

**DESTEK VEKTÖR MAKİNELERİ İLE SES TANIMA
UYGULAMASI**

**Pamukkale Üniversitesi
Fen Bilimleri Enstitüsü
Yüksek Lisans Tezi
Elektrik-Elektronik Mühendisliği Anabilim Dalı**

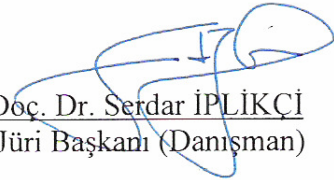
Osman ERAY

Danışman:Doç. Dr. Serdar İPLİKÇİ

**Ağustos 2008
DENİZLİ**

YÜKSEK LİSANS TEZİ ONAY FORMU

Osman ERAY tarafından Doç. Dr. Serdar İPLİKÇİ yönetiminde hazırlanan “**Destek Vektör Makineleri ile Ses Tanıma Uygulaması**” başlıklı tez tarafımızdan okunmuş, kapsamı ve niteliği açısından bir Yüksek Lisans Tezi olarak kabul edilmiştir.


Doç. Dr. Serdar İPLİKÇİ
Jüri Başkanı (Danışman)

Yrd. Doç. Dr. Aydın KIZILKAYA
Jüri Üyesi



Yrd. Doç. Dr. Sezai TOKAT
Jüri Üyesi



Pamukkale Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu'nun
.../.../..... tarih ve sayılı kararıyla onaylanmıştır.

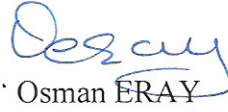
Prof. Dr. Mehmet Ali SARIGÖL
Müdür

Bu tezin tasarımı, hazırlanması, yürütülmesi, araştırılmalarının yapılması ve bulgularının analizlerinde bilimsel etiğe ve akademik kurallara özenle riayet edildiğini; bu çalışmanın doğrudan birincil ürünü olmayan bulguların, verilerin ve materyallerin bilimsel etiğe uygun olarak kaynak gösterildiğini ve alıntı yapılan çalışmalara atfedildiğini beyan ederim.

İmza

Öğrenci Adı Soyadı

:


Osman ERAY

TEŐEKKÜR

Bu tez alıőmasında bana her zaman destek olan ve üzerimde büyük emekleri bulunan babam İsmail ERAY'a ve annem Hasibe ERAY'a; bana her zaman moral veren sevgili kardeőim Glsm ERAY'a; alıőmalarım sresince bana verdiėi moral ve destekten, gsterdiėi sabır ve anlayıőtan dolayı sevgili eőim Feden ERAY'a ok teőekkr ederim.

Tez alıőmam boyunca bana yardımcı olan ve önerileri ile beni ynlendiren tez danıőmanım Do. Dr. Serdar İPLİKi'ye katkılarından dolayı ok teőekkr ederim. Ayrıca Yrd. Do. Dr. A. Kadir YALDIR, Yrd. Do. Dr. Aydın KIZILKAYA ve Yrd. Do. Dr. Sezai TOKAT'a da önerileri ve katkılarından dolayı teőekkr ederim.

ÖZET

DESTEK VEKTÖR MAKİNELERİ İLE SES TANIMA UYGULAMASI

ERAY, Osman

Yüksek Lisans Tezi, Elektrik-Elektronik Mühendisliği ABD

Tez Yöneticisi: Doç. Dr. Serdar İPLİKÇİ

Ağustos 2008, 79 Sayfa

Ses tanıma sistemleri insan-bilgisayar arası iletişimi hızlı ve etkin bir şekilde gerçekleştirmeyi amaçlar. Bu amaçla, ses tanıma teknolojilerini geliştirmek için, son yıllarda çeşitli araştırma ve çalışmalar yapılmaktadır. Bu çalışmalara örnek olarak ses tanıma, konuşmacı tanıma ve konuşmacı doğrulama verilebilir. Fakat Türkçe ses tanıma üzerine çok fazla çalışma bulunmamaktadır. Bu çalışmada ses tanıma sistemleri incelenmiş, kullanılan yöntemler araştırılmış ve bir Türkçe ses tanıma uygulaması geliştirilmiştir.

Uygulama, ses kodlama ve ses tanıma bölümlerinden oluşmaktadır.

Bu çalışmada ilk olarak, bilgisayarda sıklıkla kullanılan 20 adet Türkçe komut belirlenmiştir. Her komuttan 20 adet kayıt yapılmıştır. Toplam 400 sözcük mikrofon ile bilgisayara kaydedilmiştir.

Uygulamanın ses kodlama bölümünde, bilgisayara kaydedilen bu sözcükler Doğrusal Önkestirim Kodlama (LPC) yöntemi ile kodlanmış ve her bir sözcüğe ilişkin LPC parametreleri elde edilmiştir.

Uygulamanın ses tanıma bölümü, eğitim ve test aşamalarından oluşmaktadır. Bu bölümde Destek Vektör Makineleri (SVM) yöntemi kullanılmıştır. İki çeşit SVM sınıflandırıcısı tasarlanmıştır. Bunlar, Soft Margin Destek Vektör Makinesi sınıflandırıcısı ve Least Square Destek Vektör Makinesi sınıflandırıcısıdır. Kodlanmış 200 sözcük Destek Vektör Makinelerinin eğitim aşaması için, 200 sözcük Destek Vektör Makinelerinin test aşaması için kullanılmıştır.

Uygulamada, Soft Margin SVM sınıflandırıcısı için %91 doğru tanıma başarısı; Least Square SVM sınıflandırıcısı için %71 doğru tanıma başarısı elde edilmiştir.

Anahtar Kelimeler: Ses Tanıma, Destek Vektör Makineleri, LPC

Doç. Dr. Serdar İPLİKÇİ

Yrd. Doç. Dr. Aydın KIZILKAYA

Yrd. Doç. Dr. Sezai TOKAT

ABSTRACT**THE SPEECH RECOGNITION APPLICATION WITH SUPPORT VECTOR MACHINES**

ERAY, Osman

M. Sc. Thesis in Electric-Electronic Engineering
Supervisor: Assoc. Prof. Dr. Serdar İPLİKÇİ

August 2008, 79 Pages

Speech recognition systems aim to perform communication between human and computer quickly and efficiently. For this purpose, in recent years various researches and studies have been conducted to develop the speech recognition technologies. Exemplary to this studies, speech recognition, speaker recognition and speaker verification can be given. However, there are not so many studies on Turkish speech recognition. In this study, speech recognition systems have been examined, the methods existing in the literature have been investigated and a Turkish speech recognition application was developed.

The application consists of speech coding and speech recognition parts.

In this study, firstly, 20 Turkish commands which are often used at computer were determined. 20 records were done for each command. Total 400 words were recorded by microphone to computer.

In speech recognition part of the application, this words which were recorded to computer were coded with Linear Predictive Coding (LPC) method and thus the related LPC parameters have been obtained.

Speech recognition part of the application consists of training and testing phases. In this part, Support Vector Machines (SVM) method was used for recognition. Two types of SVM classifiers were designed. These classifiers are Soft Margin Support Vector Machine classifier and Least Square Support Vector Machine classifier. 200 words which have been coded were used for training phase of Support Vector Machines and 200 words which have been coded were used for testing phase of Support Vector Machines.

In the application, the Soft Margin SVM classifier has yielded 91% recognition performance, while the Least Square SVM classifier has 71%.

Keywords: Speech Recognition, Support Vector Machines, LPC

Assoc. Prof. Dr. Serdar İPLİKÇİ
Asst. Prof. Dr. Aydın KIZILKAYA
Asst. Prof. Dr. Sezai TOKAT

İÇİNDEKİLER

Yüksek Lisans Tezi Onay Formu.....	i
Bilimsel Etik Sayfası.....	ii
Teşekkür.....	iii
Özet.....	iv
Abstract.....	v
İçindekiler.....	vi
Şekiller Dizini.....	viii
Tablolar Dizini.....	ix
Simge ve Kısaltmalar Dizini.....	x
1. GİRİŞ.....	1
1.1. Ses Tanımanın Tarihçesi.....	1
1.2. Literatür Araştırması.....	3
2. PROBLEMİN TANIMI.....	6
3. SES.....	8
3.1. Sesin Oluşumu ve Özellikleri.....	8
3.2. Sesin Sentezi ve Analizi.....	10
3.2.1. Sesin analizi.....	11
3.2.1.1. Sesin sayısallaştırılması.....	12
3.2.1.2. Özellik vektörünün çıkarılması.....	12
3.2.1.3. Sesin spektral analizi.....	12
3.2.1.3.1. Yöntemin seçimi.....	13
3.2.1.3.2. Örnekleme hızları.....	14
3.2.1.3.3. Analiz aralığının seçimi.....	14
3.2.1.3.4. Pencereleme.....	14
3.2.1.3.5. Önvurgulama.....	15
3.2.1.4. Hızlı Fourier Dönüşümü (FFT).....	15
3.2.1.5. Dilimleme (Segmentation).....	16
3.2.1.6. Zaman normalizasyonu (Time normalization).....	16
4. SES KODLAMA.....	17
4.1. Doğrusal Önkestirim Kodlama (LPC).....	17
4.1.1. LPC yöntemindeki aşamalar.....	19
4.1.1.1. Önvurgulama.....	19
4.1.1.2. Çerçeveleme.....	19
4.1.1.3. Pencereleme.....	19
4.1.1.4. Otokorelasyon analizi.....	20
4.1.1.5. LPC analizi.....	21
4.2. Doğrusal Önkestirim Cepstrum Katsayıları (LPCC).....	21
4.2.1. LPCC yöntemindeki aşamalar.....	22
4.2.1.1. Önvurgulama.....	22
4.2.1.2. Çerçeveleme.....	22
4.2.1.3. Pencereleme.....	22
4.2.1.4. Otokorelasyon analizi.....	22
4.2.1.5. LPC analizi.....	22
4.2.1.6. LPC-LPCC parametre dönüşümü.....	23

4.2.1.7. Parametre ağırlıklandırma.....	23
4.2.1.8. Türev hesaplama	23
4.3. Mel Frekansı Cepstrum Katsayıları (MFCC).....	23
4.3.1. MFCC yöntemindeki aşamalar	24
4.3.1.1. Çerçeveleme	24
4.3.1.2. Pencereleme	24
4.3.1.3. Hızlı Fourier Dönüşümü (FFT).....	24
4.3.1.4. Mel-frekansına çevirme	24
4.3.1.4.1. Mel-filtre bankası tasarımı	25
4.3.1.5. Mel-cepstrum	25
4.4. Algısal Doğrusal Önkestirim Kodlama (PLP Coding)	25
5. SES TANIMA	27
5.1. Şablon Modeller	29
5.1.1. Vektör Uzaklık Ölçümü (VQ).....	29
5.1.2. Dinamik Zaman Uydurma (DTW).....	30
5.2. Stokastik Modeller	30
5.2.1. Gaussian Karma Modelleri (GMMs)	30
5.2.2. Gizli Markov Modeller (HMMs)	32
5.3. Yapay Sinir Ağları	33
5.4. Bulanık-Sinirsel Sistemler	35
5.4.1. Bulanık-sinirsel sınıflandırıcılar	35
5.4.1.1. Bulanık çıkarım.....	35
5.4.1.2. Bulanık-sinirsel sınıflandırıcılar	37
5.5. Destek Vektör Makineleri	38
5.5.1. Kısıtlamalı optimizasyon	38
5.5.1.1. Kısıtlamalı optimizasyon teorisi	38
5.5.1.2. Lagrange teorisi.....	40
5.5.2. İstatistiksel Öğrenme Teorisi	42
5.5.2.1. VC-boyutu.....	43
5.5.2.2. Yapısal risk minimizasyonu	43
5.5.3. Destek Vektör Makinelerine giriş	44
5.5.3.1. Destek Vektör Makinesi sınıflandırıcıları.....	44
5.5.3.1.1. Doğrusal Destek Vektör Makinesi sınıflandırıcıları	45
5.5.3.1.2. Doğrusal olmayan Destek Vektör Makinesi sınıflandırıcıları	53
5.5.4. Soft Margin Destek Vektör Makinesi sınıflandırıcıları	56
5.5.5. En Küçük Kareler Destek Vektör Makinesi sınıflandırıcıları.....	58
5.5.5.1. İkili Sınıf Tahmini için Standart Destek Vektör Makineleri.....	59
5.5.5.2. En Küçük Kareler Destek Vektör Makinesi sınıflandırıcıları.....	60
6. UYGULAMA SONUÇLARI	63
7. SONUÇ VE ÖNERİLER	74
KAYNAKLAR	76
ÖZGEÇMİŞ	79

