen ağıdır ve çoğunlukla veri kümelendirmesinde ve özellik haritalamasında kullanılır. Öğrenme, güncellenen ağ mimarisini ve bağlantı

diye ilgilendirilir. Numune tanıma problemlerinin çözümünde sinir ağı modellerinin artan popülaritesi öncelikle bölge-kesin bilgisi üzerinde görünen düşük bağımlılık (kurulu modele ve kural temelli yaklaşımlara göreceli) ve pratik kullanımda etkin öğrenme algoritmalarının kullanılabilirliği yüzündendir.

Sinir ağları; saklı katmanları kullanarak özellik çıkarımı ve sınıflandırma için non-linear algoritmalarla yeni bir gruplandırma oluşturur. Buna ilaveten mevcut özellik çıkarımı ve sınıflandırma algoritmaları etkin bir hardware gerçekleştirme için sinir ağı mimarisi üzerinde haritalanabilir. Görünen farklı prensiplere rağmen meşhur sinir ağı modellerinin çoğu eşit veya klasik istatistik numune tanıma metodlarına benzerdir. Bu ilişki çizelge 4.3.'de verilmiştir. Bu benzerliklere rağmen sinir ağları; özellik çıkarımı için birleştirilmiş yaklaşımlar, iyi, ılımlı non-linear çözümler için sınıflandırma ve esnek prosedürler gibi bir takım avantajlar sunarlar.

**Çizelge 4.3. İstatistik ve sinir ağları metodları arasındaki ilişkiler**

İstatistiksel Numune Tanıma	Yapay Sinir Ağları
Lineer Discriminant Function	Perceptron
Principal Component Analysis	Auto-Associative Network, and various PCA networks
A Posteriori Probability Estimation	Multilayer Perceptron
Nonlinear Discriminant Analysis	Multilayer Perceptron
Parzen Window Density-based Classifier	Radial Basis Function Network
Edited K-NN Rule	Kohonen's LVQ

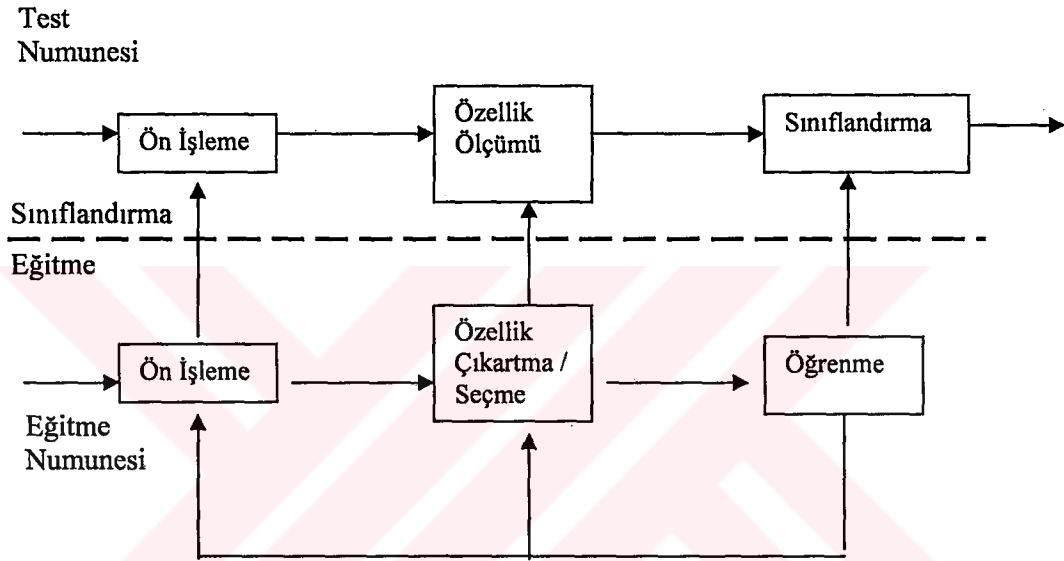
#### 4.5. Bulanık Sinir Ağları (Fuzzy Neural Network) (FNN)

Bu yaklaşım; içinde belirsizlik elemanı bulunduran enformasyon ve karar verme alanlarında ilgi çekici olmuştur.

Bu ağlar; bulanık giriş ve çıkışlardır. Veri, yapay sinir ağlarına sunulmadan önce bulanıklaştırılmalıdır. Mesela mümkünlük dağılımı kullanılarak üyelik değerinin, 0-1 arasındaki veri değerine atanması. Giriş ve çıkış transformasyonlarıyla dilbilim terimlerine bulanıklaştırılır. Bu yüzden özel bir nümerik değere sahip olmaktan çok, değerleri linguistik olarak çok düşük, düşük, orta, yüksek ve çok yüksek gibi tanımlayabiliriz. Bu çeşit bulanıklaştırma bazı uygulamalar için cezbedici olsa bile (koku kalitelerini sınıflama) sınırların güç belirlendiği durumlar için uygun değildir.

## 5. İSTATİSTİK NUMUNE TANIMA

İstatistik numune tanımada bir numune,  $d$  boyutlu özellik vektörü olarak görülen  $d$  özelliklerinin veya niteliklerinin bir setiyle temsil edilir. İstatistik karar teorisinin kavramları, numune sınıfları arasındaki karar sınırlarını kurmak için kullanılır. Tanıma sistemi şekil 5.1.'de görüldüğü gibi eğitime ve sınıflandırma modlarında çalıştırılır.



Şekil 5.1. İstatistik numune tanıma modeli

Ön işleme modülünün görevi; temelden ilgi numunesini parçalamak, gürültüyü kaldırmak, numuneyi normalize etmek ve numunenin kesin temsilini tanımlamada katkı sağlayacak diğer işlemleri normalize etmesidir. Eğitim modunda özellik çıkarımı ve seçimi modülü giriş numunelerini temsil etmek için uygun özellikler bulur ve sınıflayıcı özellik uzayının bölümlenmesi için geliştirilir. Geri besleme yolu bir tasarımcının ön işleme ve özellik çıkartma ve seçme stratejilerini optimize etmesine müsaade eder. Sınıflandırma modunda eğitilmiş sınıflayıcı, giriş numunesini ölçülmüş özellikler üzerine kurma düşüncesi altında numune sınıflarının birine atar.

Veri ön işlemede sensör dizisinin değiştirilen işlenmemiş sinyali veri analizi için bir giriş sağlayacaktır. Fakat verilen özel bir sensör dizisi için, uygun veri ön işleme tekniğini belirlemede özel bir rehber yoktur. Genellikle, uygun ön işleme tekniği bilinmez. Bazı durumlarda özel bir sensör dizisi veri analiz metodunun en uygun olanını belirlemede, bazı ön



işleme stratejilerinin keşfi faydalı olacaktır. Kullanılan veri ön işleme stratejilerinin bazılarının kısa bir özeti çizelge 5.1.'de verilmiştir.

**Çizelge 5.1.** Bazı veri ön işleme ve normalizasyon tekniklerinin tanımları

Rölatif Ölçeklendirme	$X_{ij} = X_{ij} / \max(x_i)$
	$X_{ij} = X_{ij} / \max(x_r)$ burada $x_r$ , $1 \times p$ referans cevabıdır.
	$X_{ij} = X_{ij} / \ x_i\ $ burada $\ x_i\ $ , $x_i$ 'nin Öklitik normu anlamındadır.
Temel Çıkartma	$x_i = x_i - x_b$ burada $x_b$ , $1 \times p$ boş cevaptır.
Sinyal Ortalaması	$X_{ij} = 1/N \sum_{k=1}^N X_{ij}^k$ burada $k$ , mevcut tekrarların sayısıdır.
Lineerleştirme	$X_{ij} = \log(X_{ij})$ , $X_{ij} = \sqrt{X_{ij}}$
Orta Merkezleme	$X_{ij} = X_{ij} - x_j$
Kendiliğinden Ölçeklendirme	$X_{ij} = X_{ij} - x_i / \sigma_x$
Sınır Ölçeklendirme	$X_{ij} = X_{ij} - \min(x_j) / \max(x_j) - \min(x_j)$
Kılavuz Çıkarma	$X_{ij} = X_{ij} - X_{i1}$
X bir (n x p) kalibrasyon matrisidir (n: örnek-p: sensör). $X_{ij}$ , j. sensörde i. örnek için cevaptır. $x_i$ , i. örneği için $1 \times p$ cevap vektörü ve $x_j$ , j. sensörde tüm i örneklerinin cevabını içerir.	

Veri ön işleme metotları statik ve dinamik sensör cevaplarının ikisine birden uygulanabilir. Çizelge 5.1.'deki ilk giriş, kullanılan rölatif ölçeklendirme prosedürlerinin bazılarını gösterir. Ölçeklendirme, referans cevabına veya numune cevaplarının görünüşlerine rölatif olarak yapılabilir. Bazı rölatif ölçeklendirmeler her bir sensör için, cevap yoğunluğunun konsantrasyon bağımlılığını denemek veya elimine etmek için kullanılabilirler. Bu yüzden bu yaklaşım nitelik uygulamaları için istenen bir yaklaşımdır.

Çıkarma tekniği, basitçe temel doğrulama metodudur. Matris etkilerini azaltmak için boş numune cevabı kayıt edilebilir ve her bir örnek cevaptan çıkarılabilir. Diğer bir ön işleme metodu, sinyal ortalamasıdır. Bu teknik her bir sensörle tekrarlı ölçümü gerektirir. Bu; bir dizideki her bir sensör tipinin çoklu sensörlerini kullanarak yada her bir numunenin tekrarlı ölçümlerini alarak başarılabılır. Numune cevabının sinyal / gürültü oranı  $N^{1/2}$  ile geliştirilebilir. Tanımlanan N tekrarlı ölçümlerin sayısını göstermektedir. Geliştirilmiş sinyaller bu yolla elde edilebilir.

Lineerleştirme teknikleri nonlinear bir cevap almayı arar ve bunu lineer bir temsile dönüştürür. Lineer veri analizi metotları kullanıldığında lineerleştirme istenir. Bununla birlikte, sensör cevabının non lineerlik tabiatını tanımlamak genellikle güçtür.

Orta merkezleme metodu veriyi basitçe transforme eder ve değişkenlerin merkezi orijinde birleşir. Benzer bir ön işleme metodu olan kendiliğinden ölçeklendirme verinin orta merkezlendirmesiyle ilgilidir ve özel bir sensörde tüm numune cevaplarının standart sapmasına bölünür. Kendiliğinden ölçeklendirme genellikle, ölçülen cevaplar farklı ünite ölçekleri üzerindeyken kullanılır. Kendiliğinden ölçeklendirilmiş veri her bir sensör için bir sıfır ortaya ve ünite varyansına sahip olacaktır. Sınır ölçeklendirme 0.0 ve 1.0 arasındaki tüm cevap değerlerini transforme eder. Bu şu anlama gelir; transforme edilmiş bölgede her bir sensördeki minimum cevap orijinde, maksimum cevap 1.0'dadır.

Bazı ön işleme metotları dinamik veriyi elde tutmak için dizayn edilirler. Mesela kılavuz çıkarma metodu hiçbir örnek yokken kaydedilen sinyali elimine etmede kullanılır. Bu, bir cevap için sensörde kaydedilen tüm zaman noktalarından, sensördeki ilk zaman noktası çıkartılarak elde edilir. Uygulamalar, zaman üzerinde sensörlerin veri ölçümü ile ilgilendirir. Bu, sensör başına büyük miktarda ölçüm sayısı gerektirir. Tipik olarak veri noktalarının sayısı, numune tanıma metotları için makul ebatta veri matrisi yapmak için, bazı yollarla azaltılmalıdır. En basit durumda sabit durum cevabı basitçe hesaplanır ve sensör başına bir değer üretilir. Dinamik veri için kullanılan diğer ön işleme teknikleri, dalga transformlarını, fourier transformlarını, Gram-Schmidt ortogonalizasyonunu ve oto-regresiv modeli ihtiva ederler.

İstatistik numune tanımadaki karar verme işlemini şu şekilde açıklayabiliriz: Verilen A numunesi,  $x = (x_1, x_2, x_3, \dots, x_d)$  değerlerinin d özellik vektörü üzerine kurulu c tane,  $w_1, w_2, \dots, w_c$  kategorilerinden birine atanacaktır. Özelliklerin, numune sınıfı üzerine durumlandırılmış olasılık yoğunluğu veya kütle fonksiyonuna sahip olduğu düşünülür. Bu yüzden  $w_i$  sınıfına ait x numune vektörü;  $p(x/w_i)$  sınıf durumlu olasılık fonksiyonundan gelişigüzel elde edilen bir gözlem olarak görülebilir.

İstatistik numune tanımadaki diğer ayırım denetlenmiş öğrenmeye karşı denetlenmemiş öğrenmedir. Bir eğitime numunesi etiketi numunenin ait olduğu kategoriye temsil eder. Denetlenmemiş öğrenme probleminde bazen sınıf sayısı, her bir sınıfın yapısı boyunca öğrenilmek zorundadır. İstatistik numune tanımadaki diğer ayırımlar şekil 5.2.'de gösterilmiştir.

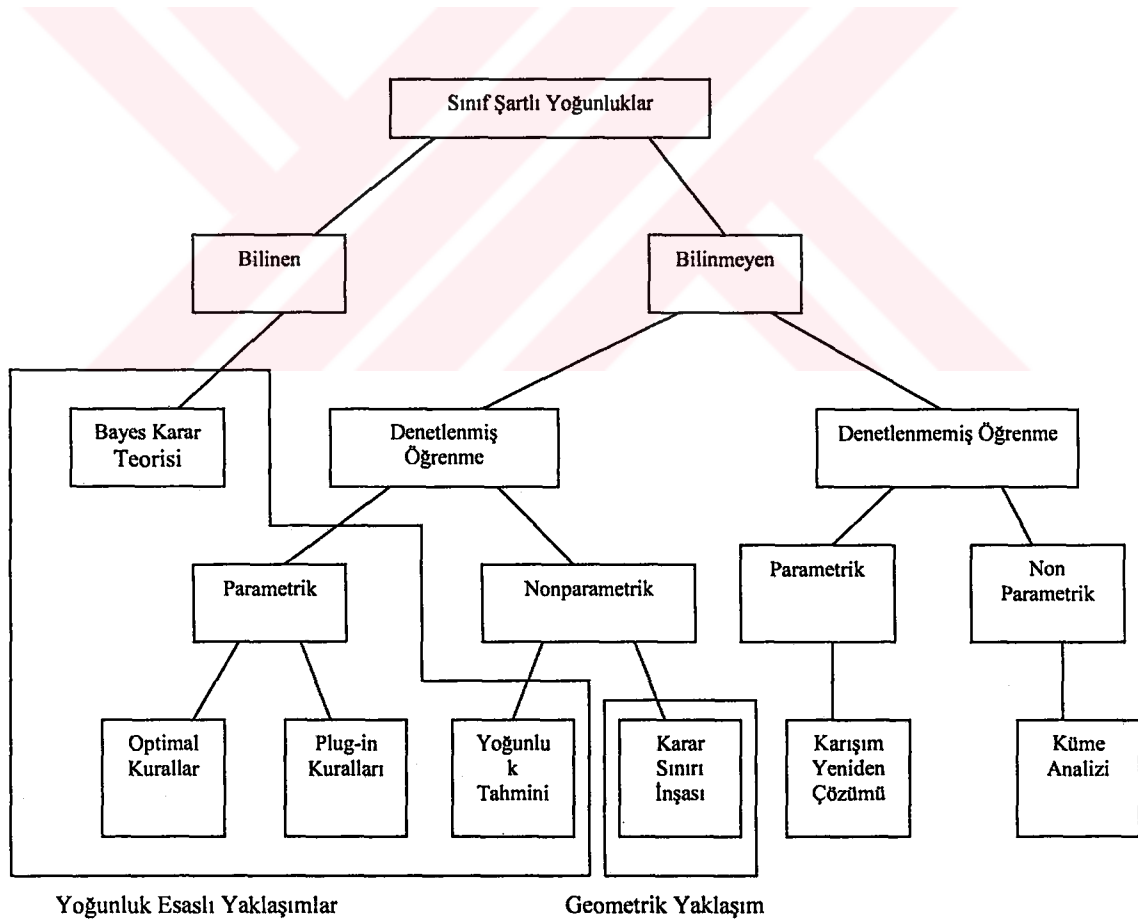
Sistem hangi sınıflandırma veya karar kullanılırsa kullanılsın, mümkün eğitime örnekleri kullanılarak eğitilmek zorundadır. Sonuç olarak sınıflandırma performansı, örnek değerlerini

belirleyiciler kadar kullanılabilir eğitime örneklerinin sayısına da bağlıdır. Aynı zamanda tanıma sistemini dizayn etmenin amacı eğitime örneklerinden farklı olan, gelecek test örneklerini sınıflandırmaktır. Bu yüzden, eğitim seti üzerinde bir sınıflamanın performansı optimizasyonu; test seti üzerinde arzu edilen performansı her zaman vermeyebilir. Bir sınıflayıcının genelleştirme yeteneği, eğitime aşaması boyunca kullanılmayan test numunelerinin sınıflamasında onun performansına başvurabilir. Bir sınıflayıcının zayıf genelleştirme yeteneği, aşağıdaki faktörlerden herhangi birine bağlanabilir.

1-Özellik sayısı eğitime örneklerinin sayısına bağlı olarak çok büyüktür.

2-Sınıflandırıcı ile ilgili bilinmeyen parametrelerin sayısı büyüktür.

3-Sınıflayıcı, eğitim seti üzerinde yoğun olarak optimize edilmiştir. Çok fazla serbest parametre varsa, regresyondaki aşırı uyum olgusuna (phenomenon of overfitting in regression) benzerdir.



Şekil 5.2. İstatistik numune tanımada değişik yaklaşımlar

## 6. BOYUTLANABİLİRLİK

Bir sınıflayıcının performansı; örnek ebatları, özellik sayısı ve sınıflayıcı kompleksliği arasındaki iç ilişkilere bağlıdır. Özellik uzayını hücelere bölen ve her bir hücrenin sınıf etiketiyle ilgilenen bir tablo bakış tekniği özellik boyutunun üstel fonksiyonu olan eğitme veri noktalarının sayısını gerektirir. Bu olgu boyutlanabilirlik olarak adlandırılır ve sınıflayıcı dizaynında artan olguya neden olur. Bir karar kuralının yanlış sınıflandırma olasılığı, özellik sayısının arttığı kadar artıramaz. Bununla birlikte pratikte genellikle şu gözlenir: Eğer sınıflandırıcı tasarımı için kullanılan eğitme örneklerinin sayısı, özellik sayısına göre küçükse eklemeli özellikler sınıflandırıcının performansının düşürebilir.

Trunk' ın [6], boyutlanabilirlik felaketini göstermek için verdiği basit bir örnek. Eşit öncelikli ihtimalli iki-sınıf sınıflandırma problemi düşünelim. Her bir sınıf için özdeşlikli kovaryans matrisinin  $d$  boyutlu çok değişkenli çok dönüşümlü Gauss dağılımı mevcut olsun. İki sınıf için orta vektörler aşağıdaki komponentlere sahiptir.

$$m_1 = (1, 1/\sqrt{2}, 1/\sqrt{3}, \dots, 1/\sqrt{d})$$

$$m_2 = (-1, -1/\sqrt{2}, -1/\sqrt{3}, \dots, -1/\sqrt{d})$$

Özellikler istatistik olarak bağımsızdır ve başarılı özelliklerin ayırıcı gücü, iki sınıf arasındaki maksimum ayrımı sağlayan ilk özellik ile monotonik olarak azalır. Yoğunluklardaki tek parametre orta vektördür,  $m = m_1 = -m_2$ . Trunk' ın olası iki durumu:

1-  $m$  orta vektörü bilinir. Bu durumda karar sınırı inşa etmek için optimal Bayes karar kuralını (0/1 kayıp fonksiyonla) kullanabiliriz. Hata olasılığı  $d$  fonksiyonu olarak şu şekilde açıklanabilir:

$$P_e(d) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\sum_{i=1}^d (1/i)}} \frac{1}{\sqrt{2\pi}} e^{-1/2z^2} dz \quad (2)$$

$\lim_{d \rightarrow \infty} P_e(d) = 0$  olduğunu doğrulamak kolaydır. Diğer bir deyişle,  $d$  özellik sayısının keyfi olarak artmasıyla iki sınıfı tam olarak ayırabiliriz.

2-  $m$  orta vektörü bilinmiyor ve  $n$  etiketli eğitme örnekleri kullanılabilir. Trunk  $m$ 'nin maksimum olasılık tahmini  $\hat{m}$ 'yi buldu ve karar kuralını harekete geçirmede kullandı. Şimdi  $n$  ve  $d$ 'nin fonksiyonu olan hata olasılığı şöyle yazılabilir.

$$P_e(n, d) = \int_{\mathcal{B}(d)} \frac{1}{\sqrt{2\pi}} e^{-1/2z^2} dz \quad (3)$$

$$\theta(d) = \sum_{i=1}^d (l/i) / \sqrt{(1+1/n) \sum_{i=1}^d (l/i) + d/n} \quad (4)$$

Trunk;  $\lim_{d \rightarrow \infty} P_e(n, d) = 1/2$  olduğunu gösterdi. Bunun anlamı: Bu iki sınıf problem için hata olasılığı, 0.5 maksimum mümkün değerine yaklaşır. Bu şunu gösterir: Sınıf şartlı yoğunlukların parametreleri eğitime örneklerinin sonlu sayısından tahmin edilirken, biz özellik sayısını keyfi olarak artıramayız. Boyutlanabilirliğin pratik iması şudur: Bir sistem sınırlı eğitime setiyle sınırlandırılırken, tasarımcı belli başlı özelliklerin az sayısını seçmeyi denemelidir.

Müşterek kullanılan sınıflayıcıların tümü boyutlanabilirlikten sıkıntı çekerler. Yanlış sınıflandırma olasılığı, eğitime örneklerinin sayısı, özellik sayısı ve sınıf şartlı yoğunlukların doğru parametreleri arasındaki tam ilişki kurulma açısından çok zorken, bazı yol göstericiler boyutlanabilirliğe bağlı olan örnek ebat oranı teklif ettiler. Genellikle şu kabul edilir ki; sınıf başına pek çok eğitime örneği gibi en az on defa özellik sayısı kullanılması  $(n/d) \geq 10$  sınıflayıcı tasarımını takip açısından iyi bir pratiktir.

### 6.1. Boyutlanabilirlik İndirgenmesi

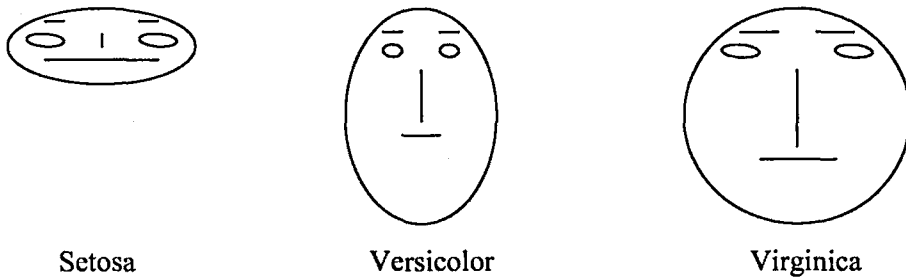
Numune temsiline boyutlanabilirliğini mümkün olduğu kadar küçük tutmak için iki ana sebep vardır. Ölçüm maliyeti ve sınıflama doğruluğu. Sınırlı çarpıcı özellik seti, seçilen temsilcilik üzerine yapılan numune temsilciliğini ve sınıflandırıcıyı basitleştirir. Neticede sonuç sınıflandırıcısı daha hızlı olacak ve daha az hafıza kullanacaktır. Özelliklerin az bir sayısı, eğitime örneklerinin sayısı sınırlı iken, boyutlanabilirliği hafifletecektir. Diğer yandan özelliklerin sayısındaki indirgeme ayırım gücünde kayıplara sebep olacak ve sonuçtaki tanıma sisteminin doğruluğu daha düşük olacaktır.

Özellik seçimi; giriş özellik setinin ümit edilen en iyi alt setlerini seçen algoritmalarla müracaat eder. Orijinal özellik setinin transformasyonları veya kombinasyonları üzerine kurulmuş yeni özellikleri meydana getiren metotlar özellik çıkarma algoritmaları olarak tanımlanır. Özellik çıkarımının özellik seçiminden önde olduğuna dikkat edilmelidir. İlk önce özellikler algılanan veriden çıkartılır daha sonra düşük ayırım yeteneği olan çıkarılan parçaların bazıları iskartaya çıkartılır. Özellik seçimi ve özellik çıkarımı arasındaki seçim kullanılan belirleyici eğitime verisine ve uygulama bölgesine bağlıdır. Özellik seçimi ölçüm maliyetindeki birikimleri idare eder ve seçilmiş özellikler orijinal fiziksel yorumlarını korurlar. Buna ilaveten korunan özellikler, numuneleri üreten fiziksel aşamaları anlamada önemli

olabilirler. Diğer yandan özellik çıkarımıyla üretilmiş dönüştürülmüş özellikler verilen en iyi alt setten daha iyi bir ayırım yeteneği sağlayabilir. Verilen özelliklerin lineer veya nonlineer kombinasyonu olan yeni özelliklerin açık bir fiziksel anlamı olmayabilir.