ŞEKİLLER DİZİNİ

Şekil 3.1 Ses oluşumu organları	9
Şekil 3.2 İnsanda ses oluşumu mekanizmasının şematik gösterimi	10
Şekil 3.3 'ball', 'bar', 'bough', 'buy' kelimelerinin spektrogramı (Hagiwara 2008)	13
Şekil 4.1 Doğrusal Önkestirim Kodlama yönteminin blok diyagram olarak gösterimi	19
Şekil 4.2 Çerçeveleme	20
Şekil 4.3 Doğrusal Önkestirim Cepstrum Katsayıları yönteminin blok diyagramı	22
Şekil 4.4 MFCC'nin blok diyagram olarak gösterimi (Demirci 2005).....	24
Şekil 4.5 Mel filtre bankası (Uzunçarşılı 2005).....	26
Şekil 5.1 Bir ses tanıma sisteminin blok diyagramı (Rabiner vd. 1993)	27
Şekil 5.2 Bir ses tanıma sisteminin modülleri (Rabiner vd. 1993)	28
Şekil 5.3 2-boyutlu Vektör Uzaklık Ölçümü (Uzunçarşılı 2005)	29
Şekil 5.4 M bileşenli Gaussian yoğunluk (Uzunçarşılı 2005)	31
Şekil 5.5 Örnek bir Gizli Markov Model (Baygün 2006).....	32
Şekil 5.6 Tek girişli bir nöron modeli	33
Şekil 5.7 Tipik çok girişli tek çıkışlı bir ANN modeli.....	34
Şekil 5.8 Bulanık mantıkta <i>Modus Ponens</i>	36
Şekil 5.9 Bir 3-sınıf sınıflandırma örneği	37
Şekil 5.10 Bir 3-sınıf sınıflandırma örneğinin ağaç olarak gösterimi	37
Şekil 5.11 Bir bulanık-sinirsel sınıflandırma ağacı	38
Şekil 5.12 Modelleme hataları	43
Şekil 5.13 Doğrusal ayrılabilme durumunda optimal ayırıcı aşırıdüzlem	46
Şekil 5.14 Doğrusal ayrılabilme durumunda optimal ayırıcı aşırıdüzlem	47
Şekil 5.15 Doğrusal ayrılamama durumunda optimal ayırıcı aşırıdüzlem	51
Şekil 5.16 Bir doğrusal olmayan Destek Vektör Makinesi.....	53
Şekil 5.17 Giriş uzayını özellik uzayına eşleme	54
Şekil 5.18 Giriş uzayını özellik uzayına eşleme	54
Şekil 6.1 Uygulamanın genel yapısı	64
Şekil 6.2 LPC analizindeki aşamalar	65
Şekil 6.3 Eğitim örüntülerinin LPC yöntemi ile kodlanması.....	67
Şekil 6.4 Test örüntülerinin LPC yöntemi ile kodlanması	67
Şekil 6.5 Test aşamasında sistemin çalışmasının şematik olarak gösterimi	69
Şekil 6.6 SM SVM sınıflandırıcısının eğitim aşaması	70
Şekil 6.7 LS SVM sınıflandırıcısının eğitim aşaması	70
Şekil 6.8 SM SVM sınıflandırıcısının test aşaması	71
Şekil 6.9 LS SVM sınıflandırıcısının test aşaması	72

TABLolar DİZİNİ

Tablo 1.1 Ses tanıma sistemleri için geliştirilen yöntemler	3
Tablo 6.1 Bilgisayar ortamına mikrofon ile kaydedilen sözcükler	64
Tablo 6.2 SM SVM sınıflandırıcısının örüntü tanıma başarısı	72
Tablo 6.3 LS SVM sınıflandırıcısının örüntü tanıma başarısı.....	73

SİMGE VE KISALTMALAR DİZİNİ

ANN	Artificial Neural Network
DTW	Dynamic Time Warping
FFT	Fast Fourier Transform
GMM	Gaussian Mixture Model
HMM	Hidden Markov Model
LPC	Linear Predictive Coding
LPCC	Linear Predictive Cepstrum Coefficient
LS	Least Square
MFCC	Mel Frequency Cepstrum Coefficient
NN	Neural Network
PLP	Perceptual Linear Prediction
SM	Soft Margin
SVM	Support Vector Machine
VQ	Vector Quantization

1. GİRİŞ

1.1. Ses Tanımının Tarihçesi

Dünyada, makine ile ses tanıma arařtırmaları, 50 yılı ařkın süredir yapılmaktadır.

Makine ile ses tanıma için sistem geliştirme çalışmaları ilk kez 1950'li yıllarda başladı. O yıllarda çeřitli arařtırmacılar ses-dağılım ve ses-bilim üzerine ilk fikirleri ortaya attı. Bell Laboratuvarlarında tek bir konuşmacı için, ayrılmıř bir rakam tanıma sistemi geliştirildi (Davis vd. 1952). RCA Laboratuvarlarında bağımsız bir çalışma daha yapıldı (Olson vd. 1956). Bu çalışma tek bir konuşmacıya ait 10 ayrı heceyi tanıma çalışmasıydı. İngiltere'de UCL'de 4 sesli ve 9 sessizi tanıyabilen bir sesbirim tanıyıcı geliştirildi (Fry vd. 1959). Bu çalışmada tanıma kararını gerçekleřtirmek üzere bir spektrum analizatörü ve bir örüntü eşleřtirici kullanıldı. Diđer bir çalışma MIT Lincoln Laboratuvarlarında yapıldı (Forgie vd. 1959). Bu çalışma konuşmacıdan bağımsız bir ses tanıma sistemi idi. Bu çalışmada spektral bilgiyi elde etmek için bir filtre bankası analizatörü ve konuşulan sesliyi tanımayı gerçekleřtirmek için de ses tüpü tınlarının zaman deęiřimlerini kestiren bir sistem kullanılmıřtı.

1960'lı yıllarda ses tanıma ile ilgili birkaç temel fikir ortaya çıktı ve yayımlandı. Bu yıllarda Tokyo'da yapılan ilk çalışma Radio Arařtırma Laboratuvarlarında gerçekleřtirilen bir sesli tanıyıcı donanım idi (Suzuki vd. 1961). Japonya'da Kyoto üniversitesinde de bir sesbirim tanıyıcı donanım gerçekleřtirildi (Sakai vd. 1962). Makine ile ses tanıma çalışması olarak ortaya atılan bir başka yaklařım, sürekli ses alanlarını tanıma için öne sürülen, ses-birimlerinin dinamik izlenmesi yöntemi idi (Reddy 1967). Benzer tarihlerde Sovyetler Birlięi'nde, bir çift ses ifadesi üzerinde zaman düzenleme (uydurma) için dinamik programlama metodları önerildi (Vintsyuk 1968).

1970'li yıllarda ses tanıma için dönüm noktası sayılabilecek birçok başarılı, önemli çalışma gerçekleřtirildi. Bunlardan ilki ayrılmıř kelime tanıma çalışmasının

uygulanabilir bir teknoloji olarak Rusya'da (Velichko vd. 1970); Japonya'da (Sakoe vd. 1978) ve ABD'nde (Itakura 1975) gerçekleştirilen temel çalışmalar sonucunda ortaya çıkması idi. Rus arařtırmacılar ses tanıma ierisinde rüntü tanıma fikrinin kullanımını geliřtirmeye yardım ettiler, Japon arařtırmacılar da dinamik programlama yöntemlerinin ses tanıma üzerinde ne kadar başarılı bir şekilde uygulanabileceğini gösterdiler ve ABD'de yapılan arařtırmalar da Doğrusal Önkestirim Kodlama (LPC)'nin ses tanıma için başarıyla uygulanabilir olduğunu gösterdi.

1980'li yıllarda ses tanımada řablon tabanlı yaklaşımların (örneğin, Dinamik Zaman Uydurma, DTW gibi) yerini istatistiksel modelleme yöntemleri (özellikle Gizli Markov Model yaklaşımı, HMM) almaya başladı. 1980'li yılların ortalarında HMM teorisi tüm dünyada birçok ses tanıma laboratuvarında yaygın bir şekilde uygulandı. Bu yıllarda ortaya atılan bir diđer yeni teknoloji veya fikir ise Yapay Sinir Ağları (ANN) idi. Yine bu yıllarda ses tanıma üzerinde birçok yeni yollar, yöntemler (örneğin, Mel Frekanslı Cepstrum Katsayıları, MFCC gibi) önerildi. 1980'li yılların sonuna gelindiğinde DARPA (İleri Savunma Arařtırma Projeleri Acentası) topluluđu geniş kelime sürekli ses tanıma sistemleri geliřtirilmesini destekledi ve bu bağlamda dünyada birçok arařtırma laboratuvarında arařtırmalar yapıldı. DARPA programı 1990'lı yıllara kadar devam etti.

1990'lı yıllara gelindiğinde artık ses tanıma teknolojileri telefon ağları için de artan bir biçimde kullanılıyor idi.

1990'lı yıllarda veri sınıflandırma ve regresyon problemlerinin çözümü amacıyla yeni bir yöntem olan Destek Vektör Makineleri (SVM) yöntemi ortaya atıldı (Vapnik 1995, 1998). Bu yöntem 2000'li yıllarda ses tanıma, konuşmacı tanıma ve doğrulama işlemleri için çeřitli arařtırmalarda kullanıldı.

Ses tanıma sistemleri için geliřtirilen yöntemler kronolojik bir sırada tablo 1.1'de verilmiştir.

Tablo 1.1 Ses tanıma sistemleri için geliştirilen yöntemler

GELİŞME	TARİH
Doğrusal Önkestirim Kodlama (LPC)	1969-1970
Dinamik Zaman Uydurma (DTW)	1970'ler
Gizli Markov Modeller (HMMs)	1975
Mel-Frekans Cepstrum Katsayıları (MFCCs)	1980
Dil Modelleri (Language Models)	1980'ler
Yapay Sinir Ağları (ANNs)	1980'ler
Destek Vektör Makineleri (SVMs)	1995, 1998
Çekirdek-tabanlı Sınıflandırıcılar	1998
Dinamik Bayesian Ağları	1999

1.2. Literatür Araştırması

Ses tanıma uygulamaları için N tane en iyi aday tabanlı bir eğitim algoritması önerildi (Chen vd. 1994). Bu çalışmada temel alınan yöntem Gizli Markov Modellerdir.

Yapay Sinir Ağları ile bir ses tanıma çalışması gerçekleştirildi (El-Ramly vd. 2002). Bu çalışma Arapça sesbirimler üzerinde denendi. Çalışmada Zaman Gecikmeli Sinir Ağları kullanıldı. Bunun nedeni ise bu yöntemin Arapça ses tanıma çalışmasında ses ile ilgili olaylar arasındaki ilişkileri temsil etmede başarılı olması idi. İki Arapça kategori seçildi. Bunlar, 'nasal' ifadeler ve 'voiced-stop' ifadelerdi. Sistemin 'nasal' ifadeleri tanıma başarısı %91 ; 'voiced-stop' ifadeleri tanıma başarısı da %82 olarak gerçekleşti.

Ses tanıma problemi için bir Geliştirilmiş Dinamik Zaman Uydurma (E-DTW) modeli önerildi (Yaniv vd. 2003). Bu çalışmada ses tanıma problemlerine etkili bir çözüm getirmek üzere Gizli Markov Modeller (HMMs) için Forward ve Viterbi algoritmaları geliştirildi. Ayrıca Gizli Markov Modeller için Baum-Welsh kestirim algoritması geliştirildi ve Baum eşitsizliği üzerine kurulu olan bu yeni modelin model parametrelerinin kestirimi için bir iteratif yöntem elde edildi.