Pek çok durumda verinin hayali incelemesine izin vermek için, çok değişkenli verinin iki veya üç boyutlu yansımalarını ( $n \times d$  numune matrisi) elde etmek yararlıdır. Çok değişkenli veriyi hayali olarak gözlemek için de bazı grafik teknikleri mevcuttur. Burada amaç her bir numuneyi d serbestlik dereceli bir resim olarak resmetmektir. Özelliklerin verilen sayısını d temsil eder. Chernoff; burun uzunluğu, ağız eğriliği ve göz ebadı gibi yüze ait özellikleri kullanarak ve her bir numuneyi bir karikatür yüzü olarak temsil eder [7]. Bunu bireysel özelliklere bağlamak amacıyla yapar. Şekil 6.1., İris verisinde İris Setosa, İris Versicolor ve İris Virginica sınıflarının orta vektörlerine bağlı üç yüzü gösterir. Setosa kategorisi dört boyutlu özellik uzayında kalan iki kategoriden iyi bir şekilde ayrılabilir. İris veri seti her bir sınıfta 50 numune olmak üzere toplam 150 tane dört boyutlu numune içerir.

Boyutluluk indirgenmesindeki ana mevzu kriter fonksiyonunun seçimidir. Yaygın olarak kullanılan kriter, bir özellik alt setinin sınıflandırma hatasıdır. Fakat örnek ebadının özellik sayısına oranı küçükken, sınıflandırma hatası kendiliğinden tahmin edilemeye bilir. Kriter fonksiyonunun seçimine ilaveten bizim, indirgenmiş özellik uzayının uygun boyutluluğunu belirlemeye de ihtiyacımız vardır. Bu sorunun cevabı verinin doğal boyutluluk notasyonunda mevcuttur. Verilen d boyutlu numunelerin, d'den daha küçük boyutluluk uzayında yeterli bir şekilde tarif edilebileceğini önemli ölçüde doğal boyutluluk belirler. Mesela, oldukça düzgün bir eğri boyunca d boyutlu numuneler d'nin önemsiz bir değeri olan doğal boyutluluğa sahiptir. Şuna da dikkat edilmelidir ki doğal boyutluluk, kovaryans matrisinin önemli özdeğerlerine ait verinin global özelliği olan lineer boyutlulukla aynı değildir. Doğal boyutluluğu tahmin etmek için bazı algoritmalar kullanılırken, bunlar tariflenmiş boyutluluğun alt uzayının nasıl kolaylıkla tanımlanabileceğini göstermezler.



Şekil 6.1. Chernoff'un orta vektörlerine bağlı yüzleri

### 6.1.1. Özellik çıkarımı

Özellik çıkarma metotları;  $d$  boyutlu orijinal özellik uzayında ( $m \leq d$ )  $m$  boyutluluğunun uygun alt uzayını lineer veya nonlineer yolla belirler. Baş bileşen analizi (principal component analysis) (PCA), faktör analizleri, lineer ayırım analizleri (linear discriminant analysis) (LDA) ve yansıtma takibi (projection pursuit) gibi lineer transformasyonlar özellik çıkarımı ve boyutluluk indirgenmesi için numune tanımada geniş ölçüde kullanılırlar. En iyi bilinen lineer özellik çıkarıcı ana komponent analizi veya Karhunen-Loeve genişlemesidir. Bunlar  $n$  tane  $d$  boyutlu numunenin  $d \times d$  kovaryans matrisinin en büyük  $m$  öz vektörlerini hesaplarlar. Lineer transformasyon şöyle tarif edilir:

$$Y = X \times H \quad (5)$$

Burada  $X$ ; verilen  $n \times d$  numune matrisi,  $Y$ ; türetilen  $n \times m$  numune matrisi,  $H$ ; kolonları öz vektörler olan  $d \times m$  lineer transformasyon matrisidir. Ana komponent analizi en anlamlı özellikleri kullandığından (en büyük öz değerli öz vektörler) bu, orta karelenmiş hata kriterini kullanan lineer alt uzayla etkin bir şekilde veriye yaklaşır. Yansıtma takibi [8] ve bağımsız komponent analizi (independent component analysis) (ICA) gibi diğer metotlar [9],[10],[11],[12] Gauss dağılımlı olmayanlar için daha uygundur. Çünkü onlar verinin ikinci mertbe özelliğine güvenmezler. ICA belirsiz kaynak ayırımı için başarıyla kullanılmıştır. Bu, bağımsız kaynakları tanımlayan lineer özellik kombinasyonlarını çıkarır. Gauss dağılımına sahip kaynaklar en az bir tane ise yukarıdaki olay mümkün hale gelir.

PCA, denetlenmemiş lineer özellik çıkarım metodu iken, ayırım analizleri en çok ayırmacı özelliklerin lineer olarak çıkarımı için her bir numuneyle ilgili kategori bilgisi kullanır. Ayırım analizinde, Fisher kriteri gibi genel ayrılabilirlik ölçüleriyle PCA'daki toplam kovaryans matrisi değiştirilerek iç sınıf ayırımı vurgulanır. Bu kriterlerle  $S_w^{-1}S_b$  öz vektörleri bulunur ( $S_w$ , sınıf dağılım matrisinde ters ürün ve  $S_b$  sınıf dağılım matrisinin arasında). Gauss olmayan sınıf şartlı yoğunluklar için diğer denetlenmiş kriter, Parzen yoğunluk tahminlerini kullanan Patrick-Fisher mesafesi üzerine kurulmuştur [13].

Nonlineer özellik çıkarma tekniklerini tanımlamak için bazı yöntemler vardır. Direk olarak PCA ile ilgili bir metot; Kernel PCA adını alır [14],[15]. Kernel PCA'nın temel fikri,  $\phi$  nonlineer fonksiyonu yoluyla giriş verisinden  $F$  yeni özellik uzaylarına tipik olarak haritalamak ve daha sonra haritalanmış uzayda lineer bir PCA gerçekleştirmektir. Bununla beraber  $F$  uzayının çok yüksek boyutu vardır.  $\phi$  haritalamasının hesaplanmasından açıkça kaçınmak için Kernel PCA iç çarpışma ayrılabilen Mercer Kernel'leri çalıştırır.

$$K(x, y) = \phi(x) \cdot \phi(y) \quad (6)$$

Netice olarak Kernel uzayının iyi tanımlı bir metriği vardır. Mercer Kernel örnekleri p. mertebe polinomial  $(x - y)^p$  ve Gauss Kernel'ini  $e^{-\|x-y\|^2/c}$  içerirler.

$X$ ; normalleştirilmiş sıfır ortalı  $n \times d$  numune matrisi ve  $\phi$ ;  $F$  uzayında numune matrisi olsun.  $F$  uzayında lineer PCA,  $\phi(X)\phi(X)^T$  korelasyon matrisinin öz vektörlerini çözer. Bu  $K(X, X)$  Kernel matrisi olarak adlandırılır. Kernel PCA'da,  $K(X, X)$ 'in ilk  $m$  öz vektörü  $E$  transformasyon matrisini tanımlamak için, elde edilir ( $m$ ; özelliklerin arzulan sayısını temsil ederken,  $m \leq d$ ,  $E$ 'nin ebadı  $n \times m$ 'dir). Yeni  $x$  numuneleri  $K(x, X)E$  ile haritalanır. Bunlar eğitime setine bağlı olan ve ölçülmüş özellik değerlerine bağlı olmayan bir şekilde temsil edilirler. Tüm temsil için  $E$ 'de,  $m$  öz vektörlerine kadar Kernel PCA ile Kernel fonksiyonuna bağlı olarak ihtiyaç duyulmayabilir. Lineer PCA'da  $d$  öz vektör seti orijinal özellik uzayını temsil eder.

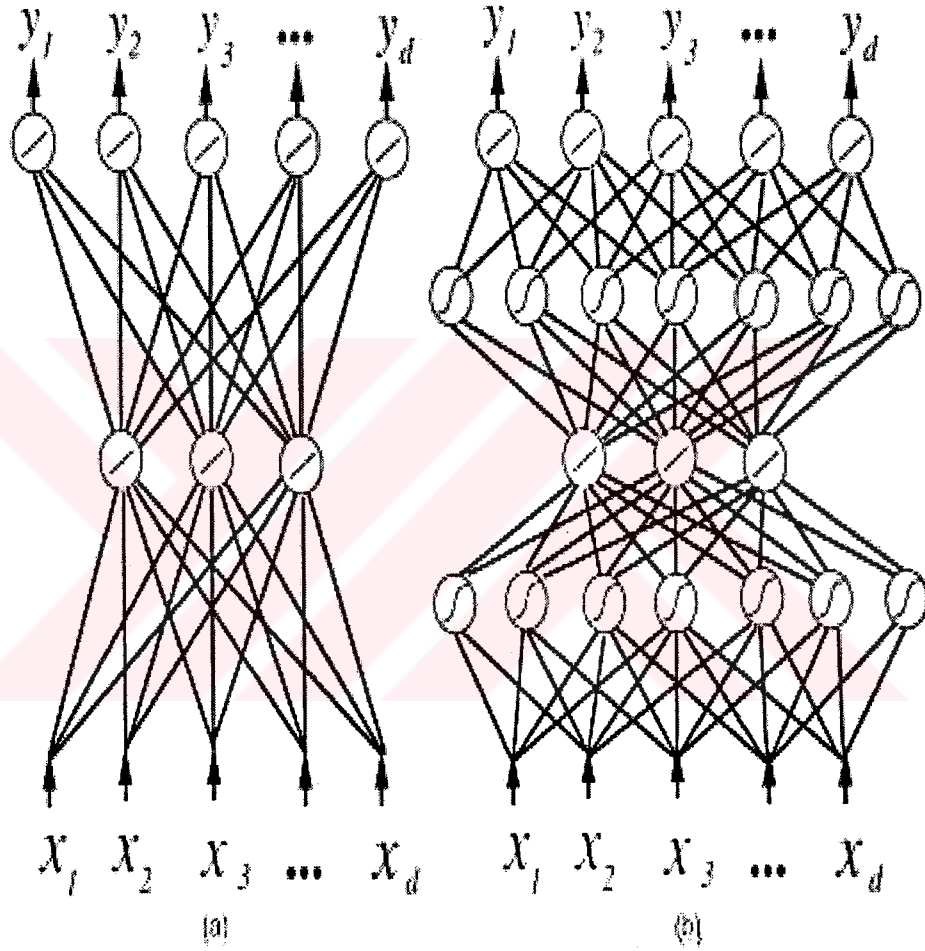
Çok boyutlu ölçekleme (multidimensional scaling) (MDS) diğer bir nonlineer özellik çıkarma tekniğidir. İki veya üç boyutta, çok boyutlu bir veri setini temsil etmeyi amaçlar. Şöyle ki; orijinal  $d$  boyutlu özellik uzayında mesafe matrisi yansıtılan uzaydaki gibidir. Bu haritalamanın performansının ölçümü için çeşitli gerilme fonksiyonları kullanılır. En popüler kriter Sammon [16] ve Niemann [17] tarafından tanımlanan gerilme fonksiyonudur. MDS; tam bir haritalama fonksiyonu vermez. Bu yüzden haritalamayı tekrar etmeksizin verilen bir eğitime seti için hesaplanan haritalamada yeni bir numuneyi yerleştirmek mümkün değildir.

Özellik çıkarımı ve sınıflandırma için besleme ön sinir ağı birleştirilmiş bir prosedür sunar. Her bir saklı tabakanın çıkışı yeni, nonlineer bir set olarak yorumlanabilir. Özellikler sınıflandırma için bir çıkış tabakası sunmuştur. Bu durumda çok katmanlı ağlar özellik çıkarıcıları olarak hizmet ederler [18].

Sinir ağları denetlenmemiş bir modda özellik çıkartmak için direkt olarak da kullanılabilir. Şekil 6.2.a, PCA alt uzayını [19] bulabilecek bir ağ mimarisini gösterir. Sigmoidlerin yerine sinirler lineer transfer fonksiyonlarına sahiptir. Verilen özellik sayısı  $d$  iken bu ağın  $d$  girişi ve  $d$  çıkışı vardır. Girişler hedef olarak da kullanılırlar. Sadece bir saklı tabaka kullanılarak giriş uzayını yeniden inşa etmek için çıkış tabakasını zorlarlar. Saklı tabakadaki üç nod, ilk üç ana komponenti yakalar [20]. Sigmoidal saklı ünitelerin iki nonlineer tabakası da dahil edilirse (Şekil 6.2.b'ye bak), orta tabakada nonlineer bir uzay bulunur, buna Bottleneck katmanı adı verilir. Nonlineerlik bu ilave katmanların ebatıyla sınırlıdır. Bu şekilde



kendiliğinden çağrışımı olanlar veya nonlinear PCA ağları, eğitime için güçlü bir araç sunarlar ve nonlinear alt uzayları tanımlarlar [21]. Kendi-Organizeli harita (SOM) veya Kohonen haritası [22] da nonlinear özellik çıkarımı için kullanılabilir.



Şekil 6.2. Üç boyutlu alt uzayın bulunması için otomatik çağrışımı ağlar  
(a) lineer (b) nonlinear

Çizelge 6.1.'de özellik çıkarma ve yansıma metotlarının bir özeti verilmiştir. Kategori bilgisinin kullanıldığı haritalamaların daha iyi ayırım yaptığı unutulmamalıdır.

Çizelge 6.1. Özellik çıkarma ve yansıtma metotları

Metot	Özellik	Yorumlar
Baş Bileşen Analizi	Lineer Haritalama, hızlı, öz vektör tabanlı	Geleneksel öz vektör temelli metot, Karhunen-Loeve genişlemesi olarak da bilinir; Gaussian veri için iyidir.
Lineer Ayrım Analizi	Denetlenmiş lineer haritalama, hızlı, öz vektör tabanlı	Sınıflama için PCA'dan daha iyidir, sıfır olmayan öz değerler ile (c-1) komponentleri ile sınırlıdır.
Yansıtma Takibi	Lineer haritalama, tekrarlı, Gaussian değil	Genellikle interaktif veri analizi keşfi için kullanılır.
Bağımsız Komponent Analizi	Lineer haritalama, tekrarlı, Gaussian değil	Kör kaynak ayrımı, Gaussian olmayan dağılımlı kaynakların ayırmak için kullanılır.
Kernel PCA	Nonlineer haritalama, öz vektör tabanlı	PCA temelli metot, numune vektörlerinin iç çarpımlarını değiştirmede kernel kullanılır.
PCA Ağı	Lineer haritalama, tekrarlı	Sadece bir saklı tabaka ve lineer transfer fonksiyonu ile otomatik çağrışımli sinir ağı
Nonlineer PCA	Lineer haritalama, Gaussian olmayan kriter, genellikle tekrarlı	Sinir ağı yaklaşımı, muhtemelen ICA için kullanılır
Nonlineer Otomatik Çağrışımli Ağ (Nonlinear auto-associative network)	Nonlineer haritalama, Gaussian olmayan kriter, tekrarlı	Çeşitli saklı tabakalarla Bottleneck ağı, bir nonlineer yeniden inşa tarafından nonlineer haritalama optimize edilir, hedef olarak giriş kullanılır
Çok Boyutlu Ölçekleme (MDS) ve Sammon Yansıtması	Nonlineer haritalama, tekrarlı	Sıklıkla zayıf genelleştirme, örnek ebatı sınırlı, gürültü hassasiyeti, çoğunlukla iki boyutlu görüntüleme için kullanılır.
Kendi-Organizeli Haritalama (SOM)	Nonlineer, tekrarlı	Özellik uzayında sinirlerin gridi üzerine kurulu, düşük boyutlu uzayların çıkartılması için uygundur

### 6.1.2. Özellik seçimi

Aşağıdaki durumlarda karşılaşılan özelliklerin çok fazla olması sebebiyle, özellik seçim metotlarını uygulamada bir ilgi canlanması olmuştur.

- 1- Çok sensör birleşmesi: Farklı sensör modellerinden hesaplanan özellikler, büyük sayıdaki komponentlerle birlikte bir özellik vektörü oluşturmak için birleştirilir.
- 2- Çoklu veri modellerinin entegrasyonu: Farklı yaklaşımlar kullanılarak sensör verisi modellenilebilir. Model parametreleri özellik olarak hizmet ederler ve farklı modellerin parametreleri yüksek boyutlu özellik vektörünün ürün vermesi için birleştirilebilir.

Verilen özellik seti  $Y$  olsun. Burada  $d$ ; kardinaliteyi ve  $m$ ; seçilen  $X, X \subseteq Y$  alt uzayında arzulanan özellik sayısını temsil eder.  $X$  seti için, özellik seçimi kriter fonksiyonu

$J(X)$  ile temsil edilir. Şunu farz edelim:  $J$ 'nin daha yüksek değeri, daha iyi özellik alt seti gösterir. Kriter fonksiyonu için tabii seçim  $J = (1 - P_e)$ 'dir.  $P_e$ ; sınıflandırma hatasını gösterir. Kriter fonksiyonunda  $P_e$ 'nin kullanımı; özellik seçimi prosedürlerini kullanılan belirleyici sınıflayıcılara, eğitime ve test setlerinin ebatlarına bağımlı yapar. Özellik seçimi probleminde en doğru yaklaşım şunları gerektirecektir:

- 1-  $m$  ebatlı tüm  $\binom{d}{m}$  mümkün alt setlerinin incelenmesi
- 2-  $J(.)$ 'nin en büyük değerli alt setinin seçilmesi

Bununla birlikte mümkün alt setlerin sayısı birleşerek büyür.

Literatürde [23] önerilen meşhur özellik seçme metodlarının çoğu çizelge 6.2.'de listelenmiştir. Sadece ilk iki metod, bir optimal alt seti garanti eder. Diğer tüm stratejiler, en iyi özellik çifti en iyi tek özelliği içermediği için alt optimaldir [24]. Genelde iyi ve daha büyük özellik alt setleri, iyi ve küçük setleri içermezler. Neticede en iyi bireysel özelliği seçmenin basit metodu başarısızlığa uğrar. Bu hala faydalı olabilir, şöyle ki; çok büyük özellik setlerinin azalmasında teker, teker bazı iyi özellikleri seçmede ilk adım olabilir. Daha fazla seçim, özellik bağımlılıklarını dikkate alan daha gelişmiş metodlarla yapılmak zorundadır. Bunlar, artan özellik setlerini değerlendirerek (ön seçim) veya daralan özellik setlerini değerlendirerek (geri seçim) çalışırlar. Sıralı ön seçim (sequential forward selection) SFS sıralı arka seçim (sequential backward selection) (SBS)'ye benzer basit bir sıralı metod bir zamanda bir özellik ekler (siler). Daha ileri teknikler al götür seçimi ((plus 1- take away r selection) ve sıralı yüzdürme araştırma metodları (sequential floating search), SFFS ve SBFS [25]'dir.

Araştırma stratejisine ilave olarak, kullanıcının  $J(.)$  uygun değerlendirme kriteri seçmeye ve  $m$  değerini belirlemeye ihtiyacı vardır. Özellik seçme metodlarının çoğu etkinliğini değerlendirmek için özellik alt setinin sınıflandırma hatasını kullanır. Mesela hata tahmininin birini dışarıda bırakma metodu kullanılarak k-NN sınıflandırıcısı ile bu yapılabilir. Bununla beraber hata oranını tahmin etmek için farklı bir metod ve sınıflayıcı kullanımı ile, seçilen farklı bir özellik alt seti idare edilebilir. Ferri ve arkadaşları ile Jain ve Zongker [23]; sınıflandırma hatası ve geçen zaman vasıtasıyla, özellik seçme algoritmalarının bazılarını karşılaştırdı. Genel sonuç şudur: Sıralı ön yüzdürme araştırması (sequential forward floating search) (SFFS) metodu, şube (kol) ve sınır algoritması kadar başarıyla çalışır ve daha düşük sayısal kaynakları talep eder.

Çizelge 6.2. Özellik seçme metotları

Metot	Özellikler	Yorumlar
Ayrıntılı Araştırma	Tüm muhtemel ( $\binom{d}{m}$ ) alt setlerini değerlendirir	Optimal alt setleri aramada garantili, m ve d'nin orta büyüklükteki tek değerleri için fizibil değil
Kol ve Sınır Araştırma	İyi bilinen kol ve sınır araştırma metodu kullanır, tüm muhtemel özellik alt setlerinin sadece bir kesri optimal alt seti bulmak için numaralandırılmalıdır.	Optimal alt setleri bulmada garantili yeter ki kriter fonksiyonu monotonluk özelliğini sağlasın, bu algoritmanın en kötü durumu eksponansiyelidir
En İyi Bireysel Özellik	Tüm m bireysel özelliklerini değerlendirir, en iyi m bireysel özelliği seçer	Basitçe hesaplama, optimal alt seti tamamen yönetmek için değil
Sıralı Ön Seçim (SFS)	En iyi tek özelliği seç ve sonra seçilen özelliğin kriter fonksiyonunu maksimize ettiği kombinasyondaki zamanda bir özellik ekle	Özellik bir kere korunduğunda ıskartaya çıkartılamaz, 2 boyutun bir alt setini seçtiği için cazip hesaplama, bu yalnızca (d-1) muhtemel alt setleri inceler
Sıralı Arka Seçim (SBS)	Tüm d özellikleri ile başlar ve bir zamanda bir özelliği başarıyla siler	Bir kez özellik silindiğinde optimal alt set geri alınamaz SFS'den daha çok hesaplama ihtiyacı duyar
Al Götür Seçimi	Önce ön seçimi kullanarak l özellikleri vasıtasıyla özellik alt setini genişlet sonra arka seçimi kullanarak r özelliklerini sil	SFS ve SBS metotlarında karşılaşılan özellik alt setinin "nesting" probleminden sakın, l ve r'nin değerlerinin seçimine ihtiyaç duyar (l>r)
Sıralı Öne Doğru Yüzdürme Araştırması (SFFS) ve Sıralı Geriye Doğru Yüzdürme Araştırması (SBFS)	Al götür metodunun geliştirilmişidir, l ve r'nin değerleri otomatik olarak belirlenir ve dinamik olarak güncelleştirilir	Gücün yettiği hesaplama maliyetinde optimal çözüme yakın çözümler sağlar

Çizelge 6.2.'deki özellik seçme metotları iyi bilinen sınıflayıcıların herhangi birisiyle kullanılabilir. Fakat çok tabakalı besleme ön ağı, numune sınıflandırma için kullanılırsa nod budama metodu simültane olarak optimal özellik alt setini ve optimal ağ sınıflayıcısını belirler [26,27]. İlk önce bir ağ eğitilir, sonra en az göze çarpan nod değiştirilir (girişte veya saklı tabakalarda). İndirgenmiş ağ yine eğitilir ve yine az çarpıcı nodun değiştirilmesiyle bu işleme devam edilir. Bu prosedür, sınıflandırma hatası ile ağ ebadı arasındaki birleşme başarılıncaya kadar tekrarlanır. Giriş nodunun budanması, buna bağlı özelliğin değişmesine eşittir.

Eğitme örneklerinin kullanılabilir sayısının, özellik sayısına oranı küçükken özellik seçme sonuçlarına nasıl güvenilebilir? Mahalonabis mesafesinin [28] özellik seçme kriteri olarak kullanıldığını düşünelim. Küçük örnek ebadı durumlarının tahminlerindeki belirsizlik, optimal özellik alt seti sonucunu verebilir. Bu alt setler kovaryans matrisi bilinirken elde edilebilecek optimal alt setten oldukça farklıdır. Jain ve Zongker [23] 20 boyutlu Gauss sınıf sathlı yoğunluklara ait iki sınıf sınıflandırma problemi için bu olayı gösterir. Umulduğu gibi

küçük eğitim setleri için seçilen özellik alt setinin kalitesi zayıftır, fakat eğitim seti ebadı artarken bu gelişir. Mesela eğitim setindeki 20 numuneyle birlikte, kol ve sınır algoritması 10 özellik alt setini seçmiştir. Bu alt setler, 10 özelliğin ideal alt setleriyle birlikte yoğunluklar bilindiğinde yaygın beş özelliği de içerirler. Eğitim setindeki 2500 numuneyle birlikte kol ve sınır prosedürü sadece bir yanlış özellikle birlikte 10 özellik alt seti seçer.

Dijit veri seti, el yazısı sayıları ('0' - '9') içerir. Sınıf başına iki yüz numune (toplamda 2000 numune), 30x48 ikili görüntü formunda kullanılabilir. Bu karakterler aşağıdaki altı özellik setiyle temsil edilirler:

1. 76 karakter biçiminin Fourier katsayısı
2. 216 profil korelasyonu
3. 64 Karhunen-Loeve katsayısı
4. 2x3 penceresinde 240 piksel ortalaması
5. 47 Zernike momenti
6. 6 morfolojik özellik

## 7. SINIFLANDIRICILAR

Pratikte sınıflandırıcı seçimi güç bir problemdir ve kullanıcı tarafından en iyi bilinen veya kullanılabilir olan sınıflandırıcı üzerine kurulurlar.

Sınıflandırıcı tasarımı için en basit ve güçlü yaklaşım benzerlik kavramı üzerine kurulmuştur. Benzer olan numuneler aynı sınıfa atanmalıdırlar. Bu yüzden, benzerliği tanımlamak için iyi bir metrik kurulduğunda, numuneler; sınıf başına bazı prototipler kullanılarak minimum mesafe sınıflayıcısı veya düzenleyici ile sınıflandırılabilir. Metrik veya prototip seçimi bu yaklaşımın başarısında çok önemlidir. En yakın orta sınıflandırıcıda seçilen prototipler çok basit ve sağlamdır. Her bir numune sınıfı tek bir prototiple temsil edilir. Prototipler bu sınıftaki tüm eğitim numunelerinin orta vektörüdür. Prototip hesaplaması için daha gelişmiş teknikler vektör kuantizasyonu (vector quantization) [29,30], öğrenme vektör kuantizasyonu (learning vector quantization) [22] ve yönetme ve yoğunlaştırma [31] gibi en yakın bir komşuluk karar kuralıyla (1-NN) ilgili veri indirgeme metotlarıdır. En kısa 1-NN kuralı, tüm sınıflandırıcılar için muhakeme işareti olarak yaygın şekilde kullanılabilir. Çünkü bu, uygulamaların çoğunda makul bir sınıflandırma performansı sağlama görüntüsü verir.

Numune sınıflandırıcı tasarımı için kullanılan ikinci ana kavram, olasılık yaklaşımı üzerine kurulur. Optimal Bayes karar kuralı (0/1 kayıp fonksiyonlu) maksimum Posterior olasılıklı sınıfa bir numune atar. Bu kural; yanlış sınıflandırmanın farklı tipleriyle ilgili maliyetleri hesaba katmak için değiştirilebilir. Bilinen sınıf şartlı yoğunluklar için Bayes karar kuralı optimum sınıflandırıcı verir. Bu durumda, verilen öncelikli olasılıklar, kayıp fonksiyonu ve sınıf şartlı yoğunluklar için diğer karar kurallarının hiçbiri daha düşük riske sahip olmayacaktır. Bu kayıp fonksiyonunun tahmin edilen değeridir mesela, hata ihtimali. Eğer öncelikli sınıf olasılıkları eşit ve 0/1 kayıp fonksiyonu adapte edilirse, Bayes karar kuralı ve maksimum olasılık karar kuralı tam olarak birleşirler. Pratikte deneyimsel Bayes karar kuralı veya plug-in kuralı kullanılır. Yoğunluk tahminleri doğru yoğunlukların yerine kullanılır. Bu yoğunluk tahminleri parametrik veya nonparametrikdir. Yaygın kullanılan parametrik modeller sürekli özellikler için çok değişkenli Gauss dağılımları [28], ikili özellikler için binomial dağılımlar ve tam değerli özellikler için çoklu normal dağılımlardır. Eğer farklı sınıflar için kovaryans matrisleri benzer olarak kabul edilirse, Bayes plug-in kuralı Bayes normal lineer olarak adlandırılır ve lineer bir karar sınırı sağlar. Eğer kovaryans matrisinin farklı olduğu kabullenilirse, Bayes normal kuadratik dediğimiz neticedeki Bayes plug-in kuralı kuadratik bir karar sınırı sağlar. Kovaryans matrisinin yaygın kullanılan maksimum olasılık tahminine ilave

olarak, küçük örnek ebadı durumlarında farklı dengeleme teknikleri [32] sağlam bir tahmin elde etmek için kullanılır.

İki meşhur parametrik olmayan karar kuralı,  $k'$  nın en yakın komşuluğu (k- nearest neighbor) (k-NN) kuralı ve Parzen sınıflandırıcısı tabiatında benzer iken pratikte farklı sonuçlar verir. İkisi de her bir serbest parametrenin birine sahiptir. Parzen Kernel' ini veya k komşuluk sayısını düzeltme parametresi hata oranının birini dışarıda bırakma tahminiyle optimize edilebilirler. Bu iki sınıflandırıcı, test numunesi ve eğitim setindeki tüm numuneler arasında mesafelerin hesaplanmasına da gereksinim duyar. Büyük hesaplama sayılarından kaçınmak için en uygun yol, eğitim setinin sistematik olarak indirgenmesidir.

Üçüncü kategori sınıflandırıcılar; kesin hata kriterinin optimizasyonu ile doğrudan karar sınırlarını inşa ederler. Bu yaklaşım seçilen metriğe bağlıken bu tip sınıflandırıcılar bazen asimtotik olarak Bayes sınıflandırıcılarına yaklaşabilirler. Bununla birlikte eğitime prosedürünün itici gücü; kriter minimizasyonudur. Örnek olarak sınıflandırıcı çıkışı ve bazı önceden ayarlanmış hedef değeri arasındaki orta kareli hata (mean squared error) (MSE) veya görünen sınıflandırma hatası verilebilir. Bu tip sınıflandırıcının klasik bir örneği, sınıflandırıcı çıkışı ve arzulanan etiketler arasındaki MSE' yi minimize eden Fisher' in lineer ayrımıdır.

Sınıflandırıcının özel bir tipi karar ağacıdır [33,34,35]. Ağacın her bir nodundaki en çarpıcı olan bireysel özelliklerin tekrarlı seçimiyle geliştirilir. Özellik seçimi ve ağaç jenerasyonu için olan kriter bilgi içeriğini, nod temizliğini ve Fisher kriterini içerir. Sınıflandırma sırasında bu özellikler hesaba katılır. Test numunesi için gerekli özellikler de incelenmektedir. Bu yüzden özellik seçimi kesin olarak içeride yapılır. En yaygın kullanılan karar ağacı sınıflandırıcıları tabiatında ikilidir ve her bir nodda tek bir özellik kullanır. Netice olarak bazı karar ağaçları çoğu uygulamalar için doğal olarak alt optimaldir. Ağaç sınıflandırıcının hızlı olmasının dışındaki diğer avantajı, bireysel özellikleri yüzünden karar kuralını yorumlama olasılığıdır. Karar ağaçları da sinir ağları gibi kolaylıkla aşırı eğitilebilirler. Bir budama aşaması kullanılarak bundan sakınılabılır [36,37,38]. Çizelge 7.1.'de en yaygın sınıflandırıcılar özetlenmiştir.

Çizelge 7.1. Sınıflandırma metotları

Metot	Özellik	Yorum
Düzenleme (Template matching)	Numuneyi en benzer şablona atar	Kullanıcı tarafından şablon ve metrik sağlanmalıdır,prosedür nonlineer normalizasyonu bulundurabilir,ölçek bağımlıdır
En Yakın Orta Sınıfı (Nearest mean classifier)	Numuneyi en yakın orta sınıfa atar	Hemen hemen geliştirmeye ihtiyaç yoktur, hızlı test, ölçek bağımlıdır
Alt Uzay Metodu (Subspace method)	Numuneyi en yakın alt uzay sınıfına atar	İnvaryantların normalizesi yerine,invaryantın alt uzayı kullanılır, ölçek bağımlıdır
1- En Yakın Komşuluk Kuralı (1-NN) (1-Nearest neighbor rule)	Numuneyi en yakın geliştirme numunesinin sınıfına atar	Geliştirmeye ihtiyaç yoktur, sağlam performans, yavaş test, ölçek bağımlıdır
k- En Yakın Komşuluk Kuralı (k-NN) (k- Nearest neighbor rule)	k'nın optimize edilmiş performans değeri kullanılarak k en yakın komşuluk arasındaki en üstün sınıfa numuneyi atar	Optimal asimtotik, ölçek bağımlı, yavaş test
Bayes Plug-in	Maksimum Posterior olasılık tahmininin sahip olduğu sınıfa numuneyi atar	Gaussian dağılımlı basit sınıflandırıcılar üretilir, yoğunluk tahmini hatasınaa duyarlıdır
Lojistik Sınıflandırıcı	Lojistik Posterior olasılıkları için maksimum olasılık kuralı	Lineer sınıflandırıcı,tekrarlı prosedür,farklı dağılımların bir ailesi için optimal, karışık veri tipleri için uygundur
Parzen Sınıflandırıcısı	Optimize edilmiş Kernel performansı ile Parzen yoğunluk tahmini için Bayes plug-in kuralı	Optimal asimtotik, ölçek bağımlı, yavaş test
Fisher Lineer Ayrımı (fisher linear discriminant)	MSE optimizasyonu kullanarak lineer sınıflandırma	Basit ve hızlı, benzer kovaryans matrisleri ile Gaussian dağılımları için Bayes plug-in'e benzer
İkili Karar Ağacı (Binary decision tree)	Özelliklerin numune bağımlı bir dizisi için eşğin bir setini bulur	Tekrarlı geliştirme prosedürü,aşırı geliştirme hassasiyeti,budama ihtiyacı,hızlı test imkanı
Perceptron	Lineer sınıflandırıcısının tekrarlı optimizasyonu	Geliştirme parametrelerine hassas,güven değerleri üretilebilir
Çok Katmanlı Perceptron (FFNN) (Multi layer perceptron )	Sigmoid transfer fonksiyonu kullanarak perceptronların iki yada daha fazla katmanının tekrarlı MSE optimizasyonu	Geliştirme parametrelerine hassas, yavaş geliştirme, nonlineer sınıflandırma fonksiyonu,güven değerleri üretilebilir,aşırı geliştirme hassasiyeti,düzenleme ihtiyacı
Radial Esaslı Ağ (Radial basis network)	Gaussian benzeri transfer fonksiyonu kullanarak sınırların en az bir katmanı ile FFNN' ün tekrarlı MSE optimizasyonu	Geliştirme parametrelerine hassas, nonlineer sınıflandırma fonksiyonu,güven değerleri üretilebilir,aşırı geliştirme hassasiyeti,düzenleme ihtiyacı,sağlam ana küme dışarısında bırakma
Destek Vektör Sınıflandırıcısı (SVC) (Support vector classifier)	Destek vektörünün minimum sayısının seçilmesi tarafından sınıflar arasındaki marjin maksimizasyonu	Ölçek bağımlıdır, tekrarlı, yavaş geliştirme, aşırı geliştirme hassasiyeti yoktur,iyi genelleştirme performansı



### 7.1. Sınıflandırıcı Kombinasyonu

Verilen bir sınıflandırma problemini çözmek için çoklu sınıflandırıcıları birleştirmenin bazı sebepleri vardır.

- 1- Bir tasarımcı, her biri farklı bölümde geliştirilmiş ve aynı problemin tamamen farklı bir temsili için olan, farklı sınıflandırıcı sayısının kullanımına sahip olabilir. Örnek olarak; ses, yüz ve el yazısı ile kişilerin tarif edilmesini verebiliriz.
- 2- Bazen, her biri farklı bir zamanda ve farklı bir çevrede toplanmış bir tek eğitim setinden daha fazlası kullanılabilir. Bu eğitim setleri farklı özellikleri bile kullanabilir.
- 3- Aynı veri üzerinde eğitilen farklı sınıflandırıcılar sadece global performanslarını farklılaştırmayabilirler aynı zamanda güçlü yerel farklar da gösterebilirler. Her bir sınıflandırıcı, en iyi performans gösterdiği özellik uzayındaki kendi bölgesine sahip olabilir.
- 4- Sinir ağları gibi bazı sınıflandırıcılar eğitime prosedüründeki serbestlik doğası yüzünden farklı başlangıçlarla farklı sonuçlar gösterirler. En iyi ağı seçip diğerlerini dışlamak yerine farklı ağlar, birleştirilebilir.

Farklı özellik setlerine, farklı eğitime setlerine, farklı sınıflandırma metotlarına veya farklı eğitim dönemlerine sahip olunabilir. Bunların tamamı, tümünden sınıflandırma doğruluğunu geliştirme ümidiyle çıkışları birleştirilebilen sınıflandırma seti sonucunu verirler. Eğer sınıflandırıcı seti sabitlenirse, problem kombinasyon fonksiyonu üzerine yoğunlaşır. Aynı zamanda sabit bir sınıflandırıcı kullanmak ve giriş sınıflandırıcılarının setini optimize etmek de mümkündür.

Tipik bir kombinasyon şeması bireysel sınıflandırıcı seti ile, son kararı vermek için bireysel sınıflandırıcıları birleştiren bir birleştiriciye sahiptir. Bireysel sınıflandırıcının ne zaman hatırlanacağı veya onların her biriyle nasıl ilişkiye gireceği, kombinasyon şemasının mimarisiyle belirlenir. Bu yüzden farklı kombinasyon şemaları; mimarilerinde, birleştirici karakteristiklerinde ve bireysel sınıflandırıcı seçiminde birbirlerinden farklı olabilirler.

Çoklu sınıflandırıcıları birleştiren farklı şemalar mimarilerine göre üç ana kategoride gruplandırılırlar.

- 1-Paralel
- 2-Kaskad
- 3-Hiyerarşik

Paralel mimaride bireysel sınıflandırıcıların tümü bağımsız olarak hatırlanır ve onların sonuçları bir birleştirici ile birleştirilir. Bireysel sınıflandırıcıların çıkışları seçilir. Kaskad mimaride bireysel sınıflandırıcılar lineer bir dizide hatırlanır. Verilen numune için mümkün sınıfların

sayısı hatırlanan dizideki sınıflandırıcılar gibi azaltılır. Randıman açısından hatalı fakat düşük sayma ve ölçme talepleri olan sınıflandırıcılar ilk önce düşünülür, daha doğru fakat yüksek sayma ve ölçme talepleri olan sınıflandırıcılar bunu takip eder. Hiyerarşik mimaride ise bireysel sınıflandırıcılar bir yapıda birleştirilir. Bu yapı karar ağacı sınıflandırıcısına benzerdir. Bununla beraber çok sayıda özellik talep eden kompleks sınıflandırıcılarla da bağlanabilir. Bu mimarinin avantajı, yüksek randıman oranı ve özelliklerin farklı tiplerinin ayırım gücünün işletilmesindeki esnekliktir. Bu üç temel mimari kullanılarak çok daha karışık sınıflandırma kombinasyon sistemleri inşa edilebilir.

Eğer bireysel sınıflandırıcılar büyük miktarda bağımsızsa bir sınıflandırıcı kombinasyonu faydalı olabilir. Eğer bu, eğitime setlerinin kullanımı ile garanti edilmezse, rotasyon ve Bootstrapping gibi yeniden örneklendirme teknikleri bazı farklılıkları yapay olarak oluşturmak için kullanılabilir. Yığmada bireysel sınıflandırıcıların çıkışları, yığılmış sınıflandırıcıyı eğitmek için kullanılır. Son karar, bireysel sınıflandırıcıların çıkışlarıyla birlikte, yığılmış sınıflandırıcıların sonuçları üzerine kurularak yapılır.

Çantalamada farklı veri setleri, orijinal veri setinden bootstrap yapılmış versiyonlarıyla meydana getirilir ve ortalamaya benzer sabit bir kural kullanılarak birleştirilir. Destekleme; eğitime veri setlerinin sıralı bir üretimi için diğer bir yeniden örnekleme tekniğidir. Sıradaki özel eğitime setinin dağılımı numunelerle temsil edilir. Bu numuneler sırada bulunan önceki sınıflandırıcılarla yanlış sınıflanmıştır. Desteklemede bireysel sınıflandırıcılar, özellik uzayındaki daha kompleks bölgeleri ayırt etmeyi öğrenmek için hiyerarşik olarak eğitilirler.

Alt sınıfların eğitime setindeki bireysel sınıflarını ayırmak için bazen küme analizleri de kullanılabilir. Netice olarak daha basit sınıflandırıcılar (lineer) kullanılabilir ve mesela parça yönünde lineer bir sonuç üretmek için daha sonra birleştirilir.

Eğitime numunelerinin farklı setleri üzerinde farklı sınıflandırıcılar yapma yerine, farklı özellik setleri kullanılabilir. Bu, bağımsız bilgi içermek için bireysel sınıflandırıcıları daha açık bir şekilde zorlar. Serbest alt uzay metodu buna bir örnek teşkil eder.

Bireysel sınıflandırıcılar seçildikten sonra, birleştirici adı verilen bir modülle birleştirilmesi gereklidir. Çeşitli birleştiriciler bireysel sınıflandırıcının çıkışı üzerinde, eğitilebilirlik, uyum ve gereksinim yönünden birbirinden ayrılabilirler. Oy verme, ortalama alma (veya toplama) ve Borda sayıcı [39] gibi birleştiriciler statiktir. Diğerleri eğitilebilirken bunların eğitmeye ihtiyacı yoktur.

Bazı kombinasyon şemaları; giriş numunesine bağlı bireysel sınıflandırıcıların kararlarını değerlendiren birleştirici durumunda uyumludur. Uyumsuz birleştiriciler ise bütün giriş numunelerine aynı muameleyi gösterirler.

Farklı birleştiriciler, bireysel sınıflandırıcılardan farklı çıkış tiplerini beklerler. Xu [40] bu beklentileri üç seviyede gruplandırdı.

1-Ölçüm veya Güven (measurement or confidence)

2-Rütbe (rank)

3-Özet (abstract)

Güven seviyesinde bir sınıflandırıcı, görüş veya ihtimal gösteren her bir sınıfın nümerik değerini çıkış olarak verir ki verilen giriş numunesi bu sınıfa aittir. Rütbe seviyesinde bir sınıflandırıcı ilk seçim olan en büyük rütbeyle her bir sınıfa bir rütbe atar. Rütbe değeri izolasyonda kullanılmaya bilir. Çünkü en yüksek rütbe sınıflandırmada yüksek güven anlamına gelmez. Özet seviyesinde bir sınıflandırıcı sadece tek sınıf etiketini veya bazı sınıf etiketlerini sonuç olarak verir. Özet seviyesi verilen kararın en az miktardaki bilgisini içerirken, güven seviyesi en zengin bilgiyi nakleder.

Çizelge 7.2., kombinasyon şemalarını ve onların karakteristiklerini listeler.

Çizelge 7.2. Sınıflandırıcı kombinasyon şeması

Şema	Mimari	Eğitme	Uyum	Bilgi Seviyesi	Yorumlar
Oylama (voiting)	Paralel	Yok	Yok	Özet	Bağımsız sınıflandırıcılar kabul eder
Toplam, Orta, medyan (sum, mean, median)	Paralel	Yok	Yok	Güven	Sağlam,bağımsız güven tahminçileri kabul eder
Genelleştirilmiş asamble (generalized ensemble)	Paralel	Var	Yok	Güven	Hata korelasyonunu düşündür
Uyumlu Ağırlık (adaptive weighting)	Paralel	Var	Var	Güven	Yerel uzmanlıkları keşfeder
Yığıma (stacking)	Paralel	Var	Yok	Güven	Geliştirme verisinin iyi kullanımı
Borda sayıcı (Borda count)	Paralel	Var	Yok	Rütbe	Rütbeyi güvene dönüştürür
Lojistik Regresyon	Paralel	Var	Yok	Rütbe güven	Rütbeyi güvene dönüştürür
Sınıf seti indirgeme (class set reduction)	Paralel Kaskad	Var/yok	Yok	Rütbe güven	Tesirli
Dempster-Shafer	Paralel	Var	Yok	Rütbe güven	İhtimalli olmayan güven sigortası
Fuzzy integrali	Paralel	Var	Yok	Güven	İhtimalli olmayan güven sigortası
Yerel uzmanın karışımı (mixture of local experts) MLE	Kapılı paralel	Var	Var	Güven	Yerel uzmanlıkları keşfeder,optimizasyonla birleşir
Hiyerarşik MLE	Kapılı paralel hiyerarşik	Var	Var	Güven	MLE'nin aynısı,hiyerarşik
Çağrışimli anahtar (associative switch)	Paralel	Var	Var	Özet	MLE'nin aynısı,fakat optimizasyonla birleşmez
Çantalama (bagging)	Paralel	Var	Yok	Güven	Kıyaslanabilir sınıflandırıcılara ihtiyaç duyar
Destek (boosting)	Paralel hiyerarşik	Var	Yok	Özet	Marjini geliştirir,aşırı geliştirmeye benzemez,yanlış etiketlemeye karşı hassastır, Kıyaslanabilir sınıflandırıcılara ihtiyaç duyar
Serbest Alt uzay (random subspace)	Paralel	Var	Yok	Güven	Kıyaslanabilir sınıflandırıcılara ihtiyaç duyar

## 8. HATA TAHMİNİ

Sınıflandırma hatası veya basit hata oranı olan  $P_e$ , sınıflandırıcı performansının nihai ölçüsüdür. Sınıflandırıcıları karşılaştırma onların hata olasılıkları üzerine kurularak da değerlendirilebilir. Diğer performans ölçüleri, ölçme özelliklerinin maliyetini ve karar kuralının hesaplanabilir gereksinimlerini içerirler. Sınıflandırma yoğunlukları yüzünden hata ihtimalini tanımlamak kolayken,  $P_e$  için kapalı form ifadesini elde etmek çok güçtür. Eşit olmayan kovaryans matrisli, çok değişkenli Gauss yoğunluklarının göreceli olarak basit olduğu durumda bile; hata oranı için basit bir analitik ifade yazmak mümkün değildir.

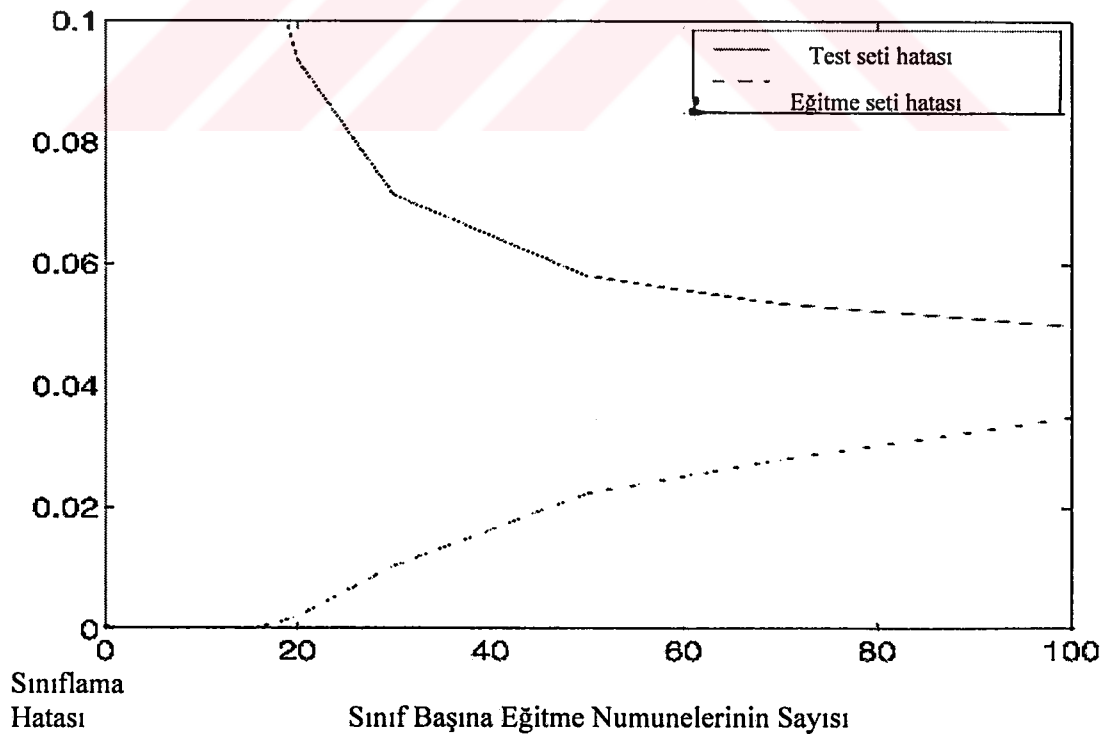
Pratikte, bir tanıma sisteminin hata oranı, eğitime ve test setlerine bölünmüş [41] tüm kullanılabilir numunelerden tahmin edilmek zorundadır. Eğitim numuneleri kullanılarak ilk önce sınıflandırıcı tasarlanır. Sonra aynı test numuneleri üzerinde onların sınıflandırma performansı üzerine kurularak değerlendirilir. Yanlış sınıflandırılmış test numunelerinin yüzdesi hata oranının bir tahmini olarak alınır. Gelecek sınıflandırma performansını tahminde bu hata tahmininin güvenilir olması için sadece eğitim ve test setinin yeterli miktarda büyük olması yetmez, aynı zamanda eğitim ve test numuneleri bağımsız olmak zorundadırlar.

Kullanılan özel eğitim ve test setlerinin bir fonksiyonu olan sınıflandırıcı hata tahmini serbest değişkenlidir. Verilen bir sınıflandırıcı için  $\tau$ ;  $n$  toplamının dışında yanlış sınıflandırılan test numunelerinin sayısıdır.  $\tau$ ' in olasılık yoğunluk fonksiyonunun, binomial bir dağılıma olduğu gösterilebilir. Maksimum olasılık tahmini  $\hat{P}_e = \tau/n$ ,  $E(\hat{P}_e) = \hat{P}_e$  ve  $Var(\hat{P}_e) = P_e(1 - P_e)/n$  ile verilir. Bu yüzden  $\hat{P}_e$ , ön yargıda bulunulmamış ve tutarlı bir tahmincidir.  $n=250$ ,  $\tau = 50$  iken  $\hat{P}_e=0.2$  ve  $\hat{P}_e$ ' nin %95 güven aralığı (0.15 , 0.25)' dir. Test numunelerinin  $n$  sayısı artarken daralan güven aralığı,  $c_1$  ve  $c_2$  sınıflandırıcılarını kıyaslamada önemli bir rol oynar. 100 test numunesinin kullanıldığı ve  $c_1$  ile  $c_2$ ' nin bu örneklerin 10' u ile 13' ünü yanlış sınıflandırdığını ele alalım.  $c_1$ ' in  $c_2$ ' den daha iyi sınıflandırdığı söylenebilir mi? Bu sınıflandırıcıların doğru hata olasılıkları için %95 güven aralıkları (0.04,0.16) ve (0.06,0.20)' dir. Bu güven aralıkları çakıştığı için  $c_1$  performansının her zaman  $c_2$ ' nin performansından daha üstün olduğunu söyleyemeyiz.

Eğer eğitim seti küçükse, sonuçtaki sınıflandırıcı çok sağlam olmayacak ve küçük genelleştirme yeteneğine sahip olacaktır. Eğer test seti küçükse, tahmin edilen hata oranındaki güvende düşük olacaktır. Hata oranını tahmin etmek için yaygın olarak kullanılan çeşitli

metotlar çizelge 8.1.'de gösterilmiştir. Bu metotlar, onların eğitime ve test setleri olarak mevcut örnekleri nasıl kullandığını ayırır. Eğer kullanılabilir örnek sayısı çok fazla ise bu metotların tamamı hata oranının aynı tahminini benzer olarak yönetir. Yeniden yerine koyma metodu (resubstitutions method), hata oranının optimistik olarak ön yargılı tahminini sağlarken, ön yargı daha küçük olur ve sınıf başına eğitime örneğinin sayısının, özellik vektörü boyutuna oranı arttıkça büyür. Kullanılan numunelerin eğitime ve test setlerine nasıl bölüneceğini gösteren iyi bir kılavuz yoktur. Veri, eğitime ve test setlerine nasıl ayrılırsa ayrılırsın farklı rasgele bölmeler farklı hata tahminlerine neden olacaktır.