Otomatik duygu tanıma sistemi geliştirildi (Lin vd. 2005). Sistem, duygu tanıma işlevini ses sinyali ile gerçekleştiriyordu. Çalışmada iki sınıflandırma metodu kullanıldı, bunlar, Gizli Markov Modeller ve Destek Vektör Makineleriydi. Sistem beş duygusal durumu sınıflandırıyordu, bunlar; öfke, mutluluk, üzüntü, şaşkınlık ve durağan olma idi. Gizli Markov Model içerisinde 39 adet aday özellik çıkarıldı ve Sıralı İleri Seçim yöntemi kullanılarak en iyi özellik alt kümesi bulundu. Ardından seçilen özellik alt kümesinin sınıflandırma performansı Mel Frekanslı Cepstrum Katsayıları (MFCC) ile karşılaştırıldı. Destek Vektör Makineleri yöntemi üzerine kurulu olan metod içerisinde Mel frekanslı ölçeği alt bant enerjileri arasındaki farkı ölçen yeni bir vektör önerildi. Aynı zamanda önerilen vektör kullanılarak En yakın K-komşu sınıflandırıcısının performansı araştırıldı. ‘Danish’ duygusal konuşma veritabanı üzerinde cinsiyete bağımlı ve cinsiyetten bağımsız uygulamalar yapıldı. Destek Vektör Makineleri ve önerilen özellik vektörünün kullanımı ile erkek için %89.4 doğru sınıflandırma oranı; kadın için %93.6 doğru sınıflandırma oranı ve cinsiyetten bağımsız durumlar için de %88.9 doğru sınıflandırma oranı elde edildi.

Ayrık Gizli Markov Model ses tanıma sistemleri için yeni bir Vektör Uzaklık Ölçümü yaklaşımı önerildi (Debyeche vd. 2006). Bu yeni yaklaşım Gizli Markov Model durumları üzerinde Vektör Uzaklık Ölçümü (VQ) kod kitabı bileşenlerinin optimal olarak dağıtılmasını gerçekleştiriyordu. Bu bağlamda bu yeni yöntem Dağıtılmış Vektör Uzaklık Ölçümü (D-VQ) olarak isimlendirildi.

Gizli Markov Model tabanlı ses tanıma için Destek Vektör Makineleri ile bir karma yaklaşım önerildi (Krüger vd. 2006). Bu çalışmada bir Gizli Markov Model tabanlı ses tanıma sistemi içerisine Destek Vektör Makineleri entegre edildi ve sürekli ses tanıma ve örüntü tanıma için başarılı sonuçlar elde edildi.

Yapay Sinir Ağları (ANN) ve mantık sistemlerinin birleşimi sonucu oluşan sinirsel-mantık sistemleri ile bir ses tanıma çalışması gerçekleştirildi (Lam vd. 2007). Bu çalışmada kullanılan temel bileşenler, Sinir Ağları, ‘VE’, ‘VEYA’, ‘DEĞİL’ mantık kapıları idi. Çalışma ‘Cantonese’ konuşma komutları üzerinde denendi. ‘Cantonese’ konuşma komutları ortalama olarak %90 doğruluk oranı ile tanındı.

Şablon-tabanlı sürekli ses tanıma araştırma ve çalışması gerçekleştirildi (Wachter vd. 2007).

Bir konuşmacı tanıma problemi üzerinde Gaussian Karma Modelleri (GMMs) için yeni bir yaklaşım önerildi (Vogt vd. 2007).

Dinamik Zaman Uydurma ve Gizli Markov Modellerin birleşimi oluşturularak bu yeni yaklaşım 'Malay' ses tanıma problemi üzerinde uygulandı (Al-Haddad vd. 2008). Çalışmanın amacı 'Malay' ses tanıma doğruluk yüzdesini arttırmaktı. Son-nokta tarama, çerçeveleme, normalizasyon, Mel Frekans Cepstrum Katsayıları (MFCC), Vektör Uzaklık Ölçümü (VQ) teknikleri, tanıma işleminde kullanılacak ses örneklerini işlemek için kullanıldı. Ses örüntülerinin tanınması aşamasında da Dinamik Zaman Uydurma ve Gizli Markov Modellerin birleşimi ile elde edilen yeni yaklaşım kullanıldı. Örüntülerin %94 doğruluk oranı ile tanındığı gözlemlendi. Oysa, bu iki yöntemin ayrı ayrı kullanımı durumunda, Dinamik Zaman Uydurma ile ses örüntüleri %80.5 doğruluk oranı ile; Gizli Markov Modeller ile ses örüntüleri %90.7 doğruluk oranı ile tanındı.

2. PROBLEMİN TANIMI

Günümüzde insanlar arası iletişimde her geçen gün yeni şeyler ortaya çıkmakta ve buna paralel olarak multimedya araçlarının kullanımı da günden güne artmaktadır. İnsanlar arası iletişim sesli olarak çok hızlı bir şekilde gerçekleşmektedir. Bu amaçla ses ile iletişim ve ses ile kontrol alanlarında önemli çalışmalar yapılmaktadır. İnsan-makine arası iletişimin ses ile gerçekleştirilmesi birçok zorlukları kolaylaştıracaktır. Ses işleme'nin alt bilim dallarından birisi olan ses tanıma biliminin uğraş alanı da insan-makine arası iletişimi gerçekleştirecek etkili ve başarılı sistemler veya yöntemler ortaya çıkarmaktır. Bu yüzden bu alanda uzun zamandan beri birçok araştırma ve çalışma yapılmış ve yapılmaya devam etmektedir.

Bu tez çalışmasının amacı da bilgisayar ortamında sıklıkla kullanılan 20 adet komutu seslendirerek, sistemi ses ile kontrol etmeye çalışmaktır. Bu, ayırık ses tanıma olarak isimlendirilen bir çalışma alanıdır. Bu ses tanıma çalışması Matlab yazılımı ile geliştirilen uygulama vasıtasıyla yapılmaya çalışılmıştır.

Geliştirilen uygulama iki ana modül üzerine inşa edilmiştir. Bunlardan ilki ses örüntülerinin kodlanması işlevini gerçekleştiren ses kodlama modülü; ikincisi de kodlanmış ses örüntülerini tanıma işlevini gerçekleştiren ses tanıma modülüdür.

Sesin kodlanmasına ihtiyaç duyulmasının sebeplerinden biri, sese ilişkin özellikleri çıkararak ses tanıma modülüne güvenilir giriş veri kümeleri sağlamaktır. Böylece tanıma başarısı daha iyi olacaktır. Çünkü ses sinyalinden alınan örnekler hem veri sayısı bakımından fazladır, hem de ses sinyalinden sesin özelliklerinin çıkarılması ses tanıma sistemlerinde daha başarılı sonuçlar doğurur.

Ayrıca diğer bir kodlama gereksinimi de daha az sayıda veri ile ses sinyalini temsil eden bir parametrik ifade elde etmektir.

Sesin tanınması aşamasında ise bir önceki adımda elde edilen kodlanmış ses sinyali yani sesin özellikleri, tanıma sisteminde giriş verileri olarak kullanılır. Burada da sistemin tanıma başarısını değiştirebilen çeşitli faktörler vardır. Örneğin, tanıma için kullanılacak yöntem içerisinde yer alan çeşitli parametre değerlerinin değiştirilmesi gibi. Parametre taraması yapılarak en iyi parametreleri bulmaktaki amaç sistemin tanıma başarısını arttırmaktır.

Ayrıca kodlama yöntemlerinin, algoritmalarının geliştirilmesi ve daha başarılı kodlama sistemlerinin ortaya çıkarılması da ses tanıma sistemlerinin tanıma başarılarını arttıracaktır.

3. SES

3.1. Sesin Oluşumu ve Özellikleri

Ses, akciğerlerden gelen havanın oluşturduğu titreşimler, duyma organları tarafından algılanan fiziksel niceliklerdir ve dili oluşturan en küçük birimdir. Harf ise bu sesleri gösteren sembollerdir ve doğal olarak bir ses için her alfabede farklı semboller kullanılabilir (Nabiyev 2005).

Ses yolu, nefes borusu, gırtlak, ağız boşluğu ve burun yolundan oluşan bir geçittir. Ses yolunun belirli yerlerinde ses telleri denilen yarım daire biçiminde iki dudakçık ile küçük dil, damak, dişler, dudaklar gibi organlar bulunur. Bu organların birbirine ya da komşu bir organa yaklaşıp uzaklaşması ile ses yolu açılıp kapanır, daralır genişler.

İnsan sesinin oluşumunda başlıca dört olay yer alır (Nabiyev 2005). Bunlar sırasıyla akciğerlerden havanın pompalanması aşaması, titreşim, rezonans ve artikülasyondur. Ham sesin oluşumu akciğerlerden başlar. Titreşim larenkste ses tellerinde oluşur ve ham ses meydana gelir. Farenks, ağız, burun boşlukları ve paranasal sinüsler rezonans bölgelerini oluştururlar ve sesin güçlendirilmesinde görev alırlar. Dil, dişler, dudak ve damak ise artikülasyon organları olarak güçlendirilmiş ham sesi işleyerek anlaşılabilir tanınan bir fonem haline getirirler. İnsanda ses oluşumu organları şekil 3.1'de görülmektedir.

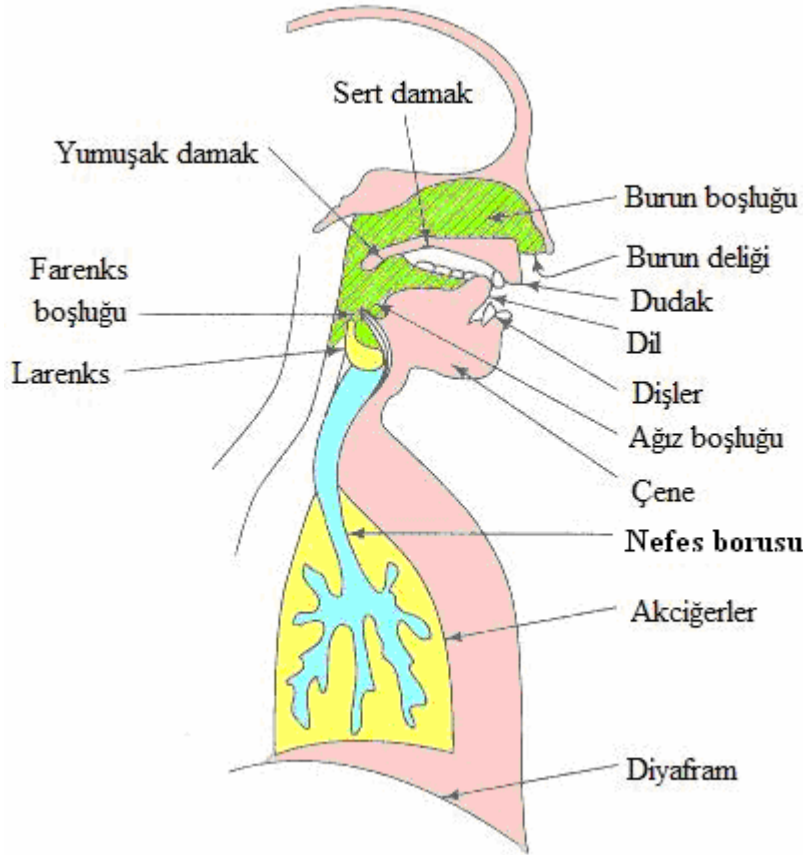
Sesbilimde rezonans ve artikülasyon, boğumlanma biçiminde birleştirilir ve sesin oluşumu için genellikle üç aşama ele alınır.

Nefes verme – Bir körük gibi işleyen akciğerler, havayı yeterince sıkıştırarak nefes borusundan ağza doğru iterler. Bunun solunumdan farklılığı ise isteme ve duruma bağlı olmasıdır.

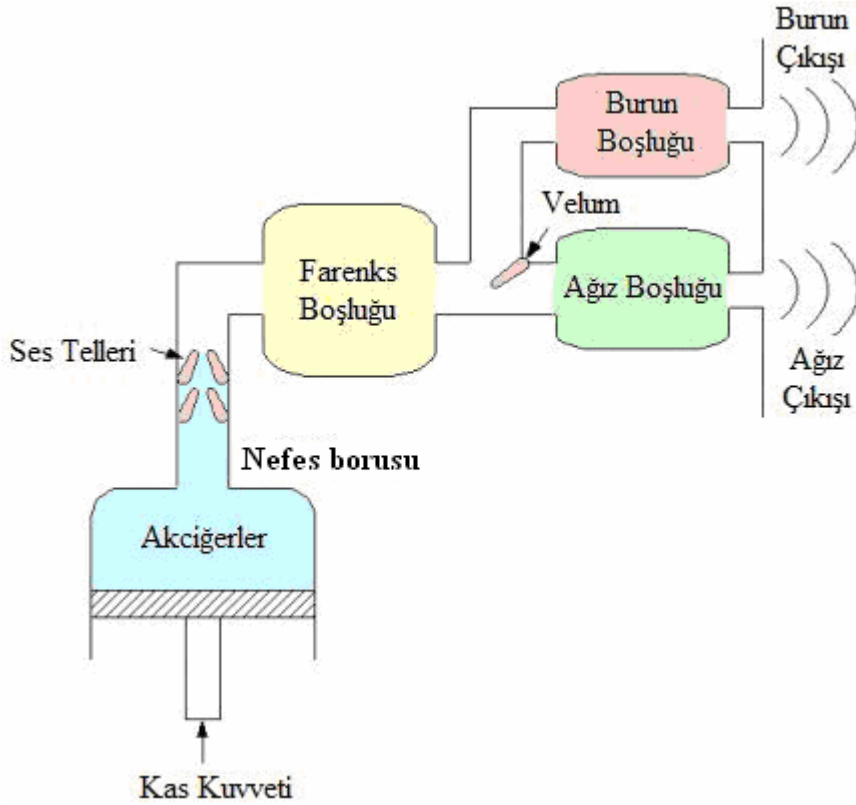
Selenleşme – ağza doğru itilen hava, nefes borusunun sonundaki gırtlakta ses telleri ile karşılaşır; bunlar havanın itişiyile birbirinden az ya da çok ayrılırken, titreşirler. Bunların titreşimi, havayı titreştirir. Böylece hava selen durumuna gelir. Sesbilimde ses, gırtlakta ses tellerinin titreşimi ile oluşan biçimlenmiş yalın selendir.