Şekil 8.1., eğitime numunelerinin sayısının bir fonksiyonu olarak dijit veri seti üzerinde Bayes plug-in lineer sınıflandırıcısının sınıflandırma hatasını gösterir. Eğitime örneklerinin sayısı artarken, test seti hatası eğitime seti hatasına yaklaşır. Sınıf başına 100 eğitime örneği için, bu iki hata oranı arasındaki göreceli olarak büyük olan fark bize şunu gösterir: Bu iki hata oranı arasındaki ön yargı, eğitime setinin büyümesiyle azaltılabilir. Bu şekildeki eğriler 50 deneyin ortalamasını temsil eder ve verilen ebadın eğitime setlerinde rasgele çizilirler. 1000 numuneli test seti sabitlenmiştir.



Şekil 8.1. Dijit veri seti üzerinde eğitime ve test setleri için eğitime örneklerinin bir fonksiyonu olarak Bayes plug-in lineer sınıflandırıcısının sınıflama hatası

Çizelge 8.1. Hata tahmin metotları

Metot	Özellik	Yorum
Yeniden Yerine Koyma (Resubstitutions method)	Hem test hem de geliştirme için kullanılabilir tüm veriler kullanılır, geliştirme ve test setleri birbirinin aynıdır	Ön yargı tahmini optimistiktir, özellikle örnek ebadının boyutluluğa oranı küçük olduğunda
Dayanma (holdout method)	Geliştirme için yarım veri kullanılır ve kalan veri test için kullanılır, geliştirme ve test setleri bağımsızdır	Önyargı tahmini pessimistiktir, farklı bölünmeler farklı tahminler verecektir
Birini Dışarıda Bırakma (leave one out method)	Bir sınıflandırıcı dizaynında (n-1) örnek kullanılır ve biri üzerinde kalan örnek değerlendirilir; bu, (n-1) ebadının farklı geliştirme setleri ile n kez tekrarlanır	Ön yargısız tahmin, fakat geniş bir varyansa sahip; geniş hesaplama ihtiyacına sahiptir çünkü n farklı sınıflandırıcı dizayn edilmek zorundadır.
Rotasyon Metodu n Katmanlı Çapraz Değer (rotation method, n- fold cross validation)	Dayanma ve birini dışarıda bırakma metotları arasında bir çelişki yaratır, P birleşme alt setlerine kullanılabilir örnekleri böler, $1 \leq P \leq n$ . Geliştirme için (P-1) alt setleri ve test için kalan alt setler kullanılır.	Dayanma metodundan daha düşük ön yargı tahmini ve birini dışarıda bırakma metodundan daha ucuz gerçekleştirir.
Bootstrap Metodu	Bootstrap örnekleri kullanarak bir takım tahminler hata oranını tanımlamıştır	Bootstrap tahminleri birini dışarıda bırakma metodundan daha düşük varyansa sahiptir, hesaplama talepleri daha fazladır, küçük örnek ebadı durumları yararlıdır

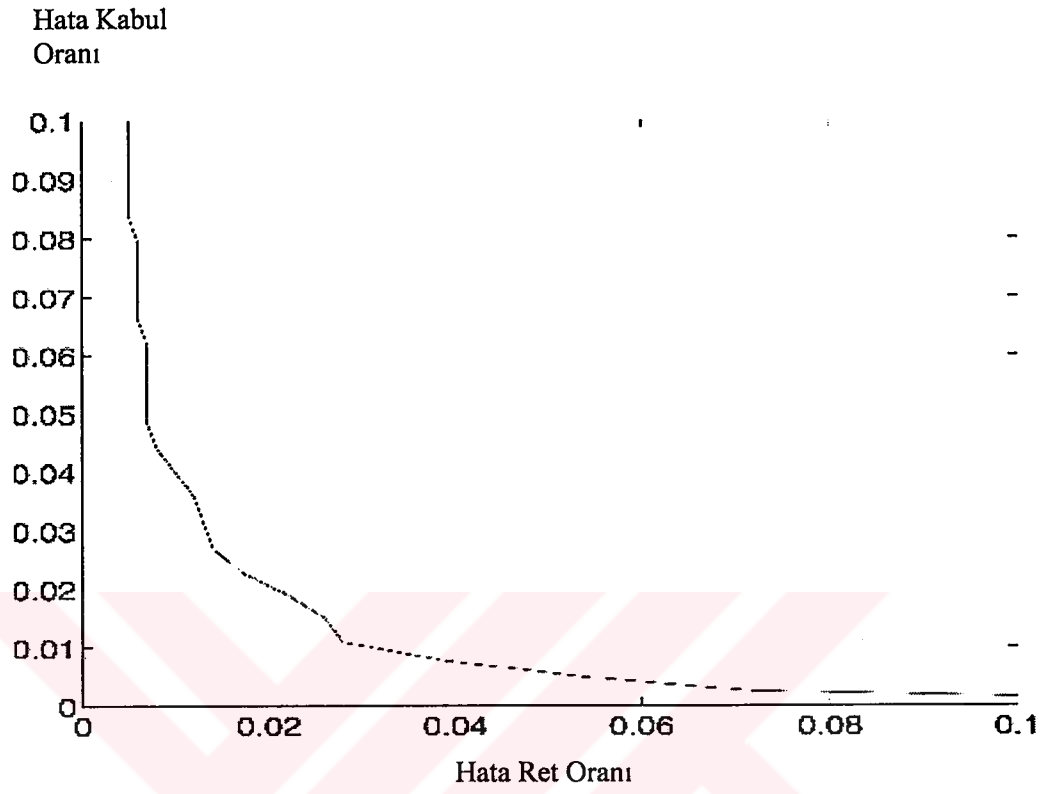
Dayanma, birini dışarıda bırakma ve rotasyon metotları; çapraz değerli yaklaşımın versiyonlarıdır. Özellikle küçük numune ebadı durumları için çapraz değerli metotların ana dezavantajı; eğitime sınıflandırıcısı için kullanılabilir olan n numunenin tamamının kullanılamaz olmasıdır. Dahası, çapraz değerliliğin iki uç durumu olan dayanma ve birini dışarıda bırakma metodu; gerek önyargıdan gerekse varyanstan sıkıntı çekerler. Bu sınırlamanın üstesinden gelmek için bootstrap metodu, hata oranını tahmin etmek amacıyla önerilmiştir. Bu metot; verilen eğitime seti gibi aynı ebadın “sahte” veri setlerinin sayısını üretmek için kullanılabilir numunelerin yer değiştirmesiyle yeniden örnekler. Bu yeni eğitime setleri sadece yeniden yerine koyma önyargısının tahmini için kullanılmaz, aynı zamanda hata oranının bootstrap tahmini olarak adlandırılan diğerini de tanımlamak için kullanılır. Bootstrap tahminleri, çapraz değerli tahminlerden ve hata oranının [42] yeniden yerine koyma tahminlerinden daha başarılıdır.

Pek çok numune tanıma uygulamasında; sistemin toplam hata oranını ölçen  $\hat{P}_e$  tek sayısı ile sınıflandırıcı performansını karakterize etmek yeterli değildir. Parmak izi çakışma (uyuşma) sisteminin değerlendirilmesi problemini ele alalım. Burada hata oranıyla ilgili iki farklı durum bizi ilgilendirir. Hata kabul oranı (false acceptance rate) (FAR); verilen bir sistemle yanlış olarak çakışmış farklı parmak izlerinin bir çiftinin, çakışma teşebbüsleri toplam

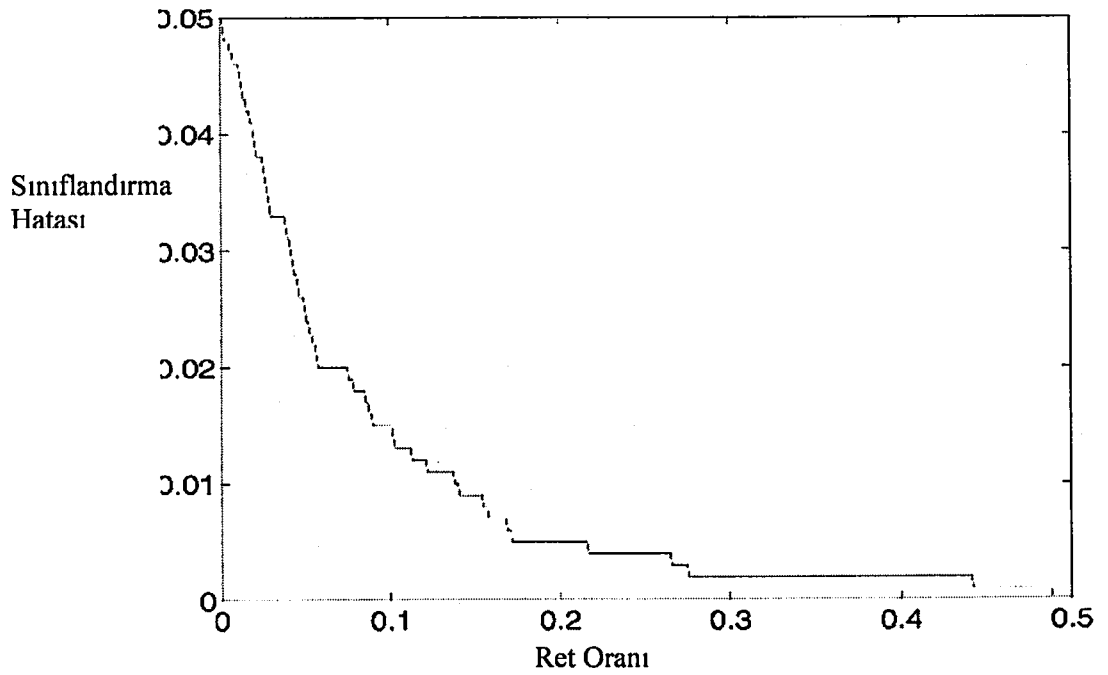
sayısına oranıdır. Hata ret oranı (false reject rate) (FRR); verilen bir sistemle uyuşmayan aynı parmak izinin bir çiftinin, çakışma teşebbüslerinin toplam sayısına oranıdır. Bir parmak izi düzenleme sistemi, FAR' ın arzulan değerinde çalışmak için ayarlanabilir. Eğer sistemin FAR' ını azaltmaya teşebbüs edilirse, FRR artacak ve bu karşılıklı olarak sürecektir. Alıcı çalıştırma karakteristiği (receiver operating characteristic) (ROC) eğrisi [43]; FAR' ın FRR' ye karşı görünüşüdür. Bu durumda ROC, sistemin eşit hata oranından (FRR=FAR) daha kapsamlı bir performans ölçümü sağlar. Şekil 8.2., dijit veri seti için ROC eğrisini gösterir. Burada Bayes plug-in lineer sınıflandırıcısı sınıf başına 100 numune üzerinde eğitilir. ROC analizlerinin kullanım örnekleri, sınıflandırıcı ve özellik seçimini birleştirir.

Hata oranına ilave olarak, sınıflandırıcının diğer faydalı performans ölçüsü, onun ret oranıdır. İki sınıf arasında karar sınırına yakın test numune düşüşlerini ele alalım. Karar kuralı bir numuneyi doğru olarak sınıflandırabilecek iken, bu sınıflandırma düşük bir güvenle yapılacaktır. Daha iyi bir alternatif; şüpheli numuneleri düşünülen kategorilerden birine atamak yerine bunları reddetmektir. Bir test numunesinin reddine nasıl karar verilir? Bayes karar kuralı için, eğer sonraki olasılık eşğin altında ise, iyi bilinen ret opsiyonu bir numuneyi reddetmek içindir. Bu ilişki; sınıflandırıcının arzulan çalışma noktasını kurmak için kullanılan hata ret trade-off eğrisi gibi temsil edilir. Şekil 8.3.; Bayes plug-in lineer sınıflandırıcısı kullanılırken dijit veri seti için hata ret eğrisini gösterir. Daha fazla numuneyi reddetme hata oranını azaltıp veya aynı tutacağından, bu eğri monoton olarak artmaz. Ret oranı için iyi bir seçim; ret veya yanlış kararlarla birleştirilmiş maliyetler üzerine kurulmuştur.





Şekil 8.2. Dijit veri seti için Bayes plug-in lineer sınıflandırıcısının ROC eğrisi



Şekil 8.3. Dijit veri seti için Bayes plug-in lineer sınıflandırıcısının hata ret eğrisi

## 9. DENETLENMEMİŞ SINIFLANDIRMA

Numune tanımının çoğu uygulamalarında, gerçek kategorili bir eğitme örneğini güvenilir olarak etiketlemek çok zordur. Denetlenmemiş sınıflandırma veri kümeleme olarak da bilinir. Bu genel bir etiket olup, numuneler arasında ölçülen veya fark edilen benzerlikler üzerine kurulu çok boyutlu bir veride, tabii grup veya kümeleri bulmak için tasarlanan bir prosedür türüdür.

Denetlenmemiş sınıflandırmada veri, farklı şekil ve ebatlarda kümeleri ortaya çıkarır. Bir kümenin fonksiyonel tanımları; 1-Küme içindeki numuneler farklı kümelere ait olanlara göre birbirlerine daha benzerdirler. 2-Bir küme; noktaların göreceli olarak düşük yoğunluğuyla diğer kümelere ayrılmış, göreceli olarak yüksek yoğunluklu noktaları içerir. Bir kümenin bu fonksiyonel tanımlamalarıyla bile, kümenin işlemsel tanımlamasının üstesinden gelmek kolay değildir.

Kümelendirme algoritmalarının çoğu, iki tip kümelendirme tekniği üzerine kurulmuştur. Bunlar; tekrarlı kare hata parçalı kümelendirmesi ve yığılmalı hiyerarşik kümelendirmedir. Hiyerarşik teknikler, bir dendogram veya ağaç formunda görülebilen grupların birleşmiş dizisindeki veriyi organize ederler. Kare hatalı parçalı algoritmalar şunu elde etmeye teşebbüs ederler: Parçalama; küme dağılımı içinde minimizasyon, küme dağılımı arasında maksimizasyon yapar. Boyut sayısı  $d$  olan numunenin  $n$  tane toplam mümkün bölümünü, hesaplanması fizibil olmayan  $K$  kümede incelemek zorundadır. Bu yüzden, bu araştırmayı azaltmak için çeşitli yöntemler kullanılır. Fakat burada optimallik garantisi yoktur.

Parçalı kümelendirme problemi şu şekilde kurulabilir:  $d$  boyutlu metrik uzayda verilen  $n$  numuneleri,  $K$  kümeye numune parçalanmasını belirler. Kümedeki numuneler, farklı kümelere göre birbirlerine daha benzerdir.  $K$  değeri belirlenebilir veya belirlenmeyebilir. İster global, ister lokal olsun kümelendirme kriteri adapte edilmelidir. Kare hata gibi global bir kriter, bir prototip vasıtasıyla her bir kümeyi temsil eder ve en benzer prototiplere göre numuneleri kümelere atar. Lokal bir kriter, veride lokal yapıyı kullanarak kümeleri teşkil eder. Mesela, numune uzayında yüksek yoğunluklu bölgeler tanımlanarak veya bir numune ve onun en yakın  $k$  komşuluğu aynı kümeye atanarak kümeler oluşturulabilir.

### 9.1. Kare Hata Kümelendirme

En yaygın kullanılan parçalı kümelendirme stratejisi; kare hata kriteri üzerine kurulmuştur. Genel amaç kümelerin sabit sayısı için kare hatasını minimize eden parçayı elde

etmektedir. Boyut sayısı  $d$  olan  $n$  numunelerinin verilen setinin,  $K$  kümelerine  $\{C_1, C_2, \dots, C_k\}$  parçalandığını düşünelim. Burada  $C_k$  kümesi,  $n_k$  numunelerine sahiptir ve her bir numune tam olarak bir kümededir  $\sum_{k=1}^K n_k = n$ .

Orta vektör veya  $C_k$  kümesinin merkezi kümenin sentroidi veya  $m^{(k)} = (1/n_k) \sum_{i=1}^{n_k} x_i^{(k)}$  olarak tanımlanır. Burada  $x_i^{(k)}$ ;  $C_k$  kümesine ait  $i$ . numunedir.  $C_k$  kümesinin kare hatası;  $C_k$  ve onun  $m^{(k)}$  küme merkezi arasındaki her bir numune arasındaki kareli öklid mesafelerinin toplamıdır. Bu kare hatası, küme varyasyonu içinde de adlandırılabilir.

$$e_k^2 = \sum_{i=1}^{n_k} (x_i^{(k)} - m^{(k)})^T (x_i^{(k)} - m^{(k)}) \quad (7)$$

$K$  kümelerini içeren tüm kümelendirme için kare hatası, küme varyasyonu içindekilerin toplamıdır.

$$E_K^2 = \sum_{k=1}^K e_k^2 \quad (8)$$

Kare hata kümelendirmenin amacı, sabit bir  $K$  için  $E_K^2$ 'yi minimize eden  $K$  kümelerini içeren bir parçalanmayı bulmaktır. Tekrarlı parçalı kümelendirme metodu için genel bir algoritma aşağıda verilmiştir.

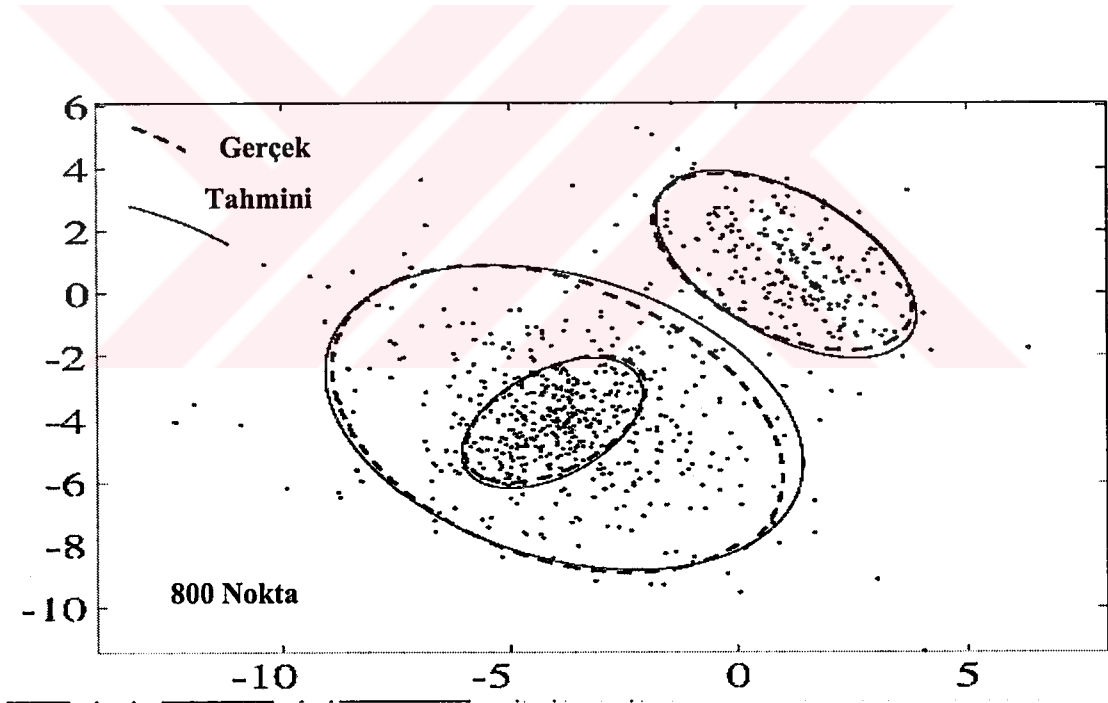
- 1.Adım:  $K$  kümeleriyle bir başlangıç parçalaması seç. Küme üyeliği kararlı oluncaya kadar  $2^2$  den  $4^2$  e kadar olan adımları tekrar et.
  - 2.Adım: Her bir numuneyi en yakın küme merkezine atayarak yeni parçalanma üret.
  - 3.Adım: Kümelerin sentroidleri olarak yeni küme merkezlerini hesapla.
  - 4.Adım: Kriter fonksiyonunun optimum değeri bulununcaya kadar, 2 ve 3 adımlarını tekrarla.
  - 5.Adım: Mevcut kümeleri birleştirerek veya parçalayarak yada küçük veya uzaktaki kümelerin kaldırılmasıyla küme sayılarını ayarla.
5. adım olmadan bu algoritma,  $K$  ortaları algoritması olarak bilinir.

Eğer kümeler sağlamsa yukarıda tarif edilen basit  $K$  ortaları parçalı kümeleme algoritması, sayısal olarak etkilidir ve sürpriz şekilde iyi sonuçlar verir. Kümelerin sağlamlığı; şekil olarak hiperküresel ve özellik uzayında iyi ayrılmış olmalarıdır. Eğer (8)'deki kareli hatayı tanımlamada Mahalanobis mesafesi kullanılırsa, algoritma hiperelipsoidal biçimli kümeleri bile bulabilir.

## 9.2. Karışım Parçalaması

Sonlu karışımlar esnek ve güçlü olasılık modelleme araçlarıdır. Karışımlar; her bir numunenin olasılıkla modellenmiş alternatif kaynak [44] setlerinden biriyle üretilen durumları, yeterli olarak modellerler. Karışımlar; kompleks olasılık yoğunluk fonksiyonlarını keyfi olarak temsil edebilecek model sınıfı olarak da görülebilir. Sonlu karışımlar, özellik seçme aracı [45] olarak da kullanılabilirler.

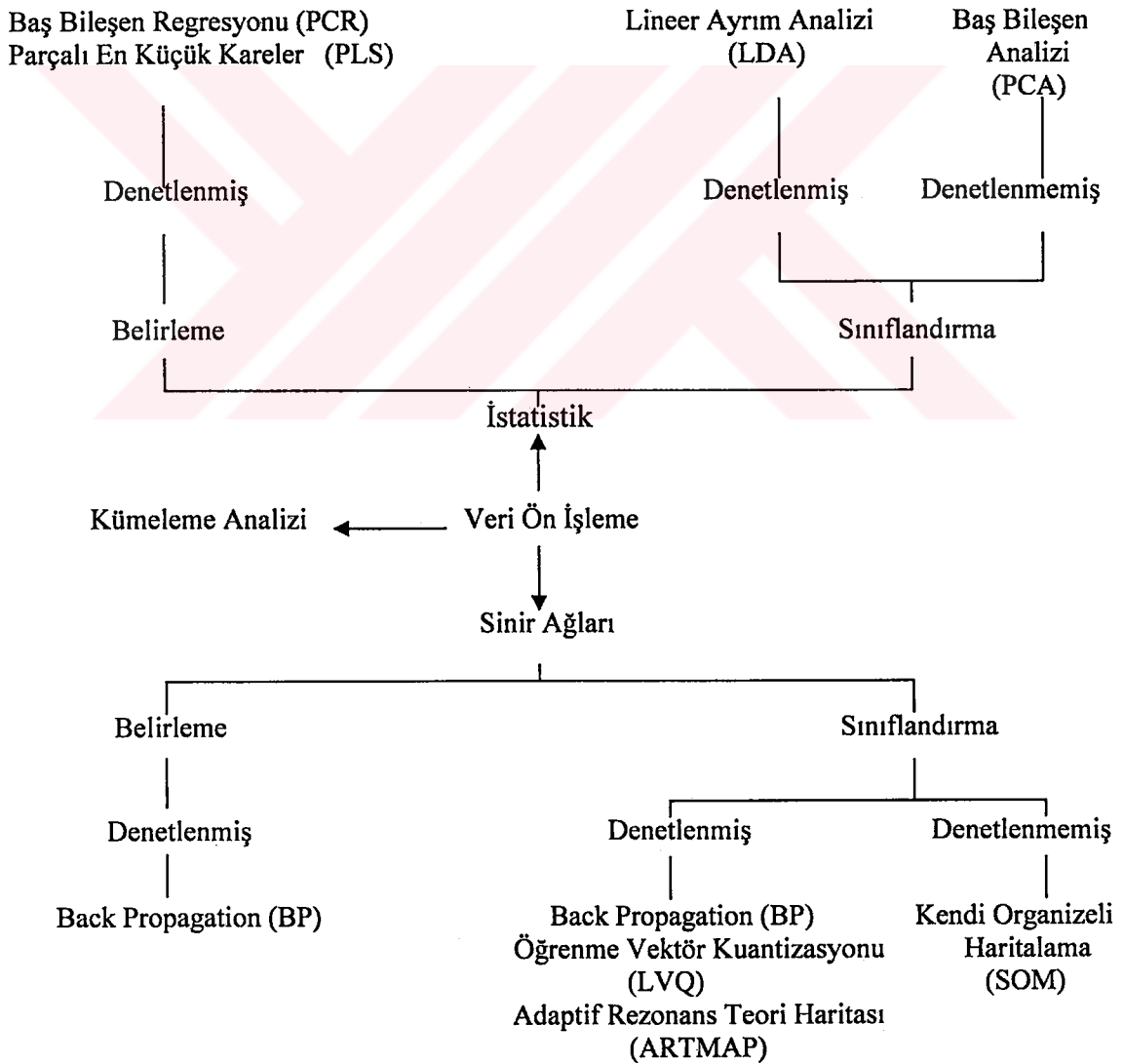
Şekil 9.1; karışım parçalama örneğini gösterir. Burada K; modifiye edilmiş minimum tarifleme uzunluğu (minimum description length) MDL kriteri [46,47] kullanılarak seçilir. Veri; üç Gauss komponenti üzerine dağılmış 800 tane iki boyutlu numuneye sahiptir. Komponentlerin ikisi aynı orta vektöre sahiptir. K ortaları gibi kümelendirme algoritması; bu komponentlerin ikisinin bir hayli çakışması sebebiyle bu üç komponenti tanımlayamayacaktır.



Şekil 9.1. Karışım parçalama örneği

## 10. KİMYASAL SENSÖR DİZİSİ VERİ ANALİZİ METOTLARI

Kimyasal sensör dizilerinin veri analizleri için kullanılabilir pek çok veri işleme yöntemi vardır. İşlenmemiş sensör cevapları sıklıkla ön işleme tabi tutulur ve bu işlenmemiş veriler daha sonra, çok değişkenli analiz tekniğinde kullanılır. Şekil 10.1. ; ister istatistik teknik olsun isterse sinir ağı esaslı yaklaşım olsun mevcut metotların bazılarının gruplaşma yollarını gösterir. İlave gruplaşmalar, teknik için gerekli bir veriyle tanımlanır. Bunlar, denetlenmiş metotlar olarak adlandırılırken bağımsız değişken bilgisini(örneğin analit sınıfları) kullanırken, sadece denetlenmemiş metotlar denem bağımsız değişken bilgisine (sensör cevapları) gereksinim duyarlar.



Şekil 10.1. Sensör dizisi veri analizi metotlarının gruplaşma yolları

Veri ön işleme diyagramının merkezinde görünür. Çünkü bu, seçilen veri analiz metodu hiç önemsenmeden tipik olarak yapılır. Diyagramın sağ üstünde görülen ana komponent analizi muhtemelen en sık kullanılan tekniktir. Ana komponentlerin hesaplanmasını takiben veriyi görüntü olarak incelemek için planlar yapılır. Ana komponent analizinin sonuçları, diğer tekniklerin herhangi birinin girişi olarak da kullanılabilir.

Şekil 10.1.'de gösterilen veri analizi metotlarının uygun olanının seçimi uygulamaya bağlıdır. Bu seçim; uygulamanın daha özel detaylarına ve operatör seçimine bağlıdır. Denetlenmemiş metotlar, verideki ilişkilerin keşfedilmesi gibi nitelik uygulamalarının en iyisidir. Denetlenmiş metotlar nicelik uygulamaları için, hangi sınıfın özel bir gözleme ait olduğunu belirlemede kullanılır.

Uygun veri analizi metodunun seçimi yüksek oranda veri tabiatına ve özel uygulamaya bağlıdır. Tüm durumlar için doğru olan genel bir metot yoktur. Bununla beraber, tipik bir analizdeki adımlar genellikle seçilen metot önemsenmeden birbirine benzerdir. İlk önce, ilgilenilen analitler için sensör dizi cevaplarının bir veri seti kurulur. İkinci olarak, uygulama ve veri analizinin seçilen bir metodu için veri, uygun bir şekilde ön işlemeye tabi tutulur. Üçüncü olarak kullanılacak özellikler seçilir. Her bir sensörün sabit durum cevabı sıklıkla, numune tanıma tekniklerinin bir girişi olarak kullanılır. Diğer sık kullanılan opsiyonlar, kullanım için sensörlerin alt setinin seçimini veya ana komponent analizinin yapımını ihtiva ederler ve daha fazla analiz için giriş olarak komponentleri kullanırlar. Dördüncü olarak, seçilen metot kullanılarak analiz gerçekleştirilir.

Beşinci bir adım her zaman olmayabilir fakat sürekli olarak sınıflandırmada gerçekleştirilmelidir. Niceliklendirme uygulaması; model oluşturmada kullanılmayan veriyi kullanarak, veri eğitime setinin teşkil ettiği modellerin geçerliliğidir. Model geçerliliği için diğer prosedür, bağımsız değişkenin mücadelesini ve mücadele edilen bağımsız değişken kullanılarak modellerin test edilmesini içerir. Eğer sonuçlar iyiye; bu, veri analizinde bir hata olduğunun göstergesidir. Eğer sonuçlar iyi değilse; bu geçerli analiz için daha fazla destektir. n'yi dışarıda bırakma geçerliliğidir. Modelleri garantilemek için kullanılan üçüncü bir araçtır ve sadece onları meydana getirmek için kullanılan eğitime seti verisine uygulanamaz.

### 10.1. Linear Ayrım Analizi (Linear Discriminant Analysis) (LDA) ve Uygulamaları

LDA; yeni objeleri uygun sınıflara atamak veya obje sınıflarını seçmek için kullanılabilir. Discriminantlar, ölçülmüş değişkenlerin lineer kombinasyonudur, mesela sensör

cevapları. Discriminant fonksiyonları; sınıfların içindeki varyasyona rölatif sınıflar arasındaki mesafeyi maksimize eden bir amaçla hesaplanırlar.

A ve B sınıflarını düşünelim. Bunlar p sensörlerinde izlenen  $n_A$  ve  $n_B$  numunelerine sahip olsunlar. Veri;  $X_A(p \times n_A)$  ve  $X_B(p \times n_B)$  matrislerinde toplanabilir. Bu veri matrislerinden örnek orta vektörler şöyle hesaplanır:

$$X_A^* = 1/n_A \sum_{i=1}^{n_A} X_{Ai} \quad X_B^* = 1/n_B \sum_{i=1}^{n_B} X_{Bi} \quad (9)$$

Burada  $X_{Ai}$  ve  $X_{Bi}$ ;  $X_A$  ve  $X_B$ 'nin i. kolonlarını gösterirler. Örnek kovaryans matrislerinden şunlar hesaplanabilir:

$$S_A = 1/(n_A - 1) \sum_{i=1}^{n_A} (X_{Ai} - X_A^*)(X_{Ai} - X_A^*)^T$$

$$S_B = 1/(n_B - 1) \sum_{i=1}^{n_B} (X_{Bi} - X_B^*)(X_{Bi} - X_B^*)^T \quad (10)$$

$X_A$  ve  $X_B$  ile temsil edilen doğru popülasyonlar için eşit kovaryans matrislerini kabul edelim [48,49]. Kovaryans matrisleri aşağıdaki gibi birleştirilebilir:

$$S_p = (n_A - 1)S_A + (n_B - 1)S_B / (n_A + n_B - 2) \quad (11)$$

Numune popülasyonlarının müşterek kovaryans matrisinin tek bir tahmini böyle şekillendirilir. Bu bilgiyle beraber discriminant fonksiyonları şu şekilde hesaplanabilir:

$$y = (X_A^* - X_B^*)^T S_p^{-1} X \quad (12)$$

Burada  $-1$ ;  $S_p$ 'nin inversini gösterir.  $S_p$ 'nin inversinin mevcudiyeti için;  $(n_A + n_B - 2) > p$  olmak zorundadır. Kalibrasyon verisindeki kümelendirmeyi gözle incelemek için discriminant fonksiyonları kullanılabilir. Buna ilaveten, bir discriminant sınıflandırıcısı hesaplanabilir ve bilinmeyen bir örneği, mevcut iki sınıftan birine atamak için kullanılabilir. Discriminant sınıflandırıcısı şöyle hesaplanabilir:

$$m = 1/2(X_A^* - X_B^*)^T S_p^{-1} (X_A^* + X_B^*) \quad (13)$$

Bilinmeyen bir bileşik için  $X_0$  cevabı kullanılarak, bilinmeyen discriminant değeri şöyle hesaplanabilir:

$$y_0 = (X_A^* - X_B^*)^T S_p^{-1} X_0 \quad (14)$$

Eğer  $y_0 \geq m$  ise  $X_0$ , A sınıfına,  $y_0 < m$  ise  $X_0$ , B sınıfına atanır.

LDA; iki sınıftan fazlasının mevcut olduğu durumlarda da kullanılabilir. Popülasyonu  $k$  olan bir durum için  $X_i^*$  ve  $S_i$  hesaplanır. Burada  $i=1$ 'den  $9$  ve  $10$  eşitliklerinin bağlı olduğu  $k$ 'ya kadardır. Bununla şu hesaplanır:

$$X^* = \sum_{i=1}^k n_i X_i^* / \sum_{i=1}^k n_i \quad (15)$$

$X^*$ , tüm kalibrasyon numuneleri üzerinde alınan ortalamadır. Birleştirilmiş kovaryans matrisi:

$$S_p = \sum_{i=1}^k [(n_i - 1)S_i] / [\sum_{i=1}^k n_i] - k \quad (16)$$

Kalibrasyon verisi için grup varyasyonundaki bir ölçüm şöyle elde edilebilir:

$$W = [(\sum_{i=1}^k n_i) - k]S_p \quad (17)$$

İlave olarak, kalibrasyon verisi için gruplar arasındaki varyasyon ölçümü şöyle hesaplanabilir:

$$B = \sum_{i=1}^k (X_i^* - X^*)(X_i^* - X^*)^T \quad (18)$$

Discriminant fonksiyonlarının katsayıları;  $W^{-1}B$  öz vektörlerinin hesaplanması ve  $e_j^T S_p e_j = 1$  ölçeklendirilmesi ile elde edilebilir. Burada  $e_j$ ;  $j$ . öz vektörü gösterir [48]. Bu, şu demektir;  $j$ . discriminant fonksiyonu için, kalibrasyon numunesinin discriminant değerleri olan  $X_i$ ;  $e_j^T X_i$  olarak hesaplanabilir.

$p > (k - 1)$  olarak kabul edelim. Sonra  $k - 1$ ; discriminant fonksiyonları olsun. Bu discriminant fonksiyonlarından tipik olarak sadece  $s$  kullanılır. Burada  $s \leq (k - 1)$ 'dir.  $s \leq 3$  için discriminant fonksiyonları, kalibrasyon verisinin grafik yerleşimi için kullanılabilir. Sınıflandırma kuralı, mevcut  $k$  sınıflarından birine bilinmeyen bir numuneyi atamak için türetilebilir. Eğer tüm  $l \neq i$  ise,  $X_0$ 'ı  $l$  popülasyonuna ata.

$$\sum_{j=1}^s [e_j^T (X_0 - X_i^*)]^2 \leq \sum_{j=1}^s [e_j^T (X_0 - X_l^*)]^2 \quad (19)$$

Bu;  $s$  discriminant fonksiyonlarıyla tanımlı uzayda bilinmeyen numuneye en yakın olan sınıf merkezini bul ve bilinmeyen numuneyi bu sınıfa ata anlamına gelir.

Dizi cevabı ve numune sınıfı arasındaki ilişkinin lineer olduğuna inanılan durumlar, LDA'nın en iyi kullanıldığı hallerdir. Discriminant fonksiyonları, geliştirme örneklerinin kovaryans matrisleri ve ortaları üzerine kurulduğu için şu çok önemlidir. Eğitim numuneleri;



karşılaşılabilecek numune popülasyonunun temsilcisi olmalıdır. Bu yüzden; eğitime verisinde her bir sınıftan büyük miktarda numuneye sahip olmak yararlıdır. Tüm denetlenmiş metotlar gibi, eğer sadece discriminant üretimde kullanılan numuneler düşünülürse, LDA optimistik sonuçlar sağlayabilir. Bundan kaçınmak için yapılacak en iyi iş, model doğruluğunu belirlemek için dış tahmin seti (discriminant üretiminde kullanılmayan numuneler) kullanmaktır.

Maricou ve yardımcıları, 12 MOS sensörü ile tertip edilmiş [50] 12 test analitli büyük bir seti analiz etti. Numuneleri; fonksiyonel olarak benzer bileşiklerin üç grubunda kümelendirmek için, iki discriminant fonksiyonu kullanıldı. 63 numunenin tamamı discriminantları üretmek için kullanılmışken, sadece biri yanlış sınıflandırıldı. Tahmin doğruluğunu ispatlamak için, discriminant üretiminde 34 numune kullanıldı. 29'unun eğitime ihtiyacı karşılanamadı. 29 dış tahmin seti numunesi için % 93 sınıflandırma oranı doğruluğu elde edildi. Bu, modelin genelleştirme yeteneğini de göstermektedir.

Yedi piezoelektrik kuartz kristali tertibi, dört yanıcı maddeden üretilen dumanı izlemek için kullanıldı. Bu maddeler; kağıt, odun, plastik ve mısır taneleridir [51]. Sensör cevapları üzerine kurulmuş numuneleri sınıflandırmada discriminant üretimi için LDA kullanıldı. Discriminantları hesaplamak için kullanılan 52 örnekten sadece biri yanlış sınıflandırıldı. Bu çalışma, hesaplanmış discriminantların tahmin yeteneğini doğrulamak için, dış tahmin seti numunelerini kullanan iyi bir örnektir. Dış tahmin setindeki 12 numunedan sadece biri yanlış sınıflandırıldı.

Bir koku analizi uygulamasında, 24'lü iletici polimer sensörlü bir tertip, üç keskin koku numunesi ile bir kontrol numunesini analiz etmek için kullanıldı [52]. Bununla beraber, tüm numuneler discriminant fonksiyonlarını üretimde kullanıldı. Bu yüzden doğru tahmin yeteneği aşırı tahmin edilmiş olabilir.

Sensör tertipleri domuz kokması için LDA ile birlikte kullanıldı. 5 MOS sensör dizisi, erkek domuzların şişman numuneleri için cevapları ölçmede kullanıldı. Numuneleri iki sınıftan birinde sınıflandırmak için LDA kullanıldı. Dış tahmin setindeki 19 numunenin 17'si geliştirilmiş discriminantlarla doğru olarak sınıflandırıldı. Anner-Frempong ve yardımcıları; andrastenon ve skatole konsantrasyonuna göre örnekleri üç sınıfta gruplandırmak için LDA kullandı [53]. Kalibrasyon numune seti kullanılarak tahmin seti numunelerinin % 84'ünün doğru olarak sınıflandırıldığı discriminant fonksiyonları geliştirildi. Temel karışıklık, normal numunelerin anormal olarak sınıflandırılması idi. Domuz kokusunun sınıflandırılmasına ait iki uygulama, modeli geçerli kılmak için dış tahmin setinin verimli kullanımını gösterir.

## 10.2. Baş Bileşen Analizi (Principal Component Analysis) (PCA) ve Uygulamaları

Sensör dizileri, çok kısa bir zamanda büyük miktarda veri üretmek için kullanılabilir. Numune ve sensör sayısı üçten büyük olduğundan, işlenmemiş verinin grafik analizi genellikle mümkün değildir. Bu yüzden, veriyi grafik olarak yerleştirebilecek boyutlara indirgeyen metotlar sıkça kullanılır. İndirgenmiş boyutlardaki sensör tertip verisinin görüntü incelemesi, numune ve sensörler hakkında faydalı bilgi sağlar.

PCA, veri setinin boyutluluğunu indirgemek için etkili bir yaklaşım sağlar. Genelde iki veya üç ana komponent, grafik çıkış için uygun olan verinin yeterli temsilini sağlar. Tipik olarak kalibrasyon verisi; satırları  $m$  numuneleri, kolonları  $n$  sensörleri olan bir  $X$  matrisinde toplanır. PCA' dan önce  $X$  veri matrisi, ön işleme tabi tutulur.  $X$ ' deki örnek sayısı sensör sayısından daha büyüktür. Bu,  $m > n$  demektir.

PCA' da ki önemli varsayım şudur:  $X$ ' deki işlenmemiş veri, öz değerlerle ilişkili öz vektörlere bölünebilir [54].  $X$ ' i parçalamak için kullanılan metotlardan en yaygın olanı, tekil değerli bölünme (singular value decomposition) (SVD)' dir [54,55]. Bu metot sıkça tercih edilir. Çünkü SVD, en geniş limitli durumlarda en kararlı alternatifi sağlar.  $X$  üzerinde SVD uygulanarak;

$$X = U \sum V^T = (u_1, u_2, \dots, u_n) \begin{bmatrix} \sigma_{11} & 0 & \dots & 0 \\ 0 & \sigma_{22} & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \sigma_{nn} \end{bmatrix} (v_1, v_2, \dots, v_n)^T \quad (20)$$

elde edilir. Burada  $U$ ;  $m \times n$  ortonormal matrisi,  $\sum$  ; diagonal ve sıfır kapalı diagonal üzerinde tek değerli ( $\sigma_{ii}$ )  $n \times n$  matrisi,  $V$ ;  $n \times n$  ortonormal matrisidir.  $\sigma_{ii}$  değerleri negatif olmayan sayılardır ve  $\sigma_{11} \geq \sigma_{22} \geq \dots \geq \sigma_{nn}$  şeklinde listelenir. Tekil değerlerin kareleri,  $X$ ' in öz değerlerine eşittir. Yani;  $\sigma_{ii}^2 = \lambda_i$ ' dir. Burada  $\lambda_i$ ;  $X^T X$  ve  $XX^T$ ' nin  $i$ . öz değerlerini gösterir.  $U$ ' nun  $u_1, \dots, u_n$  kolonları,  $(\sigma_{11})^2 \dots (\sigma_{nn})^2$  öz değerleriyle ilişkili  $XX^T$  öz vektörlerini gösterir.  $V$ ' nin  $v_1, \dots, v_n$  kolonları,  $(\sigma_{11})^2 \dots (\sigma_{nn})^2$  öz değerleriyle ilişkili  $X^T X$  öz vektörlerini temsil eder.  $U$  ve  $V$ ' yi şekillendiren öz vektörlere ana komponentler adı verilir. Sıklıkla; SVD kullanılmazsa,  $X$  bölünmesi  $U$  ve  $H$  matrislerini üretecektir. Bu durumda  $H$ ' nin  $h_1 \dots h_n$  kolonları;  $\sigma_{11} \dots \sigma_{nn}$  tekil değerleriyle ölçeklenen  $V$ ' nin  $v_1 \dots v_n$  kolonlarına eşittir. Yani  $h_i = \omega_{ii} v_i$ ' dir. Pratikte  $U$  ve  $H$ ' nin kolonları önemlidir.

Her bir öz vektörle tanımlanan varyans miktarı, öz değerlerin büyüklüğü ile belirlenir.

$$\text{var}_i = \sigma_{ii}^2 / \sum_{k=1}^n \sigma_{kk}^2 \quad (21)$$

$\text{var}_i$ ;  $i$ . öz değere katkıda bulunan varyans yüzdesini belirtir. Öz değerler azalan bir mertebede çeşitlendiği için, veri varyansının en büyük miktarı ilk PC tarafından tarif edilecektir. İkinci PC, ilk PC'ye ortogonal bir yönde veri varyansının en büyük miktarını tarif eder. Veri varyansı sadece iki boyutta iyi tanımlanabilir. Eğer sensörler nonlineer veya katkılı olmayan bir davranış gösterirse, PCA kullanarak sınıf ayrımı mümkün olmayacaktır. Bu yüzden tertipteki sensörlerin lineer davranacağı düşünülürken, PCA kullanmak en iyisidir.

$X$  bölümlenmesi bir kez yapıldığında, skorlar adı verilen  $U$ ' nun kolonları, numune kıyaslamalarının düşük boyutlu uzayına işlenmemiş veriyi yansıtmada kullanılabilirler. Mesela,  $u_1 - u_2$  eğrisi verideki en büyük miktardaki varyans için iki boyutlu hesaplamada kıyaslanabilecek  $X$ ' e sahip numunelere müsaade eder. Bu eğrilere, skor eğrileri adı verilir. Bu eğrilerin incelenmesi, ana komponent uzayında incelenmiş mevcut numune kümelerinin herhangi birini açığa çıkartır. PC' nin daha büyük sayıları boyutluluk azaltılmasını içerebilir. Fakat faydalı grafik çıkışı sadece 3 PC' ye kadar mevcut olmaktadır.

$H$  kolonlarının izafi olarak diğerine planlanmasıyla elde edilen yüklenmiş planlar, sensörleri değerlendirmek için kullanılabilirler. Her bir PC, hangi sensörün önemli şekilde bu PC' ye katkıda bulunacağını belirlemek için incelenmelidir. Yüksek bir yükleme değeri şunu gösterir: PC; orijinal sensör cevabına yakın bir yönde sıralanır. Veri varyansına yüksek oranda katkıda bulunan önemli sensörlerin seçimi yükleme planlarının incelenmesiyle gerçekleştirilebilir. İlave olarak, sensörlerin bir diğerine olan bağı keşfedilebilir. Yükleme planları; hangi sensörün kesin bilgi sağlayacağını ve hangi sensörlerin benzer bilgiyi sağlayacağını belirlemek için kullanılabilir.

PCA, bağımsız değişkenler arasında yüksek lineer bağımlı olarak genellikle veriye uygulanır. PC planları incelenirken dikkatli olunmalıdır. PC' ler veri varyasyonunu tanımlar ve sınıf ayrımı sağlamayabilirler. Sensör cevabının numune konsantrasyonuna rölaf olarak lineer davranış gösterdiği durumlarda yapılacak en iyi iş PCA' yı uygulamaktır. PCA' nın popülerliği, gerçekleştirilebilme kolaylığına bağlıdır. Eğer kümeler mevcutsa, iki veya üç boyutlu planlar belirleme amacıyla incelenebilir. Bununla beraber tekniğin grafik tabiatı, yorumlanan sonuçlarda bazı subjektiflik dereceleri ortaya koyar. Buna ilave olarak,  $X$ ' deki veriye PCA basitçe uygulanır ve eğer herhangi bir kümelendirme mevcutsa bunu görmek için PC planları

üretir. Eğer onlar uygun kümelerde başarısız olurlarsa,  $X$ ' in parçası olmayan numuneler bunu görmek için PC planları üzerine yansıtabilirler. Fakat bu doğrulama genellikle yapılmaz.

Sensör dizi performansını değerlendirmek için, test analiti olarak sıklıkla uçucu organik bileşikler seçilir. Freund ve Lewis, sekiz test analitini analiz etmek için 14' lü iletici polimer sensörü kullandı [56]. İlk önce tek sensörlerin cevapları incelendi. Fakat hiçbir tek sensör sekiz analitin tümünü ayırt edemedi. Bununla beraber, PCA' nın 14sensörün verisine uygulanması, çeşitli üç boyutlu planlardaki ilk dört PC' nin bilgisini kullanarak, tüm analitlerin ayrılması sonucunu verdi. Benzeri bir deney, 17 siyah karbon- polimer dirençli bir sensör dizisiyle yapıldı [4]. Bu çalışmada 9 organik test analiti kullanıldı. İlk üç PC kullanıldığında bazı üst üste çakışmalar mevcutken, tüm ayrılma 3, 4 ve 5 PC kullanılarak başarılabildi. Aynı tertip, metanol ve etanolün ikili karışımlarının cevaplarını incelemek için de kullanıldı. Verilen bir numune akış oranı için lineer bir davranış ilk iki PC' nin planında görüldü. Bu sırada numuneler saf metanolden saf etanol kompozisyonunda değişmekteydi. Sensörlerin lineer davranışı; PCA' nın, analiz metodu olarak uygun bir seçim olduğunu gösterir. Sadece karışım komponentleri PCA kullanarak tanımlanamaz. Aynı zamanda ikili karışımda her bir komponentin mol kesri de belirlenebilir.

Sensör dizileri genellikle meşrubatları izlemek için kullanılır. Mesela karıştırılmış içeceklerin üst boşluk (head space) numunesi, gerçek farklılaşmaya ve kusurlu yığılmalara izin verir [57]. Nanto ve yardımcıları, dört kuartz rezonatörün tertibiyle elde edilen cevaplar üzerine kurulmuş meşrubat ayrımlarında başarılıydı [58]. Hedeflenen problem, farklı tipli alkollerin numuneleri ve değişen etanol miktarlarına sahip numuneleri içerir. Her iki durumda tüm ayırım iki PC ile başarılmıştır. Özel olmayan sensör tertibi bira koku analizinde kullanılırken, sınıflar içindeki varyasyon büyük oranda sınıf varyasyonu ve PCA arasında rölatiftir. Bunlar kokuları ayırmada başarısızdırlar [52]. Di Natale ve yardımcıları şarap numunesinin cevaplarını kaydetmek için 5 MOS sensöründen oluşan bir dizi kullandı [59]. Beş bağbozumu şarabı üç konsantrasyonun her birinde incelendi. PCA, bağbozumlu numuneleri ayırt edilememiştir. Bir sensörün cevapları kaldırılırken ve kalan dört sensörün verisi üzerinde PCA çalışırken, daha iyi bir şarap ayırımı gerçekleştirildi. Bu şunu gösterir: Veri varyansına katkısı olan bilgiyle PCA ters olarak etkilenebilir. Fakat, numuneleri ayırmada faydalı bilgiye katkı sağlamaz.

PCA, bir sınıflandırma aracı olmasına rağmen bazen, sayısal bilginin bazı derecelerini elde etmek için de kullanılabilir. Hong ve yardımcıları üç konsantrasyondaki her dört gazın cevaplarını elde etmek için dört gaz sensöründen oluşan bir dizi kullandı [60]. PCA şunu açığa çıkardı: Analitler; analit benzerliği ve konsantrasyona göre kümelendiler. Daha düşük

konsantrasyonlarda numunelerin çakışması mevcuttu. İlave olarak, eğer konsantrasyon üç parçalı değerden ziyade sürekli limiti kapsıyorsa, dört gaz tipinin ayırımı için sensör tertibinin faydalılığından söz edilemez. Aynı dizi altı koku numunesinin (havuç, yeşil soğan, bayan parfümü, erkek parfümü, % 25'lik likör, % 40'lık likör) cevaplarını kaydetmek için kullanıldı. Altı koku numunesi, sadece iki PC üzerine kurulmuş hafif bir çakışmaya sahipti. Her iki örnek; konsantrasyon farkı üzerine kurulmuş aynı analiti sınıflandırmak için bazı yetenekler gösterdi. Bununla beraber, numune cevaplarının normalizasyonu cevaptan herhangi konsantrasyon etkisini kaldırabilir, konsantrasyon önemsenmeden analit tipine göre numunelerin gruplanmasına izin verir.