Boğumlanma – selenleşen hava, ses yolunda devam ederken ağız boşluğu ile burun geçidinin kesiştiği yere gelir. Burada küçük dil bulunmaktadır. Küçük dil, selenleşen havanın istenilen biçimi alabilmesi için ya burun geçidini kapatır ya da büyük dil ile birlikte ağız geçidini kapatıp burun geçidine yol verir. Selenleşen hava, burun geçidine giderse geniş sesleri oluşturur; ağız boşluğuna giderse büyük dilin kamburlaşması, yayılması, damağa ya da dişlere dokunup çekilmesi vb. sonucunda boğumlanır, yani bir engelle karşılaşınca istenilen biçimi alır.

İnsanda ses oluşumunun şematik gösterimi şekil 3.2’de görülmektedir.



Şekil 3.1 Ses oluşumu organları



Şekil 3.2 İnsanda ses oluşumu mekanizmasının şematik gösterimi

Anlaşıldığı gibi, sesin işlenmesinde bu organların uygun modellenmesi yapılmalıdır. Aksi takdirde sesin tam anlamıyla tanınması veya yapay ses üretimi mümkün olmamaktadır.

İnsan sesi temel ton ve ikincil tonlardan veya harmoniklerden oluşmaktadır. Bir titreşimin frekansı, hareketin kendi kendine özdeş olarak, 1 saniyedeki yinelenme sayısıdır. Saniyedeki bu titreşim sayısı Hz (hertz) ile ölçülmektedir. Sesin bir diğer özelliği ise şiddetidir. Sesin şiddeti onun enerjisi ile ilişkilidir. Sesin şiddeti genellikle logaritmik bir birim olan desibel (dB) ile ölçülür.

3.2. Sesin Sentezi ve Analizi

Ses işlemede genellikle sesin sentezi ve analizi üzerinde durulmaktadır. Sesin analizi aşamasında mikrofona girilen ses sinyallerinin sesin anlamsal özellikleri korunarak daha az yoğunluklu biçime dönüştürülmesine yani sesin sayısallaştırılmasına çalışılmaktadır. Sıkıştırılmış ve kodlanmış verilere göre taban bilgiler ile karşılaştırmalar yapılarak tanıma gerçekleştirilmektedir.

Ses analizatörleri (çözümleyiciler), sınırlı sözcükle işleyenler ve genel amaçlı olanlar şeklinde ikiye ayrılmaktadır. Sınırlı sözcükle işleyen sistemlerde en çok uygunluk bulunan özelliklere göre tanıma gerçekleştirilmektedir. Girilen sesli bilginin tanınması için bilgisayarın hafızasında uygun kelimelerin şablon şekilleri tutulmaktadır. Konuşma sonucunda algılanan kelime taban bilgilerle değerlendirilerek en yakın uygunluğa göre aday kelime bulunur. Ayrıca tanımanın gerçekleştirilmesinde uygunluk grafi üzerinde minimum yolun hesaplanmasına dayalı olan dinamik programlama yöntemi de kullanılmaktadır. Genellikle bu tip analizatörler için bilgi tabanı, sözcüğün bir konuşmacı tarafından sisteme girilerek öğretilmesi ile gerçekleştirilmektedir. Bu tip analizatörlerin bir diğer sınırlaması ise ses komutlarının birbirinden ayrık biçimde girilmesidir. Genellikle kelimeler arası sessizlik aralığı 0.3-1 saniye olarak sınırlandırılmaktadır. Genel amaçlı analizatörlerin tasarımında ise, bu tür analizatörlerin istenen kelimeyi veya sürekli ses sinyalini işleme yeteneğine sahip olması amaçlanmaktadır.

Ses sentezleyicilerinde ise bilgisayarda tutulan sayısal kod ardışıklığının sürekli ses sinyaline dönüştürülmesine çalışılmaktadır.

3.2.1. Sesin analizi

Sesin analizinde sesin parametrik bir ifadeye dönüştürülmesine çalışılmaktadır. Bu parametrik ifade ya da diğer ifadeyle ses sinyalinin analiz edilmesi sonucu ortaya çıkan özellik vektörleri ses tanıma sistemlerinde giriş veri kümesi olarak kullanılmaktadır. Ses tanıma sistemleri şu kategorilere ayrılabilir:

- Konuşmacı tanıma sistemleri
- Konuşmacıya bağımlı ses tanıma sistemleri
- Konuşmacıdan bağımsız ses tanıma sistemleri
- Anahtar sözcük yakalama sistemleri
- Sürekli ses tanıma sistemleri

Sesin analizinde sesin sayısallaştırılması, özellik vektörlerinin çıkarılması, sesin spektral analizi, sesin Hızlı Fourier Dönüşümü gibi işlemler yapılmaktadır.

3.2.1.1. Sesin sayısallaştırılması

Bu aşamada analog ses sinyali sayısal ses sinyaline dönüştürülür. Bu dönüşüm sırasıyla, Örnekleme (Sampling), Kuantalama (Quantize) ve Kodlama (Coding) aşamalarını içermektedir.

3.2.1.2. Özellik vektörünün çıkarılması

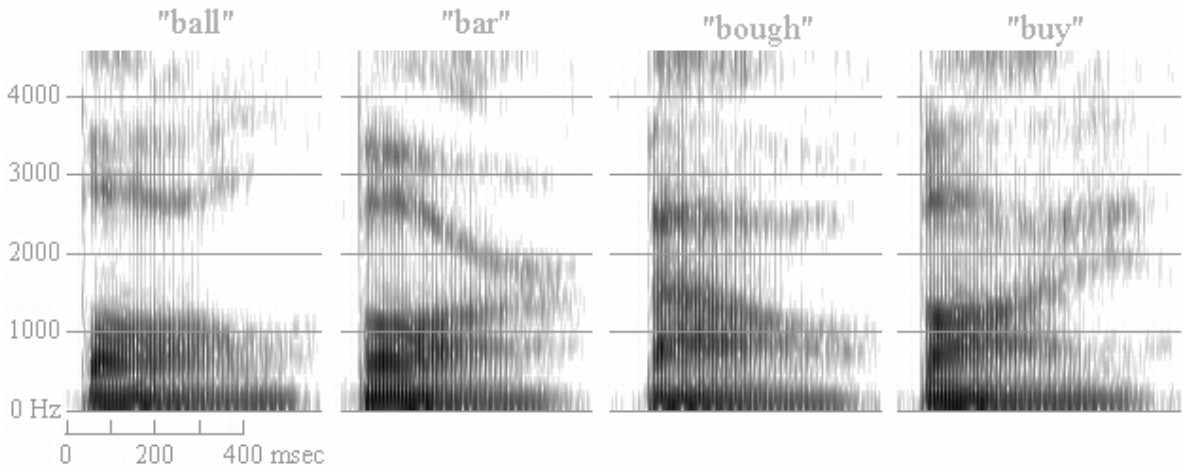
Ses tanıma, örüntü tanıma sistemlerinin özel bir uygulama alanıdır. Örüntü tanıma genellikle önce sistemin eğitilmesi (training) ve ardından eğitilmiş sistemin test edilmesi (testing) işlemleri gerçekleştirilir. Eğitim aşamasında giriş örüntüleri incelenerek özellikler çıkarılır. Bu özellikler bir kütüphane bilgisi oluşturur. Tanıma aşamasında kütüphanedeki bütün özellikler ile bilinmeyen örüntünün özellikleri karşılaştırılır. Özelliklerin belirlenmesi diğer ifadeyle özellik çıkarımı, örüntü tanımanın en önemli parçasıdır. İnsan kulağının ses işareti karşısında duyarlılığı incelendiğinde, sesin güç spektrumuna (tayfına) karşı duyarlı olduğu görülmüştür. Bu nedenle ses işareti analiz edilirken genellikle özellik çıkarımı için sesin spektral analizi yapılır.

3.2.1.3. Sesin spektral analizi

Sesin spektral analiz yöntemleri parametrik analiz ve parametrik-olmayan analiz olmak üzere iki gruba ayrılmaktadır. Parametrik olanlar, Doğrusal Öncestirim Kodlama ve Maksimum Benzerlik Yöntemi'dir. Parametrik olmayanlar da, Kısa Zamanlı Otokorelasyon (Short-Time Autocorrelation) ve Sıfır-Geçiş Analizi (Zero-Crossing Analysis)'dir.

Basit olarak ses üretim mekanizması, istenen ses için bir kaynak tarafından uyarılan akustik bir tüp ve ses yolunu içermektedir. Örneğin, sessiz harfler, ses yolunun kasılması ile daralan boşluktan geçen havanın oluşturduğu hava akımı ile elde edilir. Elde edilen sesin önemli bilgi kaybı olmaksızın iki boyutlu bir görüntüsü elde edilebilmektedir. Ses spektrogramı denilen bu görüntüde yatay eksen zaman dilimini oluşturmakta ve dikey (düşey) eksene de frekanslar yerleştirilmektedir. Herhangi bir zaman diliminde belirli frekanslı ses enerjileri uygun bölgenin işaretlenmesi ile spektrogramda görüntülenir. Gerçek zamanda kullanılan spektrogramlar genellikle işitme engelli insanların ekranda sesin görüntüsünü izleyerek konuşmaya katılımını sağlamak amacıyla da kullanılmaktadır. Konuşulan kelimelerin tanınmasında bu veya

diğer zaman-frekans diliminde büyük enerjili parçaların bulunması izlenmektedir. Buradaki zorluk, zamanla ilişkili olarak kelimelerin esnek olmasıdır ve farklı süreklilikte telaffuz edilen aynı kelimenin değişkenliğidir. Fakat frekans alanında (bölgesinde) bu esneklik görülmemektedir ve ünlü seslere kesin formant değerleri karşılık gelmektedir. Burada sözü edilen formant, ses dalgasının enerji spektrumu içindeki tepe noktalarına karşılık düşer, diğer ifadeyle ses tüpü içindeki havanın temel titreşim frekanslarına da *Formant Frekansları* denir. Ses tellerinin titreşim frekansına ise sesin *Pitch*'i denmektedir.



Şekil 3.3 'ball', 'bar', 'bough', 'buy' kelimelerinin spektrogramı (Hagiwara 2008)

Spektral analiz yaparken belirlenmesi gerekenler, hangi yöntemin kullanılacağı, örnekleme hızları, analiz aralığının seçimi, zaman pencereleri ve önvurgulama'dır.

3.2.1.3.1. Yöntemin seçimi

Spektral kestirim için otokorelasyon metodu genellikle yaygın ve genel bir metod olarak kullanılır, öyle ki ses dalgasından anlamlı sonuçlar çıkarmak için analiz penceresi, içerisinde birkaç pitch periyodu içermelidir. Diğer taraftan kovaryans metodu ise bir pitch periyodundan daha küçük bir aralıkta uygulanmaktadır dolayısıyla bu yöntem yerel bir metod olarak kullanılmaktadır. Birkaç pitch periyodu analiz edildiği zaman her iki yönteme ilişkin spektral sonuçlar birbirine oldukça benzer olmaktadır. Yani örnek sayısı yeterince büyük olduğu zaman kovaryans katsayıları, otokorelasyon katsayılarına yaklaşık olarak eşit olmaktadır. Ayrıca otokorelasyon yöntemi daha az hesaplama gerektirir. Bu yüzden uygulamalarda, otokorelasyon yöntemi daha çok tercih edilmektedir.