PCA; zamana göre toplanmış veriyi analiz etmek için de kullanılabilir. Mesela, Auge ve yardımcıları dört test analitinin verisini toplamak için 4 kuartz rezonatörlü bir dizi kullandı [61]. Analitler; başarılı zaman artışlarında veri toplanmasıyla, sıralı bir dizi sundu. Analitler kümelenebilseydi, üretilen veri matrisi bunu belirlemek için PCA ile analiz edilecekti. Sensör sinyallerinin artışı ve düşüşü ile taşıyıcı gazı temsil eden dağılma noktaları boyunca dört küme gözlemlendi. Buna ilaveten Llobet ve yardımcıları, sabit durum ve geçici ölçümler üzerine kurularak elde edilen kümelendirmeyi karşılaştırdı [62]. Dört Taguchi gaz sensörünün tertibi, değişen konsantrasyonlarda üç organik uçucuyu analiz etmek için kullanıldı. Biri sabit durumda, biri geçici bilgiyi yakalamada olmak üzere iki cevap değeri, her bir numune temsilinde her bir sensör için kaydedildi. Bağımsız PC analizleri, sabit durum ve geçici veri üzerinde kullanıldı. Geçici bilgi kullanılarak daha üst kümelendirme gerçekleştirildi. Geçici ölçümlerin üretilebilirlikleri sabit durum ölçümlerinden daha fazladır.

Nakamura ve yardımcıları, altı QCM sensör cevaplı bir diziden geçici bilgiyi çıkartmak için bir metot geliştirdi [63,64]. Onlar; sensör cevaplarını oto-regressive bir modelle tanımladı, cevabın zaman sabitlerini buradan çıkardı. Zaman sabitleri PCA ile birlikte kullanıldı ve doymuşluk kütlelerinin statik ölçümünden daha iyi test analit ayırımı sağladı [64]. Gaz konsantrasyonunun değiştiği durumlara yetebilmek için metot iyileştirildi. Değişik kaynakların dumanı, duman üzerine kurulu yangın kaynağını belirleme amacıyla incelendi. PCA, statik cevaplar üzerinde gerçekleştirilen yanan PVC kabloları ve devre bordlarının numuneleri önemli ölçüde çakıştı. Zaman sabitlerinin birleşmesi, PVC kabloların ve devre bordlarının farklılaştırılabilmesine müsaade etti.

### 10.3. Baş Bileşen Regresyonu (Principal Component Regression) (PCR) ve Uygulamaları

PCR;  $X$  cevap bilgi matrisiyle,  $C$  konsantrasyon bilgi matrisi arasında bir bağ (geçiş) sağlar.  $C$  matrisi  $m \times s$  şeklinde olup analit sayısını gösterir.  $P$ ;  $n \times s$  şeklinde regresyon katsayıları matrisi iken, matrisler  $C = XP$  şeklinde ilişkilendirilir [54,55].

$C$  ve  $X$  'deki kalibrasyon bilgisi, regresyon katsayılarına ait bir yaklaşımı bulmak için kullanılabilir. Gerçek regresyon katsayıları, deneysel hatanın kaçınılmaz mevcudiyeti yüzünden tam olarak hesaplanamazlar. Regresyon katsayılarını hesaplamada ilk adım, (20) eşitliğine göre  $X$  bölümlenmesidir.  $P$  'ye yaklaşım  $\hat{P}$  ile gösterilir ve şöyle hesaplanabilir:

$$\hat{P} = V \sum^{-1} U^T C = (v_1, v_2, \dots, v_n) \begin{bmatrix} \sigma_{11} & 0 & \dots & 0 \\ 0 & \sigma_{22} & \dots & 0 \\ \cdot & \cdot & \dots & \cdot \\ 0 & 0 & \dots & \sigma_{nn} \end{bmatrix}^{-1} (u_1, u_2, \dots, u_n)^T C$$

$$\hat{P} = \left[ \sum_{i=1}^n v_i u_i^T / \sigma_{ii} \right] C \quad (22)$$

(22) eşitliğindeki hesaplama,  $n$  tane kullanılabilir ana komponentin tümünü çalıştırır. Bununla beraber, mümkün  $n$  PC' nin tek  $k$  alt seti, hesaplama için kullanılabilir.  $k$  PC genellikle, veri varyansının en büyük miktarının hesaplanmasıdır. Bu da ilk birkaç PC demektir. Eğer çok fazla PC kullanılırsa, şekillenen model sadece verinin genel karakteristiklerini içermekle kalmayacak, aynı zamanda genelleştirme için gerekli olmayan kalibrasyon verisinin detaylarını da tanımlayacaktır. Bu, kalibrasyonda kullanılmayan verinin tahminlerini de azaltabilecektir. Sadece  $k$  PC kullanılarak (22) eşitliği yeniden yazılabilir.

$$\hat{P} = V_k \sum_k^{-1} U_k^T C = (v_1, v_2, \dots, v_k) \begin{bmatrix} \sigma_{11} & 0 & \dots & 0 \\ 0 & \sigma_{22} & \dots & 0 \\ \cdot & \cdot & \dots & \cdot \\ 0 & 0 & \dots & \sigma_{kk} \end{bmatrix}^{-1} (u_1, u_2, \dots, u_k)^T C$$

$$\hat{P} = \left[ \sum_{i=1}^k v_i u_i^T / \sigma_{ii} \right] C \quad (23)$$

Burada  $k$ ; mümkün  $n$  PC' nin  $k$  alt setinin kullanımını gösterir.

$\hat{P}_k$  hesaplandıktan sonra tahmin edilen konsantrasyonlar, verilen  $n$  sensörlü bilinmeyen cevabın, bilinmeyen numuneleri için elde edilebilir. Eğer  $y$ , bilinmeyen numunenin  $1 \times n$

cevabını gösterirse, analit konsantrasyonlarının tahmini  $\hat{c}_y = y \hat{P}_k$  olarak elde edilebilir.

Burada  $\hat{c}_y$  ; s analiti için tahmin edilen konsantrasyonların 1 x s vektörünü gösterir. Bu yüzden, kalibrasyon numunelerindeki bilinen miktarların mevcut tüm özellikleri için sayısal tahminleri elde etmede, PCR kullanılabilir.

PCR lineer bir teknik olduğundan, sensör cevaplarının lineer olduğu bilindiğinde bu, çok başarılı olacaktır. Eğer tahmin doğruluğu, sadece ana komponentleri üretmek için kullanılan veri üzerine tesis edilirse PCR, LDA' ya benzer olarak optimistik sonuçları üretebilir. Bu problem, eğitime verisiyle birlikte birini dışarıda bırakma çapraz geçerliliği prosedürünün kullanımıyla biraz yatıştırılabilir. Bununla beraber, modelin tahmin yeteneğini saptamanın daha iyi bir yolu, dış tahmin seti numuneleri ile birlikte kullanılmasıdır.

Eklöv ve yardımcıları, sos numunelerinin fermantasyon zamanını tahmin etmek için PCR kullandı [65]. Fermantasyon prosesinin çeşitli zamanlarında 14' lü sensör dizisiyle veri toplandı. Veri alt seti; modeli geçerli kılmak için kalan kısım la beraber, kalibrasyon modelini yapmak için kullanıldı. İki ana komponenti kullanan PCR modeli, fermantasyon zamanlarını tahmin etmede herhangi tek sensörlü lineer regresyondan daha başarılıydı. Zaman üzerine, sos karakteristiklerindeki sensör cevapları ve değişimleri benzer şekilde nonlineerdir. Bu yüzden, bu durumdaki optimal tahminleri PCR sağlayamayacaktır.

Balık numunelerinin birikme zamanlarının karakterize edilmesinde Schweizer-Berberich ve yardımcıları, koku üretiminin farklı bileşiklerini belirlemek için PCR kullandı [66]. Analizler gün periyodu üzerine, kokunun nasıl değiştiği izlenerek ve özel analitlerin konsantrasyonunun kokuya göre nasıl değiştiği takip edilerek gerçekleştirildi. Test gazı karışımları için sekizli amperometrik sensör dizisinin cevapları kullanılarak, makul doğrulukta bileşikleri içeren sülfür- nitrojen konsantrasyonunu tahmin etmek için modelleri inşa edebildiler. Bu onlara, balık numunelerinin aynı bileşik konsantrasyonlarının tahmin edilmesi için izin verdi. Bununla beraber diğer önemli bileşiklerin tahmin doğruluğu zayıftı. Zayıf tahminler PCR' ın başarısızlığından ziyade, önemli analitlerin sensör hassasiyeti yoksunluğu olarak görülürler.

#### 10.4. Parçalı En Küçük Kareler (Partial Least Squares) (PLS) ve Uygulamaları

PLS; PCR ve PCA' ya kadar olan metotlara benzer bilgi sağlar.  $X$  kalibrasyon matrisi  $B$  ve  $D$  matrislerine bölünür.  $b_1, b_2, \dots, b_k$  vektörleri,  $u_1, u_2, \dots, u_k$  vektörlerinin lineer

kombinasyonudur.  $d_1, d_2, \dots, d_k$  vektörleri,  $v_1, v_2, \dots, v_k$  vektörlerinin lineer kombinasyonudur. PLS;  $U$  ve  $V$  üretiminde PCR'ın yaptığından ziyade  $B$  ve  $D$  üretiminde kalibrasyon konsantrasyon değerlerinin  $C$  matrisinden daha çok bilgi kullanır.

$B$ 'nin kolonları genel olarak PLC skorları olarak adlandırılır ve  $U$ 'nun kolonları kullanarak üretilen PC planlarına benzer planları üretmek için kullanılabilir. PLS yükleme planları,  $D$  kolonlarından benzer bir biçimde elde edilebilir. Planların grafik analizi, daha geleneksel PCA planlarına benzeyen bilgiyi sağlar.

PLS modeli belirleme,  $P$ 'deki regresyon katsayılarının tahminini de gerektirir. İlk önce  $B$  ve  $D$ 'deki saklı değişkenler kullanılarak  $X$  kalibrasyon matrisi yansıtılır:

$$X_k = B_k B_k^T X D_k D_k^T \quad (24)$$

$P$  tahmini de şu şekilde elde edilebilir:

$$\hat{P}_k = X_k^+ C \quad (25)$$

Burada  $X_k^+$ ;  $X_k$ 'nin Moore-Pensore genelleştirilmiş tersini gösterir [67].

PLS; PCA ve PCR'a oldukça benzer olduğundan, PLS kullanıldığında tedbir alınmalıdır. PLS lineer bir tekniktir. Fakat nonlineer varyasyonları da mümkündür. Özellikle sayısal tahminleri elde etmek için PLS kullanıldığında, model doğruluğunu saptamak için dış tahmin seti kullanılmalıdır.

Sensör tertip verisinin cevapları üzerine kurulmuş önemli analitleri ayırmak ve sınıflandırmak için PLS kullanılmıştır. Carrasco ve yardımcıları şunu göstermek için PLS skor planlarını kullandı: 18 MOS gaz sensörlü tertip verisi, parfüm numunelerini ayırmak için kullanılabilir. Parfüm numunelerinin sınıfı; enjeksiyon için kullanılan numune üst boşluğunun hacminin belirlenmesiyle ayrılır [68]. Numunelerin tüm ayrımı gerçekleştirilmesine rağmen, LDA ve PCA'nın enjeksiyon hacmi kadar iyi olan parfüm tipi üzerine kurulmuş numuneleri ayırabildiği görüldü. Analiz sonuçları PLS kullanılarak sunulmadı. Niebling ve Müller iki sensör tertibiyle oluşturulmuş modelleri, sekiz sensörle oluşturulanlarla karşılaştırdı [69]. Sekizli sensör verisiyle oluşturulan iki faktörlü PLS modeli daha üstündü. Bu çalışma, gerekli bilgiyi ve önemsenmeyen gereksiz bilgiyi çıkartmada, PLS'nin kabiliyetini açıkça gösterir.

Grate ve yardımcıları, kalibrasyon fazında temsil edilmeyen, bilinmeyen buharları tanımlamak için, polimer kaplı akustik dalga buhar sensörlerinin dizisini kullandı [70]. Bunu başarmak için; dizideki önceden belirlenmiş polimer parametrelerinin (polimer kalınlığı gibi) ve



sensör cevabının üzerine kurulmuş, verilen buhar çözme parametrelerini tahmin etmek için modeller üretildi. Hesaplanan çözüm parametreleri daha sonra, pek çok buharın parametrelerini içeren bir veri tabanı ile karşılaştırıldı. Bilinmeyen buhar, kalibrasyon fazında tam olarak temsil edilmediğinde bile, bilinmeyen buharı tanımlamak için potansiyel araçlar sağlayan tek çalışma budur.

### 10.5. Küme Analizleri

Kümelendirme; hangi objenin diğerine en benzer olduğunu belirleyen ve buna göre onları gruplandıran bir çalışmadır. Benzerlik pek çok yöntemle belirlenebilir. En yaygınlarından biri N boyutlu uzayda mesafelerin karşılaştırılmasıdır. N; analit sensör kümelendirmesinin arzulanmasına bağlı olarak sensör cevap sayısını veya numune sayısını temsil eder. Fakat biz burada N'yi, sensör cevaplarının ve kümelendirilecek analitlerin sayısına eşit olarak tanımladık. Her bir numune sensör cevaplarının N boyutlu vektörü ile temsil edilebilir. Küme analizinde kullanmak için pek çok mesafe metriği rapor edilmiştir [71]. Fakat en yaygın kullanılanı öklid mesafesidir. Mesafe için genel formül;

$$d_{ij} = \sqrt[p]{\sum_{k=1}^N (x_{ik} - x_{jk})^p} \quad (26)$$

şeklinde. Burada  $d_{ij}$  ; iki numune arasındaki mesafedir.  $x_{ik}$  birinci numune,  $x_{jk}$  ikinci numune ve  $p$  arzulan mesafe metriğini üreten sayıdır.  $p=2$  için mesafe, öklid mesafesidir.  $p$  ' nin daha yüksek değerleri için daha küçük mesafelere daha ağırlık verilir.

Kümelendirme teknikleriyle çalışılırken akılda tutulması gereken bazı konulardan ilki; küme analizleri genellikle eğitime numunelerinin büyük sayısına ihtiyaç duyarlar. Bu gereksinimden dolayı her bir sınıfın numuneleri, ilgili N boyutlu uzayı dolduracaktır. İkinci konu ise işaretlenen farklı sonuçların kümelendirme metodunun seçimine bağlı olarak sıkça vuku bulmalarıdır.

#### 10.5.1. Haritalama ve gösterme

Sensör dizi verisi, veri seti hakkındaki kullanıcı kalite bilgisini vermek için, görüntülü teftişle grafik olarak gösterilebilir. Veri setleri üç boyuttan daha büyük iken, bu direkt olarak yapılamayabilir ( şu anlama gelir: numune başına üç sensör cevabından daha fazlası vardır). Grafik gösterim için veri, iki yada üç boyuta indirgenebilir. Veri analizinde bu amaç için en yaygın olarak kaydedilmiş metot, ana komponent analizidir. Sammon haritalaması olarak

bilinen ve nonlinear bir haritalama olan diğer teknik, bazı ticari enstrümanlar sağlayan veri analizi paketinde bulundurulur [72,73,74,75].

Sammon haritalamasında amaç; veri noktaları arasındaki orijinal ilişkileri koruyan, en iyi iki boyutlu temsili bulmaktır.  $N$  boyutlarındaki iç nokta mesafeleri bir mesafe metriği, genellikle öklid mesafesi kullanılarak hesaplanır. İki boyuttaki noktalar arasındaki mesafeler şöyle verilir:

$$d_{ij}^* = [(y_{i1} - y_{j1})^2 + (y_{i2} - y_{j2})^2]^{1/2} \quad (27)$$

Burada  $y$  ' ler ana komponent analizinden hesaplanan skorlardır.  $i$  ve  $j$  noktaları, 1 ve 2 koordinatlarını gösterir.

Amaç,  $E$  hata fonksiyonunu minimize etmek için her bir noktanın  $y$  'sini iteratif olarak değiştirmektir.

$$E = [1 / \sum_{i>j} (d_{ij})^p] \sum_{i>j} (d_{ij} - d_{ij}^*)^2 / (d_{ij})^p \quad (28)$$

$p$  eksponenti, büyük ve küçük mesafelerin ağırlığına müsaade etmek için bulundurulur. Projeksiyon üzerinde  $p$  ' nin farklı değerlerinin etkisi Bender ve Kowalski [73] tarafından tartışılmıştır. Hata fonksiyonu, en keskin düşme gibi lineer minimizasyon metodu kullanılarak minimize edilmelidir.

İletici polimer sensörlü ticari bir enstrüman kullanılarak bir çalışma yapıldı. Burada amaç, enstrümanın süt ürünlerinden su buharını ve bir süt ürününden diğerine doğru değişirken yine su buharını ayırt edebilmesiydi [74]. 32 sensörden veri alındı ve her bir vektör arasındaki öklid mesafesi hesaplandı. 32 boyutlu uzayın veri noktalarını, iki yada üç boyuta yansıtmak için hayali bir denetim amacıyla Sammon haritalaması kullanıldı. Balsam ağacının sirkesinin veri setini incelemek için, ticari olarak kullanılabilir elektronik burunla birlikte Sammon haritalaması da kullanıldı [75].

Kompleks veri setlerindeki pek çok analite ait çalışmalarda bu grafikler, analit sınıflarının iyi bir ayrımını üretemezler. Bu; analitlerin diğerine ayrılamayacağını gösteren kesin bir kanıt değildir. Bazı durumlarda yeterli ayrımını vermek için çok fazla boyutun görüldüğü haller olabilir.

### 10.5.2. Hiyerarşili kümelendirme

Tüm nokta çiftleri arasındaki çok değişkenli mesafeler ilk önce hesaplanır. Kümelendirme metoduna ve benzerlik ölçüsüne bağlı kümelendirme sonuçları veri setine uygulanır.

Bazen işlenmemiş mesafelerin kullanımından ziyade  $S_{ij}$  benzerlik değerleri aşağıdaki gibi hesaplanır:

$$S_{ij} = 1 - (d_{ij} / d_{ij}(\max)) \quad (29)$$

Burada  $d_{ij}$ ;  $i$  ve  $j$  noktaları arasındaki mesafedir.  $d_{ij}(\max)$ ; herhangi iki numune arasındaki en büyük mesafedir. Bu yüzden, veri setindeki en uzak noktalar  $S_{ij} = 0$ ' a ve benzer noktalar  $S_{ij} = 1$ ' e sahip olacaktır.

Uzaklık ve benzerlik değerleri hesaplandıktan sonra, değişik metotlardan biri kullanılarak kümelendirme gerçekleştirilebilir. Tek bağlantı; kümedeki en yakın nokta ile nokta arasındaki uzaklığa göre, kümeye rölatif olan bir noktanın yakınlığını hesaplar. Tüm bağlantı metodu; kümedeki en çok nokta ile nokta arasındaki uzaklığa göre, bir nokta ve bir küme arasındaki uzaklığı belirler. Merkez metodu, bir kümedeki noktaların ağırlık merkezi ve nokta arasındaki uzaklığı kullanır. İlave olarak hesaplama, bir kümedeki nokta sayısıyla ağırlıklandırılabilir.

Bu küme metotlarının herhangi biri, noktaların hiyerarşisini yapmak için kullanılabilir. Veri setini oluşturan noktalarla başlatılarak, en benzer iki nokta birleştirilir. Daha sonra en benzer üçüncü nokta birleştirilir sonra dördüncü ve diğerleri. Veri noktalarının tüm kümelendirme derecelerini temsil eden, hiyerarşik bir düzenleme yapılır.

Bölümlü metotlarla iyi sonuçlar elde edilebilir. Bölümlü bir metotta tüm objelerin seti ilk önce iki kümeye ayrılır. Tüm objeler tamamen ayrılana kadar her bir küme başarılı bir şekilde bölümlenir. Her bir metot kesin hesaplama avantajları sunar [71].

Minimum mesafe ağacı (minimal spanning tree) (MST) diğer bir metottur. MST; veri noktaları setinin birleştirilmesiyle bulunur ki her biri bir ağaç nodunu oluşturur ve veri seti üzerindeki tüm çizgi parçalarının toplamı minimumdur. Daha sonra kümeler, yanındaki diğer parçayla karşılaştırılarak, çizgi parçasının uzunluğu esaslı üzerinde ağaçtan budanır.

Küme oluşumunu takiben bir dendogram çizilebilir. Dendogram iki boyutlu bir şekildir. Buradaki noktalar; planın ayağı boyunca tertiplenmiş eşit boşluklu noktalara atanır ve

dikey eksenler, noktaların benzerlik derecelerine bağlıdır. Dik çizgi, planın ayağı boyunca her bir noktadan başlar ve nokta kümeye kavuştuğunda herhangi özel noktanın çizgisinde son bulur. Bu; bir veri setindeki kümelendirme ilişkilerinin tüm setinin basit, ve görsel bir incelemesine müsaade eder. Bu metotla, kümelerin otomatik olarak oluşturulmadığına dikkat edilmelidir. Onlar, mesafe veya diğer bir kritere göre atanabilirler.

Hiyerarşik kümelendirme, denetlenmemiş bir metottur. Verideki ilişkilerin kalitatif (nitelik) olarak araştırılmasıyla en iyi şekilde uygulanır.

### 10.5.3. K- en yakın komşuluk sınıflandırması (KNN)

Bu metotta bir bilinmeyen; N boyutlu uzaydaki geliştirme setinde, K' nın en yakın komşuluklarının ana sınıflandırmasına göre sınıflandırılır. Sensör cevabına veya numuneye bağlı N boyutun her biri arzulanan uygulamaya bağımlıdır. Bağlar; K' nın sadece tek sayılarla sınıflandırılmasıyla veya daha yakın komşularına daha büyük ağırlık verilmesiyle önenebilir. Yakınlık, mesafe metriği kullanılarak belirlenir.

KNN; algoritmik olarak çok basittir [76]. İlk önce, tüm nokta çiftleri arasındaki mesafe matrisi hesaplanır. En yakın komşulukları belirlemek için; diyagonal olarak simetrik mesafe matrisinin alt ve üst parçası, en küçük K mesafe değerleri için taranır. Hangi sınıf için olursa olsun, bir gözlemin K en yakın komşuluğu onun sınıfına ait olarak alınır. Algoritma; birini dışarıda bırakma çapraz geçerlilik ile birlikte kullanılır. Sınıf üyeliği; en yakın komşuluğu incelemek için veri setinin kalanları kullanılarak, bir zamanda, bir üye için belirlenir. Tüm veri setinin tahmin yeteneğinin bir ölçümü olan yüzde hata; gözlemlerin toplam sayısının bölünmesiyle doğru olarak sınıflandırılmış gözlem sayısıdır. Kolaylığına rağmen KNN sınıflandırması, çapraz reaktif sensör dizi verisiyle birlikte çok verimli olarak kullanılabilir.

KNN; denetlenmiş bir metottur. Bu; bağımlı değişken bilgisinin kullanılacağı anlamına gelir. Önemli parametreler; kullanılan eğitim numunelerinin toplam sayısını, her bir sınıfta ve kullanılan eğitim metodundaki eğitim örnek sayısını içerir. Oluşturulduktan sonra KNN model performansını değerlendirmek için geçerlilik yerine getirilmelidir. Geçerlilik metotları; eğitimde kullanılmayan gözlem sınıflarını tahmin etmek için model kullanımını ve bağımlı değişkenle yeniden yapma analizlerinin mücadelesini içerirler. Mücadele deneyi için arzulanan sonuç, düşük doğrulukta sınıflandırma oranıdır ve serbest sınıf atamasıyla sürdürülmesi umulan tahminlere tercihen yakındır.

Nakamura ve yardımcıları otoregresif model katsayılarından, bir sensör cevap modeli türetti [77,64]. Plazma polimer film tabakasıyla kaplı 6 AT kesici kuartz kristalli bir dizi kullanılarak, beş organik çözücü için veri alındı. Normalize edilmiş doymuş kütle, kendisi ve zaman sabitiyle birlikte kullanılarak ana komponent analizi yapıldı. Parametrelerin ikisi de bulunduğunda, analit sınıflarının ayırımındaki gelişmeyi belirlemek için KNN kullanıldı. Verinin iki boyutlu planlarını yapmak için PCA kullanımını gösterildi daha sonra PC' lerle analit ayırımını belirlemek için KNN kullanıldı. Grafik gösterim için sadece iki yada üç PC' nin kullanılması uygunken, KNN PC' lerin herhangi sayısını barındırabilir.

#### 10.5.4. Görülebilir olarak-etkinin deneysel bölgesi (Visually-empirical region of influence) (VERI)

İki boyutlu insan kümelendirme yargılamasını, k boyutlu veriye genişleten bir yapı, kümelendirme kriteri olarak kullanıldı. Verinin arkasındaki kullanıcının hiçbir girişi gerekli değildir. Birikme ve çok sınıflı veri noktaları için tedarik yapılır.

VERI; etki bölgesi (region of influence) (ROI) kavramı üzerine kurulmuştur. Eğer diğer veri noktaları etki bölgesi içinde uzanmıyorsa, iki boyutta her bir çiftin kümelendirildiği noktaların her bir çifti etrafında bir bölge tanımlanır. Bölgenin şekli insan konulu deneylerden tecrübe olarak belirlenir. ROI' nin ölçeği ve düzenlemesi, noktaların her bir çifti için ayarlanır. Metodu k boyuta genişletmek için, üç noktanın tüm kombinasyonları düşünülür. Eğer ROI' de sadece iki nokta uzanıyorsa ve biri yapmıyorsa, iki nokta bir küme oluşturamaz. Bu yüzden, k boyutlu bir problem, iki boyutlu problem serisine indirgenir.

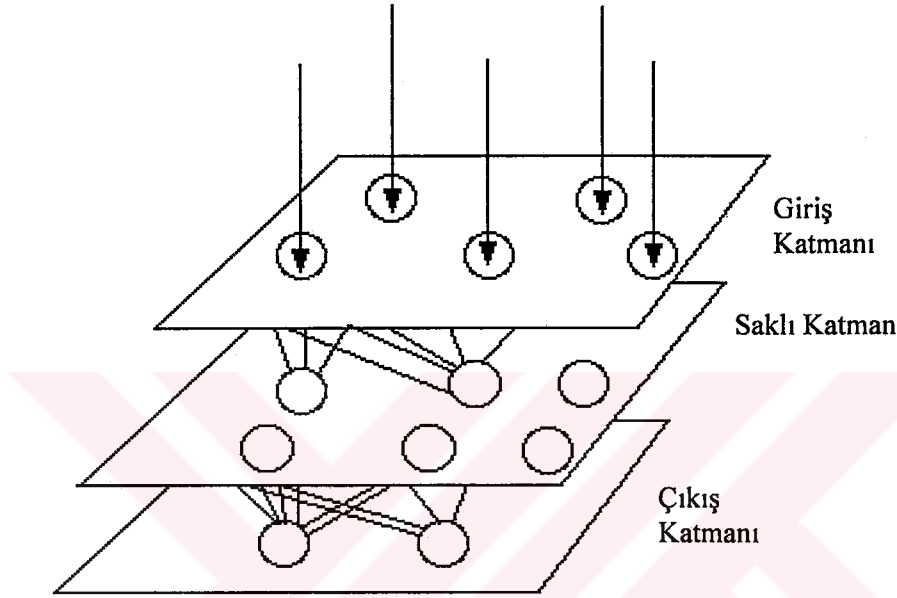
Kümelendirme, bir numune tanıma metoduna da genişletilebilir. İlk aşama, her bir özellik için değerleri normalize etmektir. Daha sonra, verinin birini dışarıda bırakma prosesi, yığıntıları ve çoklu sınıflardaki noktaları tanımlamak için çalıştırılmalıdır. Bir bilinmeyen veya test noktası sınıfı, onun kümelendirdiği eğer varsa eğitim noktalarına göre sınıflandırılır. Herhangi tanımlanmış sınıfta düşmeyen noktalar yığıntılar olarak düşünülür.

#### 10.6. Besleme Ön Sinir Ağı (Feed Forward Neural Network) (FFNN) ve Uygulamaları

Bu ağlar; çok katmanlı perceptronlar ve back-propagation sinir ağları olarak da bilinirler. Bunlar, sınıflandırıcı ve belirleyici olarak kullanılırlar.

FF ağları; şekil 10.2'de görüldüğü gibi tipik olarak bir giriş tabakası, bir saklı tabaka ve bir çıkış tabakasını içerirler. Sensör dizileri için giriş tabakasındaki nöron sayısı, genellikle

dizideki sensör sayısına veya dizideki sensörlerin optimal sayısına eşittir. Her bir sensörden alınan bilgi parçalarının sayısı sıklıkladır ama asla her zaman bir değildir. Giriş tabakasını besleyen veri de sıklıkla ölçeklenmiştir.



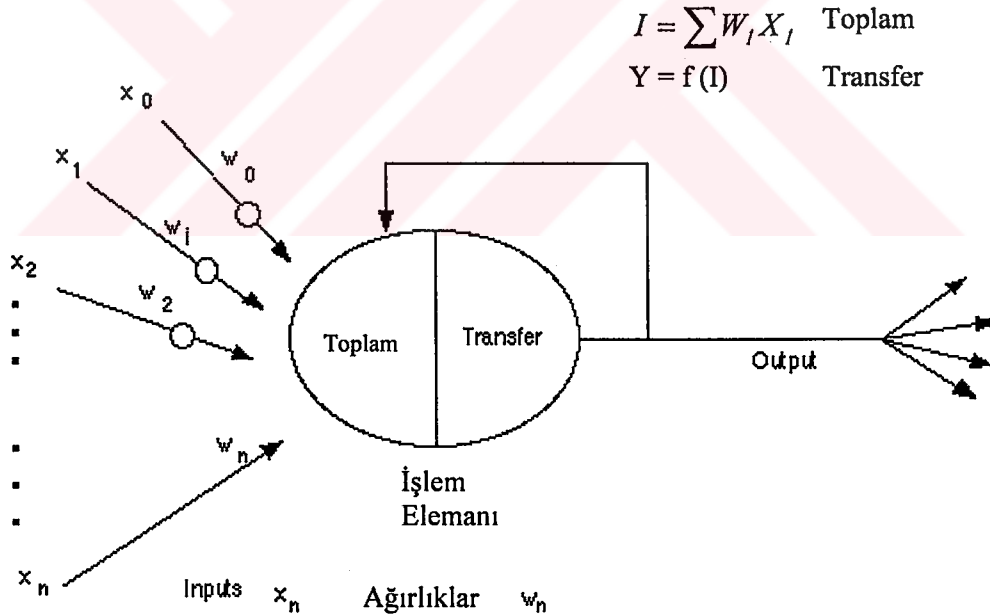
Şekil 10.2. Bir Perceptronun katman yapısı

Her bir veri, ağırlık boyunca saklı tabakaya doğru geçirilir. Saklı tabakada girişlerin ağırlıklı toplamı hesaplanır ve bir bias terimi ilave edilir. Bu, bir sabit değerli ilave girişe sahip olmaya eşittir. Bu değer; tipik olarak bir sigmoid veya hiperbolik tanjant fonksiyonu olan bir transfer fonksiyonuna bağlanır. Şekil 10.3'de bu genel işlem elemanının yapısı gösterilmiştir. Saklı tabakaya sahip olmak için nöronların optimal sayısı, deneysel olarak belirlenmek zorundadır.

Saklı tabakanın çıkışından sonra çıkış tabakasına geçilir. Çıkış tabakasındaki nöron sayısı genellikle uygulamaya bağımlı değildir. Eğer ağ bir sınıflandırıcı olarak kullanılırsa, nöron sayısı tipik olarak sınıfların sayısına eşit olur. Aynı işlemler, saklı tabaka nöronlarındaki gibi çıkış nöronlarında da gerçekleştirilir. Daha sonra sonuç ölçeklendirilir (ağın başlangıcındaki ölçekleme prosesinin tersi) ve kullanıcı cevabı alır.

Ağırlık ve biasları ayarlamak için FF ağırları geliştirilirken, bu amaçla çeşitli algoritmalar ortaya çıkarıldı. En popüler olanı; back propagation metodudur [78]. Bu tekniğin kullanımı; öğrenme oranı ve momentum teriminin optimizasyonunu gerektirir.

Seçilen algoritma önemsenmezse eğitime prosesi, tekrarlı bir prosedürdür. Gözlemler birer birer ağa sunulur ve doğru değerle ağın tahmin ettiği değer arasındaki farkı minimize eden bir yöndeki her bir gözlemden sonra, ağırlık ve biaslar ayarlanır. Eğer ağ bir sınıflandırıcı olarak kullanılıyorsa, ağın yerine sınıflandırma kullanılır. Bazı eğitime metotları, her biri tüm eğitime setini geçtikten sonra, ağırlık ve biasları her bir epoch sonrasında ayarlamayla ilgilenir. Bu süreç, eğitime tamamlanincaya kadar devam eder. Eğitime tamamlandığında belirleme metodu, ağırlık ve biasları ayarlamak için, seçilen optimizasyon algoritmasına bağlı olur. Eğitime ağında kullanılmayan gözlemleri tahmin etmek için, ağın faydalarını belirlemek amacıyla tipik olarak bir dış seti kullanılır.



Şekil 10.3. Genel işlem elemanının yapısı

NN' nin performansı, eğitmeden sonra ağırlık ve biasların başlatılmasına bağlıdır. Rasgele seçilen ağırlıkların bir seti, düşük hata oranlı bir üretebilir ve diğer bir set daha düşük sonuçlar veren bir ağ üretebilir. Başlama noktası bağımlılığını çözmek için bir yolu, değişik

rasgele başlama noktalarından pek çok ağ eğitmek ve en iyi performansı veren birini seçmektir. Diğer seçenek, ağırlık ve biasların başlatılmalarına dair makul bir set seçmek için bir optimizasyon algoritması kullanmaktır.

FFNN kullanıldığında düşünülmesi gereken diğer faktör bunun aşırı eğitilmesidir. Ağ; eğitime verisinin özelliklerini hafızaya alır ve genelleştirme yeteneğini kaybederse bu vuku bulur. Aşırı eğitmeden sakınmak için kullanılan bir metot, ağın eğitilmesini izlemek için çapraz geçerlilik (cross validation) (CV) kullanılmasıdır. Her bir epochtan sonra CV seti tahmin edilir ve eğitime; CV set hatasının artmaya başladığı geliştirme noktasında durdurulur.

Sınıflandırma amaçları için FFNN kullanılmasının bir dezavantajı şudur: Sensör sayısı veya sınıflar artarken, ayarlanabilir parametrelerin sayısı da hızla artar. Bir NN' de eğitime gözlemlerinin sayısının, ayarlanabilir parametrelerin sayısına oranı 2' nin üzerinde olmalıdır [79]. Ayarlanabilir parametre sayısı olan P şu şekilde verilir:

$$P=(I \times H)+(H \times O)+H+O \quad (30)$$

Burada I giriş tabakasındaki ünite sayısı, H saklı tabakadaki nöron sayısı ve O çıkış tabakasındaki ünite sayısıdır. Eğitime seti veri gözlemlerinin büyük sayısına sahip olmayı istemeyen birisi, sınıflandırma için muhtemelen FFNN kullanılmasını istemeyecektir.

Arzulanan eğitime veri gözlem sayısına ilave olarak, ilave numuneler çapraz geçerlilik seti için alınmalıdır. Bu, eğitime verisinin depoladığı noktadaki ağı aşırı eğitmeden korumak için kullanılır. Ağın, tahmin yeteneğine sahip olmaması için aşırı uydurma (overfitting) sonucunu verir. Çapraz geçerlilik için her bir epochtan sonra, ağın çıkışını tahmin amacıyla eğitime prosesi izlenir. Eğitime ağının hareketi boyunca çapraz geçerlilik seti hatası azalacak ve bu aşırı eğitmenin başladığı ağ noktasına kadar devam edecektir. Bu noktada çapraz geçerlilik seti hatası artmaya başlayacak ve hata artmaya başlamadan önce eğitmenin tamamlanması doğru olacaktır.

NN' nin eğitilmesini takiben, geçerliliğin bazı formları şunu göstermek için arzulanır: Model; daha çok şans eseri olarak, veri ile bağımlı değişken arasındaki doğru ilişkiden ortaya çıkar. Eğitime veya çapraz geçerlilikte kullanılmayan gözlem seti, veri setinin ihtiyacını karşılamalı ve bu amaçla kullanılmalıdır. Oluşturulan model daha sonra tahmin setinin üyelerinin bağımlı değişkenini tahmin etmek için kullanılır. Eğer tahmin seti hatası eğitime ve çapraz geçerlilik seti hatasına yakınsa, model geçerli olmuştur. Eğer tahmin seti hatası, eğitime seti hatası çok iyiye çok büyükse, ağ aşırı eğitilmiş ve eğitime seti verisini basitçe hafızaya almış demektir. Bu durumdaki diğer ihtimal şudur: Çok fazla ayarlanabilir parametre kullanıldı, yine aşırı eğitime kolayca hafızaya alındı ve ağın tahmin yeteneği yoktur.



Monte Carlo deneyleri; modeli geçerli kılmak için ilave bir metot olarak kullanılmaktadır. Burada bağımlı değişkenle mücadele edilir ve daha sonra orijinal modelde kullanılan bu bağımlı değişken ve tanımlayıcılar kullanılarak ağ geliştirilir. Eğer bu, zayıf tahmin seti hatası sonucunu verirse daha fazla delil sağlar ki model geçerlidir ve şansa bağlı olarak ortaya çıkmamıştır.

Llobet arkadaşları; sabit durum cevapları kadar iyi olan sensörlerin geçici bilgi kullanımını araştırdı. Birini dışarıda bırakma geçerliliği yapılmasına rağmen, gözlemlerin ayarlanabilir parametrelere oranı küçük olduğu için sonuçlar dikkatle yorumlanmalıdır. Bu araştırmada, FFNN analizlerindeki sabit durum cevabı kadar iyi olan dinamik verinin kullanımının değerinin bir örneği verilmiştir [62]. Dinamik verinin kullanımı; üç tip çayın ve bakteri ile maya üst boşluğunun sınıflandırılmasına ait uygulamalarda da araştırıldı [80,81].

$\text{CH}_3\text{SH}$ ,  $(\text{CH}_3)_3\text{N}$ ,  $\text{C}_2\text{H}_5\text{OH}$ ,  $\text{CO}$ ' in değişen konsantrasyonlarını sınıflandırmak için bir FFNN modeli geliştirildi. Yazarlar %100 sınıflandırmayı rapor ettiler. Bununla birlikte, sonuçlarla ilgili bazı yorumlara dikkat edilmelidir. Eğitim setinin boyutu belirtilmemiştir, tahmin setinin kullanılmasından bahsedilmemiştir ve belirlenmiş durdurma kriteri kullanılmadan ağ, 10000 iterasyon için eğitilmiştir. 148 ayarlanabilir parametreye bağlı 4:8:12 ağ mimarisi kullanılmıştır. Bu şu anlama gelir: Eğitim seti en az 296 gözlemle oluşturulmalıydı. Bu çalışmada eğitim için kullanılan numune sayısı yaklaşık olarak 60 olduğu, gözlemlerin ayarlanabilir parametrelere oranının 0.4 ve tavsiye edilen minimum 2' nin altında olduğu açıkça görülebilir[79].

### 10.7. Kendiliğinden Organize Haritalar (Self Organizing Maps) (SOM)

Kohonen ağları olarak da bilinen [82] SOM, denetlenmemiş numune tanıma tekniğidir. Eğitim; bir gözlemle başlar ve N boyutlu bir vektördür. Genellikle, dizideki sensör sayısına bağlıdır. Her  $x_i$  vektör elemanı, ağa girişi temsil eder. Çıkış üniteleri bir tertiple düzenlenir ve girişlere  $w_{ij}$  ağırlıklarıyla bağlanır. Giriş gözlemi, her bir çıkış vektörünün ağırlık vektörüyle karşılaştırılır ve en yakın ağırlık gözlemlerli çıkış vektörü galip olarak belirlenir.

Ağırlıklar öğrenme kuralı ile güncelleştirilir;

$$\Delta w_{ij} = \eta \wedge (i, i^*) (x_j - w_{ij}) \quad (31)$$

bu tüm i gözlemleri ve j elemanları için geçerlidir.  $\eta$ ; öğrenme oranıdır.  $i^*$ ; galip çıkış ünitesidir.  $\wedge (i, i^*)$ ; komşuluk fonksiyonu  $i = i^*$  için 1' dir ve gözlem çıkışlardan dolayı

artarken bu azalır.  $\eta$  ve  $\wedge(i, i^*)$ ' nın her ikisi de kullanıcının belirlediği parametrelerdir ve tipik bir şekilde eğitime ilerlemesi olarak ayarlanırlar. Bu üniteler artan önemsiz değişimlerin yükünü çekerken, galibe yakın üniteler önemli bir şekilde değişen kendi ağırlıklarına sahiptirler. Bu yöntemde galip sinirin ağırlık vektörü ve onun komşulukları, giriş vektörüne daha benzer olarak yapılırlar. Denetlenmiş numune tanıma metotları durumunda olduğu gibi sonuçların geçerliliği, eğitime ağında kullanılmayan test setlerinin sonuçlarıyla, eğitim seti ağından elde edilen sonuçların karşılaştırması ile yapılır.

Marko ve yardımcıları [83] SOM' un sensör dizisi işlemi boyunca öğrenmeye devam etmesine izin verdi. Öğrenme oranı minimumda tutuldu ve sadece galip sinirin ağırlıklarının güncelleştirilmesine izin verildi. Benzetilmiş sensör dizisi birikmesinin %20' ye ulaşmasıyla, SOM' un başarı oranı %80' in üzerinde kaldı. Bununla birlikte, eğer kümelenme boyunca SOM' un öğrenmeye devam etmesine izin verilmezse, ağ performans kaybı olmaksızın bir tecrübeye sahip olacaktır. Sensör dizi verisi kullanılarak analitleri sınıflandırmak içinde SOM kullanıldı. Altı kalay oksit sensörü kullanılarak yapılan yanıcı gazların sınıflandırılması %3' den küçük bir hata oranıyla başarıldı. Sınıflandırmayı yapmak için SOM eğitildikten sonra, son denetlenmiş adım gerçekleştirildi. Eğitime numuneleri yine ağa sunuldu ve galip sınırlar bağlı gözlemlerin sınıfına göre etiketlendi. Gözlemler arasındaki mücadelenin ortaya çıktığı durumlarda bir prosedür gerçekleştirildi. Yazarlar; veri setini, eğitime ve test seti olarak ayırmışken, her bir deneyde tahmin için kullanılan gözlem ve toplam gözlem sayıları açık bir şekilde kurulmamıştır. Birikmeli sensör sensör verisinin diğer bir simülasyonu, birikme etkilerinin kompanze edilmesinde SOM etkisini buldu [84]. Benzetilmiş birikmeli sensör tertip verisi, bazı şartlar altında operasyonlar için optimal SOM parametrelerini deneysel olarak belirlemek amacıyla kullanıldı [85].

### 10.8. Öğrenme Vektör Kuantizasyonu (Learning Vector Quantization) (LVQ)

LVQ; Kohonen tarafından önerilen denetlenmiş SOM' un bir versiyonudur. Algoritma, her bir sınıf için ağırlık vektörlerinin başlangıç oluşturması ile başlar. Bunların sayısı kullanıcı tarafından belirlenir. Ağırlık vektörleri küçük rasgele sapmaya ilave olarak, veri sentroidleri olarak başlatılabilirler. Sonra gözlem ağa sunulur ve ağırlıklı öklid uzaklığı, gözlem ve her bir başlangıç ağırlık vektörü arasında hesaplanır. En yakın gözlemin ağırlık vektörü galip sinir olarak isimlendirilir. Gözleme ikinci olarak en yakın ağırlık vektörüne ise runner-up siniri denir.

Eğer galip sinir ve gözlem aynı sınıfa aitseler, sadece bu sinir güncelleştirilir. Eğer galip sinir, gözlem ve runner-up' la aynı sınıfta değilse, her iki sinir birden güncelleştirilir. Güncelleştirmeler öğrenme fonksiyonuna göre yapılır.

$$w_{ij}^{new} = w_{ij}^{old} + \delta \alpha(t) [w_{ij}^{old} - x_j] \quad (32)$$

$w_{ij}^{new}$ ; öğrenmeden sonraki i. sinir için, ağırlık vektörünün j. elemanıdır.  $w_{ij}^{old}$ ; öğrenmeden önce i. sinir için ağırlık vektörünün j. elemanıdır.  $\alpha(t)$ ; monoton olarak azalan öğrenme oranı ve  $x_j$ ; gözlem için sunulan j. tanımlandırıcı değerdir. Eğer galip sinir, gözlem ve  $-1$  değilken aynı sınıftansa,  $\delta = +1$ ' dir. Runner-up siniri için,  $\delta = +1$  olan durumda gözlem olarak aynı sınıftan değilse,  $\delta = 0$ ' dır. Eğer sonuç doğru ise; güncelleştirmeler, siniri gözleme daha benzer yapma sonucunu verir. Eğer yanlışsa, sinir daha az benzetilir. Eğitim sürecinde nadiren seçilen veya hiç seçilmeyen sinirler, küçük rasgele sapmaya ilaveten zayıf tanımlanmış sınıfın merkezine doğru olan eğitim süreci boyunca yeniden başlatılırlar.

Eğitim verilerinin tamamı ağırlık bir kez sunulduktan sonra mevcut performans, bir maliyet fonksiyonu kullanılarak değerlendirilir [86]. Bu doğru olarak sınıflandırılmış bir gözlem yüzdesi veya ondan bazı varyasyonlardır. Bu iterasyonun maliyeti, önceki maliyetten geliştirilmişse bu kabul edilir. Eğer o zararlı ise, belli bir ihtimal üzerine kurularak kabul edilir. Eğitim; epochların belirlenen sayısı geçene kadar sürer. LVQ; aşırı eğitime yeterlidir. Aşırı eğitimden korumanın bir yolu, çapraz geçerlilik kullanılmasıdır. Tahmin fazı boyunca bilinmeyen vektör, ağırlık vektörleri ile karşılaştırılır ve bilinmeyen vektöre en benzer bir tanesi galiptir ve bilinmeyen sınıfını belirler.

20 organik buhar verisini toplamak için, polimer kaplı fiber optik sensör dizisi kullanıldı [87]. Konsantrasyon etkilerini kaldırmak için, veri otomatik olarak ölçeklendirildi ve tanımlandırıcılar hesaplandı. Her bir analit ve LVQ için, iki bireysel FF ağı kullanılarak sınıflandırma gerçekleştirildi. LVQ modelleri için özellik seçimi, genetik algoritma kullanılarak yapıldı [88,89]. Aşırı eğitimin vuku bulmasından önce, bitme noktasını belirlemek ve ağ eğitimini izlemek amacıyla çapraz geçerlilik kullanıldı. Bireysel FF ağları kolektif olarak tahmin seti gözlemlerinin sadece %73' ünü doğru olarak sınıflandırırken, LVQ' nun daha üst performansı, dış tahmin setindeki gözlemlerin %90' ının doğru olarak sınıflandırılmasıyla doğrulanmış oldu. LVQ, FF ağları ve fuzzy uyumlu rezonans teori haritasının (fuzzy adaptive resonance theory map) (ARTMAP) performansları, kalay oksit sensör tertibi verisi kullanılarak muz olgunluğunu belirlemek amacıyla karşılaştırıldı [90]. Fuzzy ARTMAP ve FF ağları, %90

ve %83 doğru sınıflandırma oranlarına sahipken, LVQ muz numunelerinin %92' sini yedi hamlık aşamasını içeren sınıflarda doğru olarak sınıflandırdı.

### 10.9. Uyumlu Rezonans Teorisi (Adaptive Resonance Theory) (ART)

FFNN' nin zıddına, ART ağları az eğitime verisine gereksinim duyar. Çünkü ayarlanabilir parametrelerin sayısı hızla artmaz. ART ağları, LVQ ağlarından daha hızlı eğitilirler. Sadece bir miktar özellik seti kullanarak eğitime istendiğinde, eğitime zaman farkı makul karşılanabilir. Özellik setlerini karşılaştırmak için pek çok eğitilmeyi gerektiren daha yoğun özellik seçimi arzulanırsa, LVQ için hesaplama zamanı ART ağlarından önemli ölçüde yüksek olacaktır.

Yeni veri noktasını yerleştirmek amacıyla yığılmalı veri sunulduğunda, LVQ gibi diğer algoritmalar onların ağırlık vektörlerini ayarlayacaktır. Bu; mevcut sınıfları tanımlamak için ağ yeteneğinin düşmesi sonucunu verebilir. LVQ analogisinde, ART ağları galip sinirin yeni gözlemine karşılaştırır. Eğer yeterli derecede benzer değilse, ağ mimarisi adapte edilebilir. ART1, ART2 [91], fuzzy ART, ARTmap ve fuzzy ARTmap [92] ihtiva eden pek çok varyant mevcuttur. Diğer NN metotları gibi bu NN' ler de çapraz geçerlilik, aşırı eğitilmeyi koruma tahmini (denetlenmiş metotlarda) ve model geçerliliğine ihtiyaç duyarlar.

Shukla ve yardımcıları, sensör dizi verisini sınıflandırmak için denetlenmiş ART ağının faydasını araştırdı [93]. Kalay oksit sensörleriyle güçlendirilmiş dört dirençteki değişim; dizi, test buharlarının değişen konsantrasyonlarına maruz bırakılırken ölçüldü. Aseton, etil metil keton, karbon tetraklorid ve oksijen; test buharlarıydı. Sensör çıkışlarının üçünün, dördüncü sensör çıkışına oranları, ART ağının girişi olarak kullanıldı. ART ağı için sonuçlar, %100 doğru sınıflandırma şeklinde idi ve aynı veriyle FFNN kullanılarak elde edilen %90 doğru sınıflandırmadan daha iyiydi.

## 11. SONUÇLAR

Numune tanıma metotları ile birleştirilmiş kimyasal sensör dizilerinin (MOS, MOSFET, QCM, SAW, polimer kaplı sensörler ve optik sensörler) kullanımı ve sınıflandırma metotları tanımlandı.

Bir sensör dizisinin veri analizi, ön işleme yaklaşımı ve numune tanıma metodunun seçimini ilgilendirir. İyi bir ön işleme metodu; kullanılan sensörlerin tipine ve analiz amacına bağlıdır. Örneğin; analit cevaplarından konsantrasyon etkilerini kaldırmak için kullanılacak olan ön işleme tekniği; normalizasyondur.

Her bir sensörün sabit durum cevapları, hangi metot için seçilmiş olursa olsun giriş olarak kullanılırsa, genellikle düşük özellik seçimine gerek duyacaktır. Bununla birlikte, sensör sinyallerinin zamanla değişen hali hesaba katılırsa, verinin boyutluluğu; PCA veya bir ön işleme metodu kullanılarak indirgenmelidir.

LDA, PCA, PLS gibi metotlar, veri setinin boyutluluğunu indirmek için faydalıdır. Bu metotların tamamı lineer metotlar oldukları için, sensör dizilerinin lineer cevap vereceği bilinen durumlarda en iyi şekilde kullanılırlar. Ayrıca bu metotlar, nonlineer veriyle de kullanılabilirler fakat analiz sonuçlarının yorumlarına dikkat edilmesi gereklidir.

Küme analizleri ve sinir ağları, meydana gelebilecek nonlineerliği doğru bir şekilde modellemeye çalışırlar. NN yaklaşımları modelleri tanımlamak için hiçbir istatistik bilgi üretmezler. Buna ilave olarak, bir NN için ayarlanabilir parametre sayısı artarken, bu yüzden eğitme numune sayısı ile eğitme zaman miktarına gerek duyulur. Bu, NN'lar için bir dezavantaj oluşturur.