3.2.1.3.2. Örnekleme hızları

Örnekleme hızı arttırıldığı zaman sürekli ses sinyalinin temsili daha doğru olmaktadır. Fakat daha çok örnek, daha büyük depolama alanı gereksinimi doğurmakta ve daha çok hesaplama neden olmaktadır.

3.2.1.3.3. Analiz aralığının seçimi

Analiz aralığının seçimi iki faktörü içerir, bunlar; aralığın yeri ve aralığın uzunluğudur.

3.2.1.3.4. Pencereleme

Pencereleme, ses sinyalinin başındaki ve sonundaki süresiz bölümleri minimize etmek için, ses sinyalinin bir pencere fonksiyonu ile çarpılmasıdır. Pencere fonksiyonlarına örnek olarak Dikdörtgen veya Hamming pencere verilebilir.

Dikdörtgen pencere;

$$w(n) = \begin{cases} 1, & 0 \leq n \leq N-1 \\ 0, & n < 0, n > N-1 \end{cases} \quad (3.1)$$

şeklinde tanımlanır (Eşitlik 3.1).

Hamming penceresi ise;

$$w(n) = \begin{cases} 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), & 0 \leq n \leq N-1 \\ 0, & n < 0, n > N-1 \end{cases} \quad (3.2)$$

şeklinde tanımlanır (Eşitlik 3.2). Genel bir kural olarak uygulamalarda doğrusal önkestirim yaparken aralık uzunluğu 15 ms'den büyük seçilmektedir veya pencere birkaç pitch periyodu içerecek şekilde pencere uzunluğu belirlenmektedir.

3.2.1.3.5. Önvurgulama

Eğer ses tüpünün spektral özellikleri glottal dalga biçiminin sonuçları ve dudak yayılma parametreleri olmaksızın kestirilecek ise ses sinyaline önvurgulama işlemi uygulanmalıdır. Önvurgulama, ses sinyalini bir basit bir-sıfır filtresinden geçirerek gerçekleştirilebilir. Buradaki bir-sıfır filtresi, $1 - \mu.z^{-1}$ ile karakterize edilir ise, μ parametresi 1 olarak seçilebilir.

3.2.1.4. Hızlı Fourier Dönüşümü (FFT)

Mikrofona gelen ses gürültüler içermektedir. İşlenecek ses işaretinde genellikle beyaz gürültüye rastlanmaktadır. Beyaz gürültünün frekansı ana bileşene göre çok yüksektir. Doğadaki işaretlerin tümü sonsuz sinüs ve kosinüs toplamları biçiminde ifade edilebilmektedir. Matematiksel olarak periyodik $x(t)$ fonksiyonu sonsuz sayıda \sin ve \cos fonksiyonlarının toplamı biçiminde fourier serisi yardımıyla gösterilir (Eşitlik 3.3);

$$x(t) = c_0 + \sum_{k=0}^{\infty} (a_k \cos(k\omega_1 t) + b_k \sin(k\omega_1 t)) \quad (3.3)$$

Burada $x(t)$ zamanla belirlenmiş ayrık işareti, c_0 , a_k , b_k ise fourier katsayılarını ifade etmektedir. $\omega_1 = 2\pi f_0$ 'a eşit olmakta ve f_0 ise temel frekans olarak adlandırılmaktadır. Sayısal işaret işlemede ayrık zaman işaretleri frekans alanında ifade edilir ve işaretin frekans bandı hakkında bilgi edinilir. Fourier dönüşümü, Ayrık Fourier Dönüşümü (DFT) ve bunun hesaplanmasını sağlayan Hızlı Fourier Dönüşümü (FFT) biçiminde kullanılmaktadır. İşaretin FFT'si alındıktan sonra bir alçak geçiren süzgeç (filtre) kullanılır ve böylece işaretin yüksek frekanslı gürültü bileşenlerini ayırmak mümkün olabilmektedir. Fourier dönüşümü ile frekans alanına (bölgesine) taşınmış veriler ters fourier dönüşümü ile tekrar zaman bölgesine aktarılabilmektedir.

İşaretin Hızlı Fourier Dönüşümü;

$$X_n = \sum_{k=0}^{N-1} x_k e^{-2\pi jkn/N}, \quad n = 0, 1, \dots, N-1 \quad (3.4)$$

3.4 eşitliđi ile tanımlanmaktadır.

3.2.1.5. Dilimleme (Segmentation)

Dilimleme işlemi cümledeki kelimeleri ayırma işlemidir. Konuşulan kelimenin başının ve sonunun bulunması başlıca bir problemdir. Bu probleme ilişkin varolan çözümlerin çođu, cümledeki kelimelerin bitiş noktalarını tahmin etmeye çalışmaktadır.

3.2.1.6. Zaman normalizasyonu (Time normalization)

Ses tanımada diđer bir problem, zaman normalizasyonudur. Bir kelimenin sürekliliđinde ve zaman dağılımında farklılıklar bulunmaktadır. Bu farklılıklar, yalnızca deđişik kişiler için deđil, aynı kişiden farklı zamanlarda alınan örnekler için de geçerlidir. Kütüphane bilgileri ile tanıma sistemine girdi olarak alınan kelime arasında zamansal sıraya koyma çok önemlidir. Zamansal sıraya koyma için iki yöntem sıkça kullanılmaktadır. Bunlar, Dinamik Zaman Uydurma (Dynamic Time Warping-DTW) ve ses verisinin olasılıklara dayalı Markov Modelinin kullanımüdür.

4. SES KODLAMA

Ses sinyalini parametrik olarak temsil etme için kullanılabilir çeşitli yöntemler vardır (Rabiner vd. 1993). Bu yöntemler ses sinyaline ilişkin, kısa zaman enerjisi, sıfır geçiş hızları, seviye geçiş hızları ve diğer ilgili parametreleri içerebilir. Ses sinyalinin parametrik olarak temsilinde en önemli olanı *Kısa Zamanlı Spektral Analiz*'dir. Bu bölümde ses sinyalinin parametrik olarak temsil edilmesi, yani ses sinyaline ilişkin özelliklerin çıkarımı için geliştirilmiş yöntemlerden en çok kullanılanları açıklanacaktır. Bu yöntemler şunlardır:

- Doğrusal Önkestirim Kodlama (LPC)
- Doğrusal Önkestirim Cepstrum Katsayıları (LPCC)
- Mel Frekansı Cepstrum Katsayıları (MFCC)
- Algısal Doğrusal Önkestirim Kodlama (PLP)

4.1. Doğrusal Önkestirim Kodlama (LPC)

Doğrusal Önkestirim Kodlama yönteminde, n zamanda verilen $s(n)$ ses örnekleri, önceki p tane ses örneğinden yaklaşık olarak eşitlik 4.1'deki gibi elde edilebilir.

Bu yöntem insan gırtlığı ve ağız yapısı özelliklerinin yanı sıra ses özelliklerini de dikkate alır. Doğrusal Önkestirim Kodlama yöntemi, sesin periyodik uyarma (dürtü) ve rasgele gürültü ile uyarılan, doğrusal ve zamana göre değişen bir sistemin çıktısı ile modellenebileceği prensibine dayanır (Baygün 2006). LPC, doğrusal bir filtre olarak eşitlik 4.2 ile ifade edilmektedir.

$$s(n) = a_1 s(n-1) + a_2 s(n-2) + \dots + a_p s(n-p)$$

$$s(n) = \sum_{i=1}^p a_i s(n-i) \quad (4.1)$$

Buradan ses sinyalinin yeniden ifade edersek,

$$s(n) = \sum_{i=1}^p a_i s(n-i) + Gu(n) \quad (4.2)$$

eşitliği elde edilmiş olur. Burada $u(n)$ normalize edilmiş uyarma fonksiyonu, G ise uyarma kazancıdır. Eşitlik 4.2'yi z - uzayında yeniden ifade edersek eşitlik 4.3 elde edilir.

$$S(z) = \sum_{i=1}^p a_i z^{-i} S(z) + GU(z) \quad (4.3)$$

Buradan transfer fonksiyonunu çekersek,

$$H(z) = \frac{S(z)}{GU(z)} = \frac{1}{1 - \sum_{i=1}^p a_i z^{-i}} \quad (4.4)$$

eşitliği elde edilir. Eşitlik 4.4'e ters z -dönüşümü uygulandığında eşitlik 4.2 elde edilir.

Sonuç olarak denilebilir ki, LPC, sıradaki örneğin, önceki bir dizi örnekten elde edilebileceği prensibi ile çalışır.

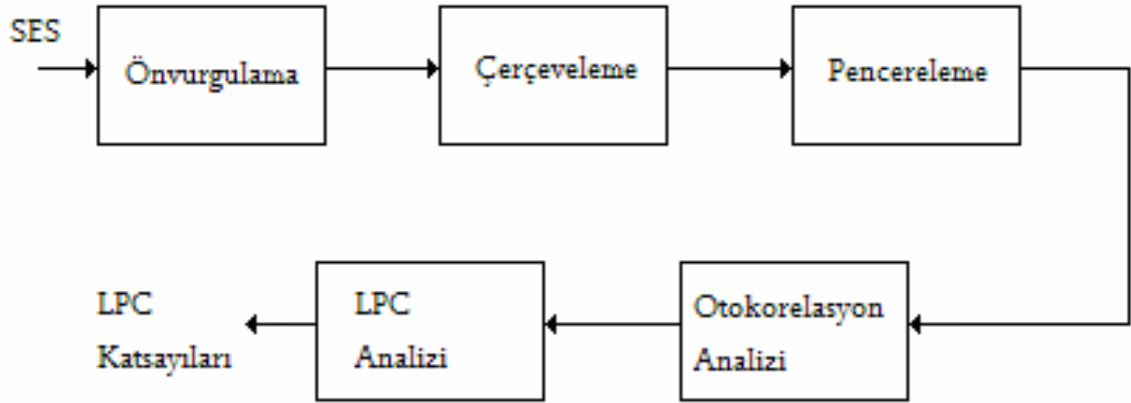
Tahmin sonucu elde edilen örneğin asıl örnekle olan farkının; yani hataların kareleri toplamının minimize edilmesi için bir dizi parametre hesaplanır.

Doğrusal Önkestirim Kodlama yöntemi ile p sayıda parametre hesaplanır. Burada p LPC kodlayıcının mertebesi (derecesi), a_1, a_2, \dots, a_p ise LPC katsayıları olarak ifade edilmektedir.

LPC parametreleri LPC kodlayıcının çıkışı olup, bu çıkış LPC kodlayıcının girişine verilen bir çerçeve örneğe karşılık gelen çıktıdır.

4.1.1. LPC yöntemindeki aşamalar

LPC yönteminin aşamaları şekil 4.1’de görülmektedir.



Şekil 4.1 Doğrusal Önkestirim Kodlama yönteminin blok diyagram olarak gösterimi

4.1.1.1. Önvgulama

Sayılaştırılmış ses sinyali düşük dereceli bir sayısal sistemden (genellikle bir birinci-dereceden FIR filtreden) geçirilir. Buna *Önvgulama* denmektedir.

4.1.1.2. Çerçeveleme

Bu aşamada ses sinyali üzerinde kısa zamanlı spektral analiz yapabilmek için sürekli ses sinyali, N adet örnekten oluşan çerçevelere bölünür. İlk çerçeve N örnekten oluşurken ikinci çerçeve ilk çerçeveden M örnek sonra başlar ve ilk çerçevenin $N-M$ örnek kadar üzerine biner. Aynı şekilde üçüncü çerçeve ilk çerçeveden $2M$ örnek, ikinci çerçeveden M örnek sonra başlar ve ilk çerçevenin $N-2M$ örnek kadar üstüne biner. Bu işlem tüm ses sinyali boyunca yapılır. Şekil 4.2’de çerçeveleme işlemi görülmektedir.