Bir uygulamadaki analit sayısı yükselirken, veriyi analiz etmek için kullanılan algoritmanın karmaşıklığı da bu yüzden artar. KNN sınıflandırma şeması bu olay için önemli bir istisnadır. Değişik uygulamalarda daha komplike metotlardan daha iyi sonuçlar vermiştir. KNN için, hesaplama zamanı gözlem sayısının karesi ile artar. Zaman sebepleri yüzünden çok büyük bir veri seti için ARTMAP gibi farklı bir sınıflandırma şeması gerekli olabilir. Bu zorluk, birini dışarıda bırakma yerine n'yi dışarıda bırakma eğitim şeması kullanılarak hafifletilebilir. Çünkü algoritma çok basittir ve bilinmeyen sınıfların gözlemleri hesaba katılmazlar.

LVQ; back - propagation sinir ađına kıyasla daha az ayarlanabilir parametreye gereksinim duyar ve daha hızlı eđitme zamanı sonucunu verirler. Bu da daha gvenilir bir sınıflayıcı anlamına gelir.

ART tabanlı yaklaşımlar; nceden đrenilmiř bilgiyi unutmadan anında đrenmeyi sađlar. nceden đrenilmiř kategorileri evredeki deđişikliklere uyarlayarak yeniden kodlayabilirler. Eđitmek iin gereken zaman ok hızlı bir byklktr.

Fuzzy sinir modelleri, non fuzzy benzerlerine kıyasla yanlış sınıflandırılmıř numunelerin sayısının yarısı kadar numuneye sahiptir. Buna ilaveten, daha az zamanda daha azalmıř hata ile birleřirler. Fuzzy sinir ađlarının non fuzzy benzerlerinden daha stn olduđu grlr. Yaygın sinir ađlarına fuzzy parametrelerinin giriři, zor sınıflandırma problemlerinin zmnde nemli bir avantaj sunar.

Kimyasal sensr dizilerinden elde edilen veri analizleri iin pek ok metot mevcuttur. Sensr dizisi fabrikasyonundaki geliřmeler arttırılacak metotlara bađlıdırlar. Hesaplama bakıřları; kimyasal olarak benzer analitlerin geniř bir oranını izlemede kullanılacak sensr zerine kurulan diziye izin vermede anahtar bir rol oynayacaktır. Kimyasal sensr dizilerini kullanarak elde edilen bařarı, nemli analitleri belirlemek ve tanımlamak iin hesaplamalı tekniklerle birleřmiřtir. Bu arařtırma alanı gelecekte de byk ilgi grmeye devam edip dikkat ekecektir.

## KAYNAKLAR DİZİNİ

- [1] Singh, S., Hines, L. and Gardner, J.W., 1996, Fuzzy neural computing of coffee and tainted water data from an electronicnose, *Sensors and Actuators*, B30, 185-190.
- [2] Tsung Tan, T., Schmitt, V.O., Lucas, Q. and Isz. S., September/October 2001, *Electronic noses and electronic tongues*, LabPlus International, 16-19.
- [3] Pardo, M., Niederjaufner, G., Comini, E., Faglia, G. and Sberveglieri, G., 1999, Use of an electronic nose to classify different types of Italian cheeses, *Proceedings of the 5<sup>th</sup> international symposium on olfaction and the electronic nose*, Technomic Publishing Company Inc., 43-50.
- [4] Lonergan, M.C., Severin, E.J., Doleman, B.J., Beaver, S.A., Grubbs, R.H. and Lewis, N.S., 1996, *Chem. mater.*, 8, 2298- 2312.
- [5] Schalkoff, J. R., 1992, *Pattern recognition: statistical, structural and neural approaches*, Clemson University John Wiley & Sons Inc., 364 p.
- [6] Trunk, G.V., July 1979, A problem of dimensionality: a simple example, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.1, no.3, 306-307.
- [7] Chernoff, H., June 1973, The use of faces to represent points in k- dimensional space graphically, *J. Am. Statistical Assoc.*, vol.68, 361-368.
- [8] Friedman, J.H., 1987, Exploratory projection pursuit, *J. Am. Statistical Assoc.*, vol.82, 249-266.
- [9] Comon, P., 1994, Independent component analysis a new concept?, *Signal Processing*, vol.36, no.3, 287-314.
- [10] Bell, A. and Sejnowski, T., 1995, An information- maximization approach to blind separation, *Neural Computation*, vol.7, 1004-1034.
- [11] Cardoso, J., 1998, Blind signal separation: statistical principles, *Proc. IEEE*, vol.86, 2009-2025.
- [12] Lee, T.W., 1998, *Independent component analysis*, Dordrech Kluwer Academic Pub.
- [13] Devroye, L., Györfi, L. and Lugosi, G., 1996, *A probabilistic theory of pattern recognition*, Berlin: Springer- Verlag.
- [14] Haykin, S., 1999, *Neural networks, a comprehensive foundation*, second ed., Englewood Cliffs, N.J.: Prentice Hall.
- [15] Schölkopf, B., Smola, A. and Muller, K.R., 1998, Nonlinear component analysis as a Kernel eigenvalue problem, *Neural Computation*, vol.10, no.5; 1299- 1319.
- [16] Sammon Jr, J.W., 1969, A nonlinear mapping for data structure analysis, *IEEE Trans. Computer*, vol.18, 401- 409.

**KAYNAKLAR DİZİNİ (devam)**

- [17] Niemann, H., 1980, Linear and nonlinear mappings of patterns, *Pattern Recognition*, vol.12, 83- 87.
- [18] Lowe, D. and Webb, A.R., April 1991, Optimized feature extraction and the Bayes decision in feed- forward classifier networks, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.13, no.4, 264- 355.
- [19] Oja, E., 1992, Principal components and linear neural networks, *Neural Networks*, vol.5, no.6, 927- 936.
- [20] Bishop, C.M., 1995, *Neural networks for pattern recognition*, Oxford: Clarendon Press.
- [21] Lerner, B., Guterman, H., Aladjem, M. and Dinstein, I., 1999, A comparative study of neural network based feature extraction paradigms, *Pattern Recognition Letters*, vol.20, no.1, 7- 14.
- [22] Kohonen, T., 1995, *Self- organizing maps*, Springer Series in Information Sciences, vol.30, Berlin.
- [23] Jain, A.K. and Zongker, D., Feb. 1997, Feature selection: evaluation, application, and small sample performance, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.19, no.2, 153- 158.
- [24] Cover, T.M., 1974, The best two independent measurements are not the two best, *IEEE Trans. Systems, Man, and Cybernetics*, vol.4, 116- 117.
- [25] Pudil, P., Novovicova, J. and Kittler, J., 1994, Floating search methods in feature selection, *Pattern Recognition Letters*, vol.15, no.11, 1119- 1125.
- [26] Castellano, G., Fanelli, A.M. and Pelillo, M., 1997, An iterative pruning algorithm for feedforward neural networks, *IEEE Trans. Neural Network*, vol.8, no.3, 519- 531.
- [27] Mao, J., Mohiuddin, K. and Jain, A.K., Oct. 1994, Parsimonious network design and feature selection through node pruning, *Proc. 12<sup>th</sup> Int'l Conf. Pattern on Recognition*, 622- 624.
- [28] Fukunaga, K., 1990, *Introduction to statistical pattern recognition*, second ed., New York: Academic Press.
- [29] Oehler, K.L. and Gray, R.M., 1995, Combining image compression and classification using vector quantization, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.17, no.5, 461- 473.
- [30] Xie, Q.B., Laszlo, C.A. and Ward, R.K., 1993, Vector quantization technique for nonparametric classifier design, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.15, no.12, 1326-1330.



### KAYNAKLAR DİZİNİ (devam)

- [31] Devijver, P.A. and Kittler, J., 1982, Pattern recognition: a statistical approach, London: Prentice Hall.
- [32] Friedman, J.H., 1989, regularized discriminant analysis, J. Am. Statistical Assoc, vol.84, 165- 175.
- [33] Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J., 1984, Classification and regression trees, Wadsworth California.
- [34] Chou, P.A., 1991, Optimal partitioning for classification and regression trees, IEEE Trans. Pattern Analysis and Machine Intelligence, vol.13, no.4, 340- 354.
- [35] Quinlan, J.R., 1993, C4.5: Programs for machine learning, San Mateo California: Morgan Kaufmann.
- [36] Gelfand, S.B., Ravishankar, C.S. and Delp, E.J., Feb. 1991, An iterative growing and pruning algorithm for classification tree design, IEEE Trans. Pattern Analysis and Machine Intelligence, vol.13, no.2, 163- 174.
- [37] Mehta, M., Rissanen, J. and Agrawal, R., Aug. 1995, MDL- based decision tree pruning, Proc. First Int'l Conf. Knowledge Discovery in Databases and Data Mining, Montreal Canada.
- [38] Quinlan, J.R., 1987, Simplifying decision trees, Int'l J. Man- Machine Studies, vol.27, 221- 234.
- [39] Ho, T.K., Hull, J.J. and Srihari, S.N., 1994, Decision combination in multiple classifier systems, IEEE Trans. Pattern Analysis and Machine Intelligence, vol.16, no.1, 66- 75.
- [40] Xu, L., Krzyzak, A. and Suen C.Y., 1992, Methods for combining multiple classifiers and their applications in handwritten character recognition, IEEE Trans. Systems, Man and Cybernetics, vol.22, 418-435.
- [41] Hand, D.J., 1986, Recent advances in error rate estimation, Pattern Recognition Letters, vol.4, no.5, 335- 346.
- [42] Jain, A.K., Dubes, R.C. and Chen, C.C., May 1987, Bootstrap techniques for error estimation, IEEE Trans. Pattern Analysis and Machine Intelligence, vol.9, no.5, 628- 633.
- [43] Metz, C.E., 1978, Basic principles of ROC analysis, Seminars in Nuclear Medicine, vol.VIII, no.4, 283- 298.
- [44] Titterington, D., Smith, A. and Makov, U., 1985, Statistical analysis of finite mixture distributions, Chichester U.K.: John Wiley & Sons.

**KAYNAKLAR DİZİNİ (devam)**

- [45] Pudil, P., Novovicova, J. and Kittler, J., 1995, Feature selection based on the approximation of class densities by finite mixtures of the special type, *Pattern Recognition*, vol.28, no.9, 1389- 1398.
- [46] Barron, A., Rissanen, J. and Yu, B., Oct. 1998, The minimum description length principle in coding and modeling, *IEEE Trans. Information Theory*, vol.44, no.6, 2743- 2760.
- [47] Figueiredo, M., Leitao, J. and Jain, A.K., 1999, On fitting mixture models, *Energy Minimization Methods in Computer Vision and Pattern Recognition*, E. Hancock and M. Pellillo, eds., Springer- Verlag.
- [48] Johnson, R.A. and Wichern, D.W., 1982, *Applied multivariate statistical analysis*, Prentice-Hall: Englewood Cliffs, NJ.
- [49] Brereton, R.G., 1992, *Multivariate pattern recognition in chemometrics*, Elsevier: Amsterdam, vol.9.
- [50] Maricou, H., Preira, D., Verschuere, L., Philips, S. and Verstraete, W., 1998, *Water air soil pollut*, 107, 423- 442.
- [51] He, X.W., Xing, W.L. and Fang, Y.H., 1997, *Talanta*, 44, 2033.
- [52] Pearce, T.C. and Gardner, J.W., 1998, *Analyst*, 123, 2057- 2066.
- [53] Annor- Frempong, I.E., Nute, G.R., Wood, J.D., Whittington, F.W. and West, A., 1998, *Meat sci.*, 50, 139- 151.
- [54] Malinowski, E.R., 1991, *Factor analysis in chemistry*, 2<sup>nd</sup> ed., John Wiley & Sons: New York.
- [55] Baltes, H., Göpel, W. and Hesse, J., Eds, 1996, *Sensors update*, VCH Publishers, vol.2.
- [56] Freund, M.S. and Lewis, N.S., 1995, *Proc. Natl. Acad. Sci. U.S.A.*, 92, 2652.
- [57] Byfield, M.P. and May, I.P., 1996, *GEC J. Res.*, 13, 17.
- [58] Nanto, H., Kondo, K., Habara, M., Douguchi, Y., Waite, R.I. and Nakazumi, H., 1996, *Sens. actuators*, 35- 36, 183.
- [59] Di Natale, C., Davide, F.A.M., D'Amico, A., Sberveglieri, G., Nelli, P., Faglia, G. and Perego, C., 1995, *Sens. actuators*, 24- 25, 801.
- [60] Hong, H.K., Shin, H.W., Yun, D.H., Kim, S.R., Kwon, C.H., Lee, K. and Moriizumi, T., 1996, *Sens. actuators*, 35- 36, 338.
- [61] Auge, j., Hauptmann, P., Hartmann, J., Rosler, S. and Lucklum, R., 1995, *Sensors and Actuators B*, 26-27, 181-186.

**KAYNAKLAR DİZİNİ (devam)**

- [62] Llobet, E., Brezmes, J., Vilanova, X., Sueiras, J.E. and Correig, X., 1997, *Sens. actuators*, 41, 13.
- [63] Nakamura, M., Sugimoto, I. and Kuwano, H., 1996, *Sens. actuators*, 33, 122- 127.
- [64] Nakamura, M., Sugimoto, I., Kuwano, H. and Lemos, R., 1994, *Sens actuators*, 20, 231-237.
- [65] Eklov, T., Johansson, G., Winqvist, F. and Lundstrom, I.J., 1998, *Sci. Food Agric.*, 76, 525-532.
- [66] Schweizer- Berberich, P.M., Vaihinger, S. and Gopel, W., 1994, *Sens. actuators*, 18- 19, 282- 290.
- [67] Kalivas, J.H. and Lang, P.M., 1993, *Mathematical analysis of spectral orthogonality*, Marcel Decker: New York.
- [68] Carrasco, A., Saby, C. and Bernadet, P., 1998, *Flavour Fragrance J.*, 13, 335- 348.
- [69] Niebling, G. and Muller, R., 1995, *Sens. actuators*, 24- 25, 805.
- [70] Grate, J.W., Wise, B.M. and Abraham, M.H., 1999, *Anal. chem.*, 71, 4544- 4553.
- [71] Massart, D.L. and Kaufman, L., 1983, *The interpretation of analytical chemical data by the use of Cluster analysis*, John Wiley and Sons: New York.
- [72] Sammon, J.W., 1969, *IEEE Trans. Comput.*, C-18, 401.
- [73] Kowalski, B.R. and Bender, C.F., 1973, *Am. chem. soc.*, 95, 686.
- [74] Visser, F.R. and Taylor, M., 1998, *Sens. stud.*, 13, 95- 120.
- [75] Anklam, E., Lipp, M., Radoviç, B., Chiavaro, E. and Palla, G., 1998, *F. chem.*, 61, 243-248.
- [76] Cover, T.M. and Hart, P.E., 1967, *IEEE Trans. Inform. Theory*, 13, 21.
- [77] Kay, S.M. and Marple, S.L., 1981, *Proc. IEEE*, 69, 1380.
- [78] Rumelhart, D.E. and McClelland, J.L., 1986, *Parallel distributed processing, Explorations in the Microstructure of Cognition, vol.1: Foundations*, MIT Press: Cambridge.
- [79] Livingstone, D.J. and Manallack, P.T., 1993, *J. med. chem.*, 36, 1295- 1297.
- [80] Gibson, T.D., Prosser, O., Hulbert, J.N., Marshall, R.W., Corcoran, P., Lowery, P., Ruck- Keene, E.A. and Heron, S., 1997, *Sens. actuators*, 44, 413.
- [81] Corcoran, P., Lowery, P. and Anglesea, J., 1998, *Sens. actuators*, 48, 448- 455.

**KAYNAKLAR DİZİNİ (devam)**

- [82] Kohonen, T., 1997, Self organizing maps, 2<sup>nd</sup> ed., Springer: Berlin.
- [83] Marco, S., Ortega, A., Pardo, A. and Samitier, J., 1998, IEEE Trans. Instrum. Meas., 47, 316- 321.
- [84] Davide, F.A.M., Di Natale, C. and D'Amico, A., 1994, Sens. actuators, 18- 19, 244- 258.
- [85] Di Natale, C., Davide, F.A.M. and D'Amico, A., 1995, Sens. actuators, 26- 27, 237- 241.
- [86] Johnson, S., 1999, Ph.D. Thesis, The Pennsylvania State University.
- [87] Johnson, S.R., Sutter, J.M., Engelhardt, H.L., Jurs, P.C., White, J., Kauer, J.S., Dickinson, T.A. and Walt, D.R., 1997, Anal. chem., 69, 4641- 4648.
- [88] Hibbert, D.B., 1993, Chemom. Intell. Lab. Sys., 19, 227.
- [89] Lucasius, C.B. and Kateman, G., 1993, Chemom. Intell. Lab. Sys., 19, 1.
- [90] Llobet, E., Hines, E.L., Gardner, J.W. and Franco, S., 1999, Meas. Sci. Technol., 10, 538- 548.
- [91] Vandeginste, B.G.M., Massart, D.L., Buydens, L.M.C., De Jong, S., Lewi, P.J. and Smeyers- Verbeke, J., 1998, Handbook of chemometrics and qualimetrics: Part B, Elsevier: Amsterdam.
- [92] Kasuba, T., 1993, AI Expert, 8, 18.
- [93] Shukla, K.K., Das, R.R. and Dwivedi, R., 1998, Sens. actuators, 50, 194- 203.

**EKLER**



## 1. Cyranose 320

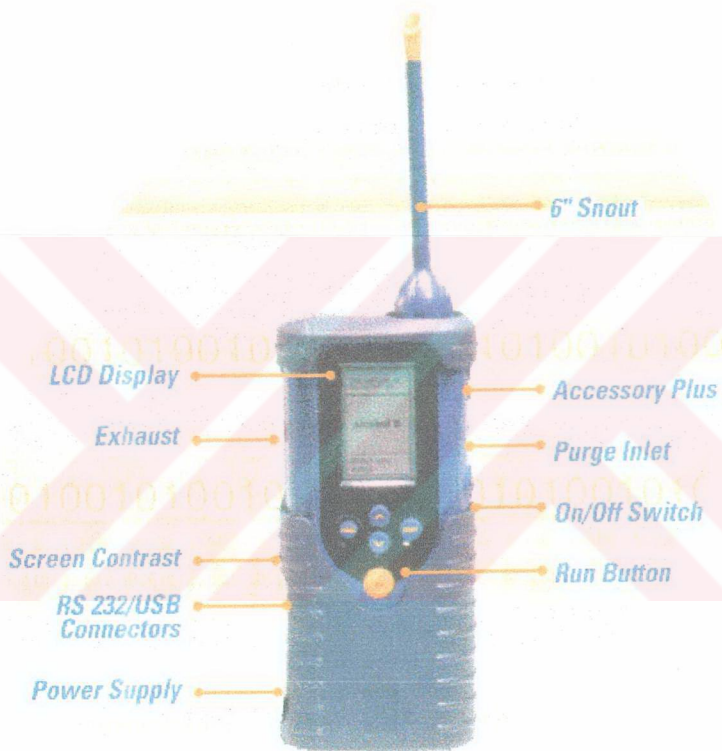
Cyrano Science firmasının bir ürünü olan Cyranose 320 isimli elektronik burun, 32 polimer kompozit sensörden oluşan bir diziye sahiptir. Bu cihaz bir kokuya maruz bırakıldığında polimer kompozit direncinde değişimler meydana gelir ve kullanılacak olan veri bu değişimlerin bir setiyle üretilir. Toplanan işlenmemiş veri yüksek frekans gürültüsünden kurtulmak için filtre edilir daha sonra filtreden geçirilen veri zayıflatılır. Bir koku izi veya numune; dizide yer alan 32 adet sensörün teker, teker verilerinin zayıflatılmasıyla elde edilir. Elde edilen koku izine normalizasyon ve ölçekleme teknikleri uygulanır ve numune tanımlama için cihazın üzerinde bulunan lineer ayırım analizleri kullanılarak koku izi sınıflandırılır.

Cyranose 320; yiyecek, paketlenme, plastik, kimya endüstrilerinde ve tıpta kullanılmaktadır. Bu teknoloji laboratuvar ortamındaki materyallerin analizinde zaman ve para kaybı miktarlarını önemli ölçüde düşürür. Yiyecek şirketlerinde geliştirme, hazırlama ve dağıtma aşamalarında ham maddedeki bozulma ve ürün tutarlılığının sürekli incelenmesi için kullanılırlar. Ayrıca kara mayınlarının yer tespitinde, yayılan zehirli maddede kimyasal teşhisinde, vücut akışkanlarıyla ilgili özel metabolik durumlarda belirli kimyasal bileşiklerin teşhisinde, belirli deri hastalıkları veya bakteriyel enfeksiyonların teşhisinde kullanılmaktadırlar. Cyrano Sciences'ın Nose-Chip teknolojisi ile buzdolabı ve mikro dalga fırın gibi herkes tarafından bilinen uygulamaların birleştirilmesi sonucu gıda zehirlenmesi veya bozulmuş gıdanın tüketilmesi riski minimize edilebilir.

### 1.1. Cyranose 320'nin özellikleri

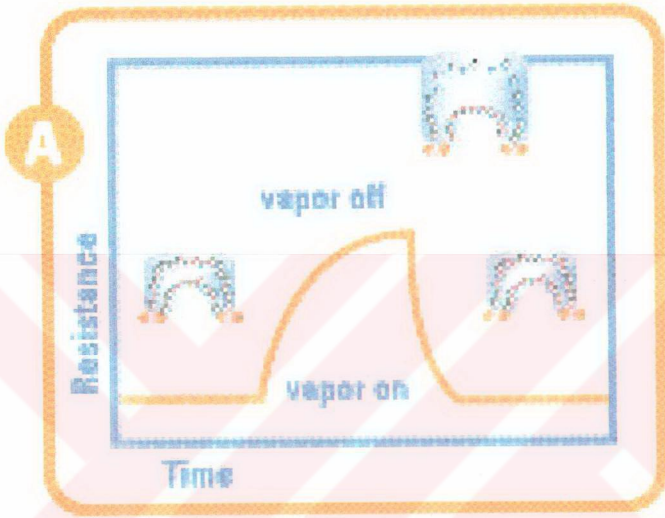
Ağırlık	32 Ons (0.91 kg)
Sensör Tipi	32 Kanallı Polimer Kompozit sensör dizisi
Batarya Tipi	NiMH, 4 AA batarya paketi veya 4 AA Alkalın pil
Batarya Ömrü	Normal çalışma şartlarında 3 saat
Batarya Şarjı	3 Saat harici adaptör ile
Boyutları	10 x 22 x 5 cm
Genel Güç Adaptörü	110 - 220 V AC harici güç adaptörü
Display	320 x 200 Grafik ile LED arka fon
Tepki Zamanı	10 Saniye
Alan Kalibrasyonu	TDB
Giriş Probu	2" & 4" Birbirinin yerine kullanılabilir iğne ile standart dişi Luer kilidi veya fiş adaptör
Tuş Takımı	Toplam 5 buton, Scroll Up/Down, Select, Cancel, Run
İletişim	RS-232 Link, 57600 bps 'ye kadar
Sampling Pump	50 - 180 cc/ min
Algoritma	K-NN, Kmeans, PCA, CDA (Canonical Discriminant Analysis)
Çalışma Sıcaklığı	0 ile 40 °C (32 - 104 °F) arasında
Nem Oranı	0 - 95 %, yoğunlaşma olmaz
Depolama Sıcaklığı	-20 ile 50 °C (-4 - 122 °F) arasında
<b>VERİ SAKLAMA KAPASİTESİ</b>	
Metot Sayısı	Cyranose 320 üzerinde 5 adet, PC üzerinde ise sonsuz
Metot Başına Sınıf Sayısı	6
Sınıf Başına Ortaya Konan Eğitim Sayısı	10
Teşhisi Saklama Sayısı	Cyranose 320 üzerinde 100'e kadar, PC'ye kaydedilip Cyranose320'den silinebilir.

## 1.2. Cyranose 320'nin görünümü

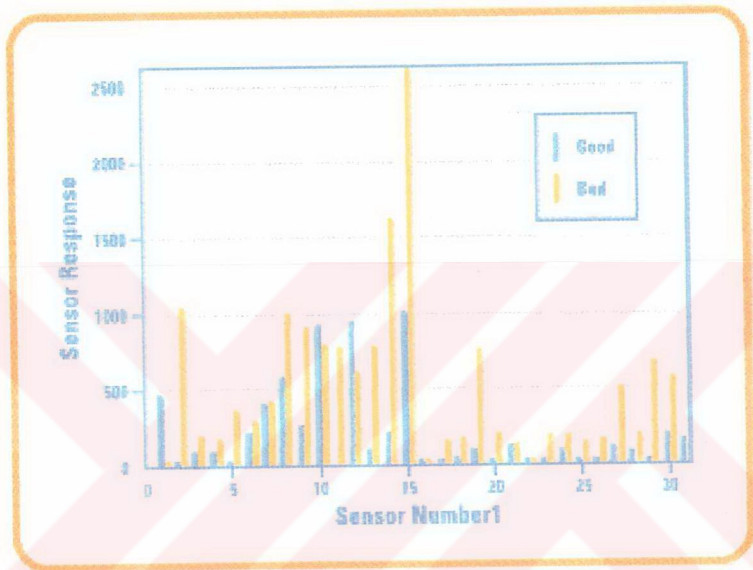




1.3. Kompozit malzemenin analite maruz kalma durumunda cevabi



#### 1.4. Dizideki 32 adet sensörün cevabı



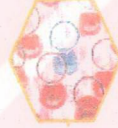
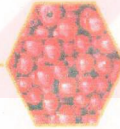
## 1.5. Cyranose 320'de işlem süreci



## 1.6. Uygulama alanları



## 1.6. Uygulama alanları



## 1.6. Uygulama alanları