4.1.1.3. Pencereleme

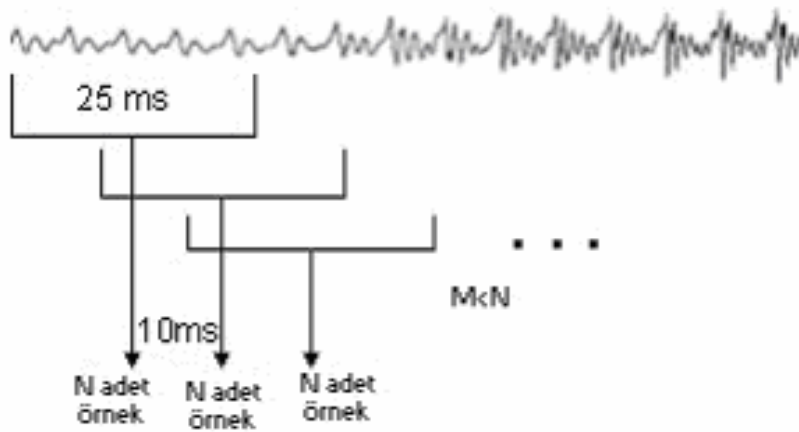
Sıradaki işlem, her çerçeveyi pencereleyerek, sinyalin başındaki ve sonundaki süreksiz bölümleri minimize etmektir. Buradaki amaç, pencereyi kullanarak çerçevelerin başı ve sonundaki bilgi içermeyen bölümleri kırmak, dolayısıyla spektral

bozulmayı engellemektir. Eğer pencereyi $w(n)$, $0 \leq n \leq N - 1$, her çerçevedeki örnek sayısını da N olarak kabul edersek, sinyalin pencerelenmiş hali (Eşitlik 4.5);

$$y(n) = x(n).w(n) \quad (4.5)$$

olur. Pencereleme çeşitleri, Hamming, Hanning, Dikdörtgen ve Blackman pencerelemedir. Yapılan uygulamada Hamming Pencerelemesi kullanıldı. Hamming pencerelemesinin formülasyonu şu şekildedir (Eşitlik 4.6);

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \quad (4.6)$$



Şekil 4.2 Çerçeveleme

4.1.1.4. Otokorelasyon analizi

Pencerelenmiş sinyalin her bir çerçevesine otokorelasyon analizi uygulanır. Burada p LPC analizinin derecesidir. Genellikle p , 8 ile 16 arasında bir değer seçilir. Yapılan uygulamada $p = 8$ seçildi. Otokorelasyon;

$$r_l(m) = \sum_{n=0}^{N-1-m} \hat{x}_l(n) \cdot \hat{x}_l(n+m), \quad m = 0, 1, \dots, p \quad (4.7)$$

şeklinde tanımlanır (Eşitlik 4.7). Otokorelasyon analizinin bir yararlı tarafı da, sıfıncı otokorelasyonun ilgili çerçevenin enerjisini ifade ediyor olmasıdır. Bir çerçevenin enerjisi ses tanıma sistemleri için önemli bir parametredir.

4.1.1.5. LPC analizi

Bu bölümde her bir çerçeveye ait $p+1$ otokorelasyondan LPC parametre kümesi hesaplanır. LPC parametre kümesi LPC katsayılarından, PARCOR katsayılarından, cepstral katsayılardan veya yukarıdaki parametre kümelerinin istenen herhangi bir dönüşümü sonucu elde edilen katsayılardan oluşabilir. Genellikle LPC analizi sonucunda elde edilen LPC katsayıları LPC parametre kümesi olarak sıkça kullanılmaktadır.

Otokorelasyon analizinden LPC analizine geçişte Durbin metodu gibi bir yöntem kullanılabilir.

LPC analizine geçişte otokorelasyon analizinin yerine kovaryans analizi de yapılabilir, fakat genellikle yaygın olarak kullanılan otokorelasyon analizidir.

Sonuç olarak elde edilen LPC parametreleri, a_m LPC katsayılarından oluşmaktadır (Rabiner vd. 1993).

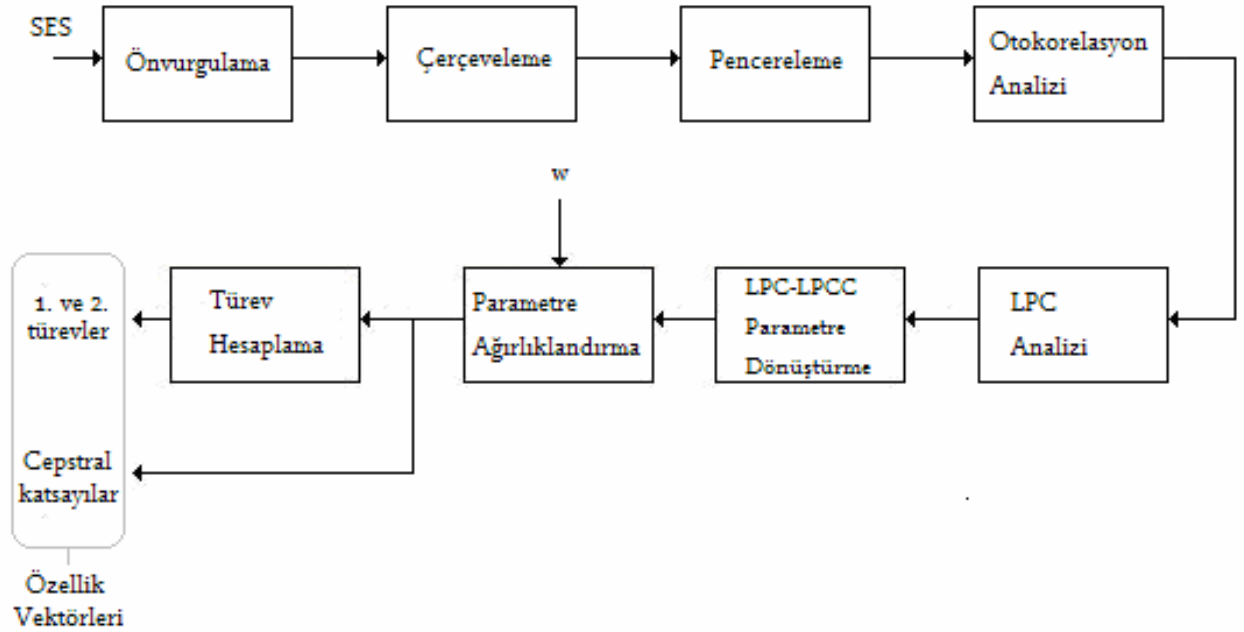
4.2. Doğrusal Önkestirim Cepstrum Katsayıları (LPCC)

Doğrusal Önkestirim Cepstral Katsayılar (LPCC) yöntemi temel olarak LPC katsayılarının Fourier dönüşümü ile cepstral katsayılara dönüştürülmesi prensibine dayanmaktadır. Dolayısıyla bu yöntemde ilk olarak LPC katsayılarının elde edilmesi ve ardından diğer işlemlerin yapılması gerekmektedir. LPC katsayılarının elde edilmesi işlemi Bölüm 4.1’de anlatıldığı için bu bölümde önvurgulama, çerçeveleme, pencereleme, otokorelasyon analizi, LPC analizi işlemleri tekrar anlatılmayacaktır. LPCC’deki aşamalar şekil 4.3’te görülmektedir.

LPCC yönteminin aşamaları Önvurgulama, Çerçeveleme, Pencereleme, Otokorelasyon analizi, LPC analizi, LPC-LPCC parametre dönüştürme, Parametre ağırlıklandırma, Türev hesaplama’dır. Sonuç olarak elde edilen parametrik ifadeler *Özellik vektörleri* olarak adlandırılmaktadır.

4.2.1. LPCC yöntemindeki aşamalar

LPCC yönteminin aşamaları şekil 4.3'te görülmektedir.



Şekil 4.3 Doğrusal Önkestirim Cepstrum Katsayıları yönteminin blok diyagramı

4.2.1.1. Önvgulama

Bölüm 4.1'de anlatılmıştır.

4.2.1.2. Çerçeveleme

Bölüm 4.1'de anlatılmıştır.

4.2.1.3. Pencereleme

Bölüm 4.1'de anlatılmıştır.

4.2.1.4. Otokorelasyon analizi

Bölüm 4.1'de anlatılmıştır.

4.2.1.5. LPC analizi

Bölüm 4.1'de anlatılmıştır.

4.2.1.6. LPC-LPCC parametre dönüşümü

Bu aşamada Fourier dönüşümü ile LPC parametre kümesinden LPCC'a dönüşüm gerçekleştirilir. Genellikle elde edilen cepstral katsayılar ses tanıma sistemleri için özellik kümesi olarak daha güvenilir ve sağlam olmaktadır.

4.2.1.7. Parametre ağırlıklandırma

Bu aşamada önceki adımda elde edilen cepstral katsayıların gürültü vb. gibi çeşitli etkenlere duyarlılığını minimize etmek amacıyla bu cepstral katsayılara bir ağırlıklandırma işlemi uygulanır. Buradaki ağırlıklandırma fonksiyonu cepstral uzayda bir bant geçiren filtre ile gerçekleştirilmektedir.

4.2.1.8. Türev hesaplama

Sesin spektrumunun cepstral temsili, verilen analiz çerçevesi için sinyalin yerel spektral özelliklerinin iyi bir temsilini sağlar (Rabiner vd. 1993).

4.3. Mel Frekansı Cepstrum Katsayıları (MFCC)

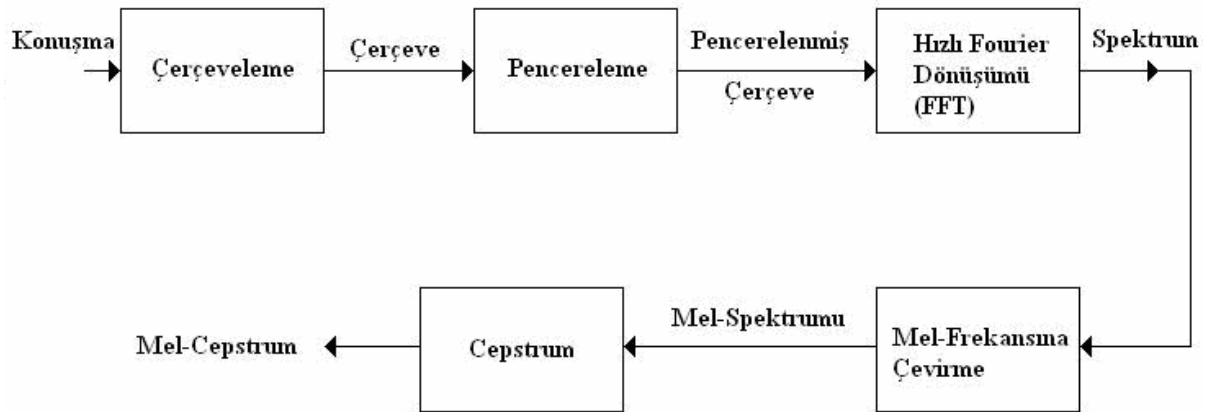
Mel ölçeklendirme, insan kulağının duyma özelliğine göre oluşturulmuş algısal bir ölçeklendirmedir. İnsan kulağı frekansları doğrusal olarak algılayamaz. Mel, algılanan ses sinyal frekansı ölçüm birimidir.

Referans olarak 1 KHz seçilmiş ve 1000 mel'e karşılık geldiği kabul edilmiştir (Uzunçarşılı 2005). Bu dönüşümün matematiksel formülasyonu (0-4 KHz arasında) şu şekilde tanımlanmıştır (Eşitlik 4.8);

$$M(f) = 2595 \cdot \log_{10} \left(1 + \frac{f}{700} \right) \quad (4.8)$$

4.3.1. MFCC yöntemindeki aşamalar

MFCC yönteminin aşamaları şekil 4.4'te görülmektedir.



Şekil 4.4 MFCC'nin blok diyagram olarak gösterimi (Demirci 2005)

4.3.1.1. Çerçeveleme

Bölüm 4.1'de anlatılmıştır.

4.3.1.2. Pencereleme

Bölüm 4.1'de anlatılmıştır.

4.3.1.3. Hızlı Fourier Dönüşümü (FFT)

Hızlı Fourier Dönüşümü (FFT), her çerçeveyi zaman bölgesinden frekans bölgesine taşır. FFT, kesikli (ayrık) fourier dönüşümünü hızlı bir şekilde yerine getirir. FFT sonrasında sinyalin FFT spektrumu elde edilmiş olur. Sinyalin Hızlı Fourier Dönüşümü,

$$X_n = \sum_{k=0}^{N-1} x_k e^{-2\pi jkn/N}, \quad n = 0, 1, \dots, N-1 \quad (4.9)$$

4.9 eşitliği ile tanımlanmaktadır.

4.3.1.4. Mel-frekansına çevirme

Sinyalin Mel spektrumunu elde etmek için FFT Spektrum, Mel-filtre bankası adı verilen bir dizi filtreden geçirilir. Mel-filtre bankasını bant-geçiren üçgen dalgalar

oluşturmaktadır. Hz türünden verilen frekans değerlerini Mel ölçeğine çevirmek için 4.8 eşitliği kullanılmaktadır.

4.3.1.4.1. Mel-filtre bankası tasarımı

Mel-Filtre Bankasını bant-geçiren üçgen dalgalar oluşturmaktadır. Elde edilmeye çalışılan katsayı sayısı kadar üçgen filtre kullanılır. Filtrelerin genişlikleri %50 oranında birbiri üstüne getirilir. Filtre bankasında üçgen bant geçiren dalgalar kullanılır. Genellikle 20 filtre kullanılır. Şekil 4.5'te Mel-filtre bankası görülmektedir.

4.3.1.5. Mel-cepstrum

Mel-filtre bankasından geçirilen sinyalin logaritması alındıktan sonra Kesikli (Ayrık) Kosinüs Dönüşümü (DCT) ile tekrar zaman bölgesine geçilir. Elde edilen katsayılara *Mel Frekans Cepstrum Katsayıları* (Mel Frequency Cepstrum Coefficients) denir. Ses spektrumunun cepstral gösterimi verilen sinyal çerçevesi için ses sinyalinin yerel spektral özelliklerinin iyi bir gösteriminin yapılmasını sağlar. Çünkü mel spektrum katsayıları reel sayılardır.

$$c(n) = \sqrt{\frac{2}{K}} \sum_{i=1}^K \left(\log \tilde{S}_i \right) \cos \left[n \left(i - \frac{1}{2} \right) \frac{\pi}{K} \right], \quad n = 0, 1, \dots, K - 1 \quad (4.10)$$

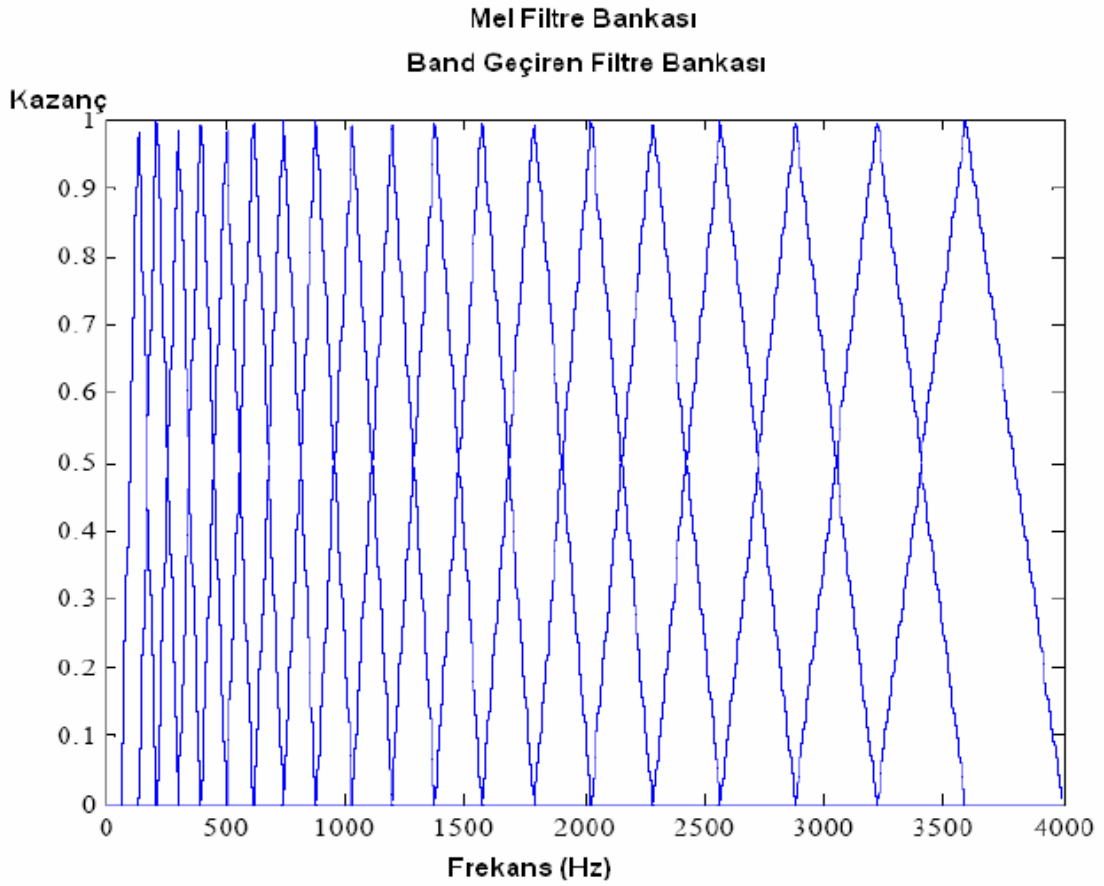
Eşitlik 4.10'da kullanılan, \tilde{S}_i mel spektrumu çıkış sinyalini ve K da mel filtre bankasında kullanılan bant geçiren üçgen filtre sayısını gösterir.

Böylece 30 ms uzunluğundaki birbiri üzerine bindirilmiş her ses çerçevesi için bir dizi MFCC hesaplanır.

4.4. Algısal Doğrusal Önkestirim Kodlama (PLP Coding)

LPC'nin bir varyasyonudur ve ilk olarak (1990) Hermansky tarafından ortaya atılmıştır. Bu yöntemdeki temel fikir, insan kulağının işitilebilir aralıkla ilgili, fiziki özelliklerinden türetilen bazı karakteristikleri dikkate almasıdır (Baygün 2006).

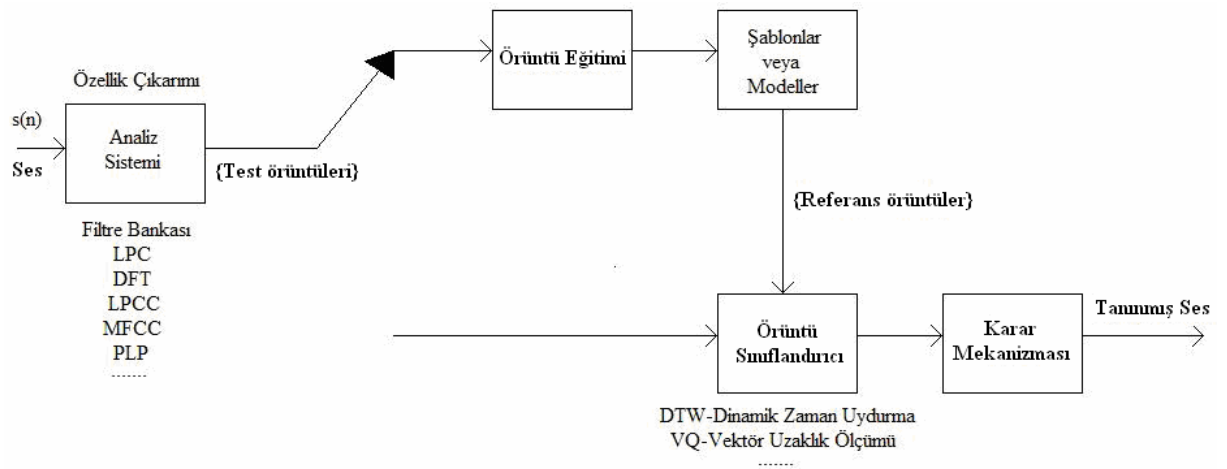
PLP'de de, LPC'de olduğu gibi bir dizi parametre hesaplanmaktadır. PLP parametreleri, DFT (Ayrık Fourier Dönüşümü) ve LP (Doğrusal Önkestirim) tekniklerinin birleştirilmesi ile hesaplanır.



Şekil 4.5 Mel filtre bankası (Uzunçarşılı 2005)

5. SES TANIMA

Ses tanıma teorisi oldukça geniş bir konudur ve genel olarak bir örüntü tanıma problemi olarak görülebilir (Rabiner vd. 1993). Şekil 5.1’de örüntü tanıma yaklaşımını kullanan bir ses tanıma sisteminin blok diyagramı görülmektedir.



Şekil 5.1 Bir ses tanıma sisteminin blok diyagramı (Rabiner vd. 1993)

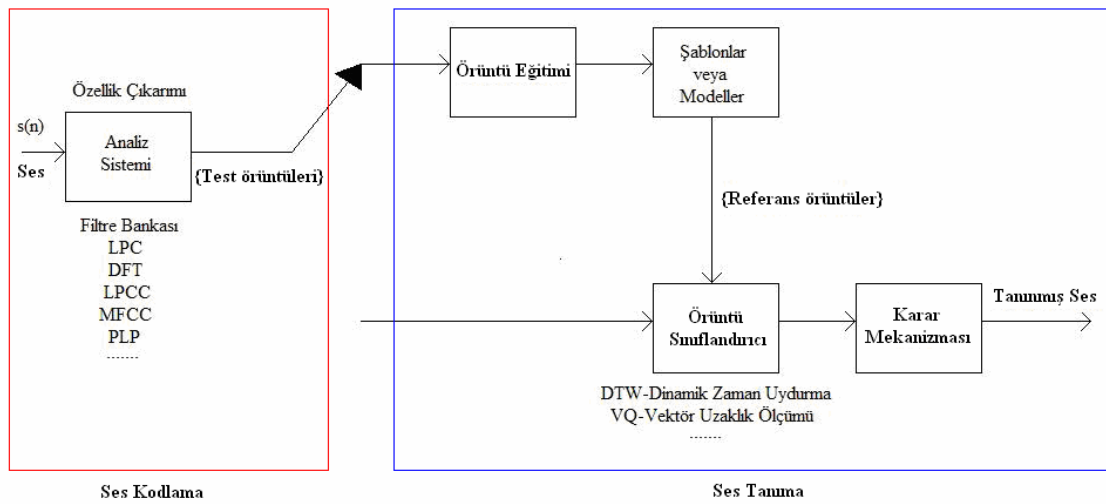
Sesli ifadelerin bir mikrofona ile örneksel sinyallere dönüştürülmesi, sayısallaştırılması ve gerekirse gürültü gibi bozucu etkenlerden arındırılmasının ardından elde edilen ses sinyali analiz edilmek üzere özellik çıkarımı modülüne verilir.

Bu modülün amacı, çeşitli spektral analiz tekniklerini kullanarak (örneğin, Filtre Bankası, LPC, DFT, LPCC, MFCC, PLP gibi) ses sinyalinden, bir parametrik ifade elde etmektir. Bu parametrik ifadeye *Özellik Vektörü* (ya da *Özellik Vektörleri*) denir. Özellik vektörleri ses sinyalini temsil eden parametreler kümesidir.

Özellik vektörleri ya da diğer ifadeyle örüntüler, sistemi eğitmek üzere örüntü (pattern) eğitimi modülüne giriş veri kümeleri olarak verilir. Oluşan referans örüntüler şablonun veya modelin çıktısı olmaktadır.

Örüntü sınıflandırıcı, bilinmeyen bir test örüntüsü ile her referans örüntü sınıfını karşılaştırır diğer ifadeyle bilinmeyen test örüntüsü ile her referans örüntü arasındaki mesafe ölçümünü gerçekleştirerek örüntülerin arasındaki benzerliği ölçer. Bu benzerlik ölçümü için kullanılacak çeşitli yöntemler vardır, bunlardan birkaçı; Dinamik Zaman Uydurma ve Vektör Uzaklık Ölçümü olarak sayılabilir.

Sonuç modülü olarak yer alan karar mekanizması modülü ise bir önceki aşamada ölçülen benzerliği değerlendirerek bilinmeyen test örüntülerine en iyi eşleşen referans örüntüyü seçer. Yani bu modül adından da anlaşılacağı gibi karar verme işlevini yerine getirmektedir.



Şekil 5.2 Bir ses tanıma sisteminin modülleri (Rabiner vd. 1993).

Ses tanıma yöntemlerini şu şekilde sınıflandırmak mümkündür:

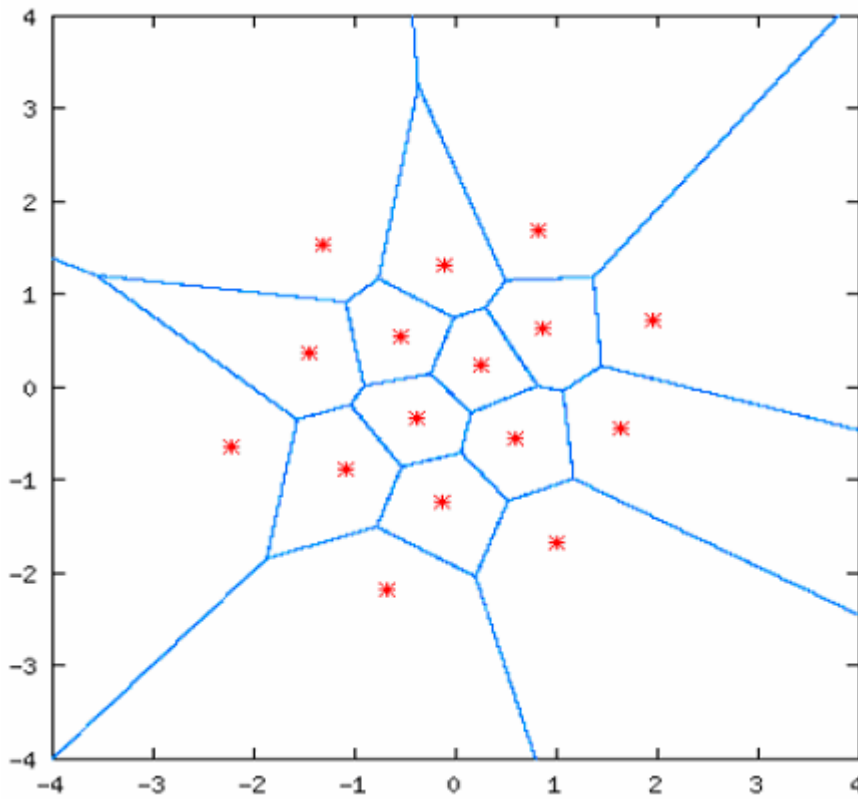
- Şablon Modeller
 - Vektör Uzaklık Ölçümü (VQ-Vector Quantization)
 - Dinamik Zaman Uydurma (DTW-Dynamic Time Warping)
- Stokastik Modeller
 - Gaussian Karma Modelleri (GMMs-Gaussian Mixture Models)
 - Gizli Markov Modelleri (HMMs-Hidden Markov Models)
- Yapay Sinir Ağları (Artificial Neural Networks)
- Bulanık-Sinirsel Sistemler (Fuzzy-Neural Systems)
- Destek Vektör Makineleri (Support Vector Machines)

5.1. Şablon Modeller

Bu bölümde şablon model olarak isimlendirilen Vektör Uzaklık Ölçümü (VQ) ve Dinamik Zaman Uydurma (DTW) yöntemleri açıklanacaktır.

5.1.1. Vektör Uzaklık Ölçümü (VQ)

Vektör Uzaklık Ölçümü ses tanıma sistemlerinde kullanılan yöntemlerden birisidir. 2-boyutlu bir vektör uzaklık ölçümü şekil 5.3'te gösterilmektedir. Burada * sembolü ile gösterilenler *Kod Vektörü*, kod vektörüne yaklaştırılan bölgeler *Kodlama Bölgesi* ve kod vektörlerinin oluşturduğu gruba da *Kod Kitabı* denir.



Şekil 5.3 2-boyutlu vektör uzaklık ölçümü (Uzunçarşılı 2005)

Vektör uzaklık ölçümüne dayalı ses tanıma sistemlerinde eğitim örüntüleri kullanılarak her bir örüntüye ilişkin özellik vektörleri çıkarılır. Özellik vektörlerini elde etmek için, FFT, LPC, LPCC, MFCC, PLP gibi yöntemlerden herhangi biri kullanılabilir. Her bir örüntüye ilişkin özellik vektörleri çıkarıldıktan sonra, bu vektörler, kümelendirme algoritmalarından biri kullanılarak eğitilir ve her bir örüntü için optimum referans model yani kod kitabı tasarlanır.

Sistemin test aşamasında ise test örüntülerine ilişkin özellik vektörleri de çıkarıldıktan sonra referans model olarak hazırlanan kod kitabı vektörleriyle en yakın uzaklığı veren kod vektörleri bulunur. Bilinmeyen örüntü, belirlenen karar kriterine göre örüntülerden birine atanır.

5.1.2. Dinamik Zaman Uydurma (DTW)

Dinamik Zaman Uydurma yöntemi ses tanıma sistemlerinde kullanılan yöntemlerden birisidir. Bu yöntemde, ses ifadeleri, seslendirme süreleri sıkıştırılarak ya da genişletilerek referanslar ile karşılaştırılır. Belirli bir sözcüğün seslendirilmesi, aynı kişi seslendirse bile zaman içinde farklılık gösterebilmektedir. Dinamik Zaman Uydurma yöntemi ile, bu iki seslendirme, zaman içinde genişletilerek ya da daraltılarak birbirine yaklaştırılmaya çalışılır. Yani bu iki seslendirmenin, zaman olarak birbiri ile örtüştürülmesi işlevi gerçekleştirilmeye çalışılır. Bu yöntem dinamik programlama tekniklerinin kullanımı ile gerçekleştirilmektedir.

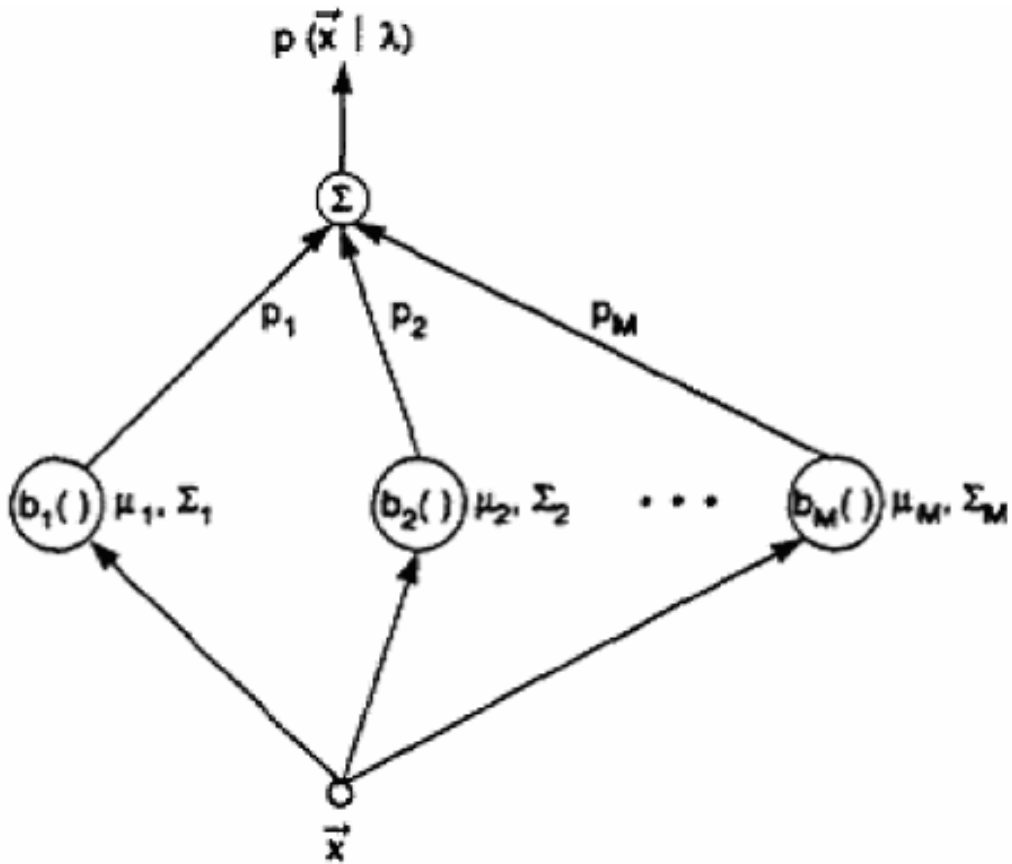
5.2. Stokastik Modeller

Bu bölümde stokastik (olasılıksal) model olarak isimlendirilen Gaussian Karma Modelleri (GMMs) ve Gizli Markov Modeller (HMMs) açıklanacaktır.

5.2.1. Gaussian Karma Modelleri (GMMs)

Gaussian karma yoğunluk fonksiyonu şekil 5.4'de gösterildiği gibi M bileşenli yoğunluk fonksiyonunun ağırlıklandırılmış toplamıdır (Uzunçarşılı 2005) ve eşitlik 5.1 ile ifade edilir.

Bu eşitlikte \vec{x} D-boyutlu rastgele bir vektör, $b_i(\vec{x})$, \vec{x} rastgele değişkenin i. küme için ortak yoğunluk fonksiyonu ve p_i 'ler de karma ağırlıklarıdır ($i=1,2,\dots,M$). Her bir küme için \vec{x} rastgele değişken dağılımı, ortalama vektörü μ_i ve kovaryans matrisi Σ_i olan çok değişkenli Gaussian dağılımları ile ifade edilir. Karma ağırlıkları $\sum_{i=1}^M p_i = 1$ şartını sağlar. $|\Sigma_i|$ işareti determinantı göstermek üzere, i. kümenin ortak olasılık yoğunluk fonksiyonu eşitlik 5.2 ile tanımlanır.



Şekil 5.4 M bileşenli Gaussian yoğunluk (Uzunçarşılı 2005)

$$p(\bar{x}/\lambda) = \sum_{i=1}^M p_i b_i(\bar{x}) \quad (5.1)$$

$$b_i(\bar{x}) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2}(\bar{x} - \bar{\mu}_i)' \Sigma_i^{-1} (\bar{x} - \bar{\mu}_i)\right) \quad (5.2)$$

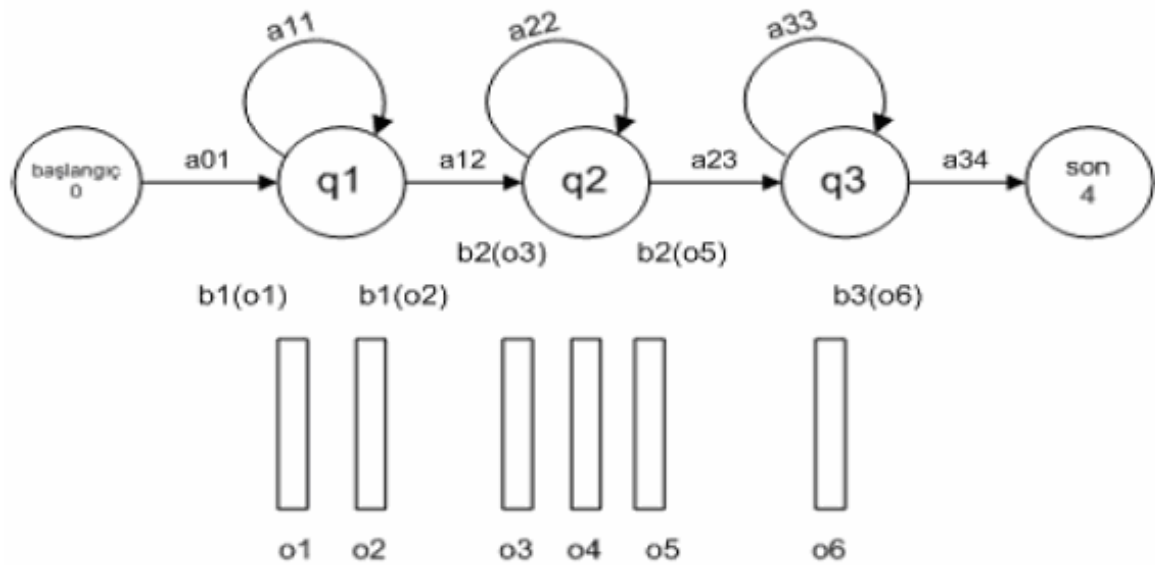
Tam bir Gaussian karma yoğunluk fonksiyonu; her bir yoğunluk fonksiyonunun ortalama vektörleri, kovaryans matrisleri ve karma ağırlıklarıyla parametrelendirilir. Bu parametrelendirmenin ortak gösterimi eşitlik 5.3 ile yapılmaktadır.

$$\lambda = \left\{ p_i, \mu_i, \Sigma_i \right\}, \quad i = 1, 2, \dots, M \quad (5.3)$$

Sonuç olarak, Gaussian karma modeli λ ile ifade edilmektedir.

5.2.2. Gizli Markov Modeller (HMMs)

Gizli Markov Model (HMM), ayrıık zamanlı serilerin, gözlenmiş veri örneklerinin sınıflandırılması için güçlü bir stokastik (olasılıksal) yöntemdir (Baygün 2006). Ses tanıma sistemlerinde kullanılan Gizli Markov Modelin amacı, bilinmeyen (gizli) durum dizisini gözlemlere dayanarak bulmaktır. Şekil 5.5’de örnek bir Gizli Markov Modelin gösterimi sunulmaktadır.



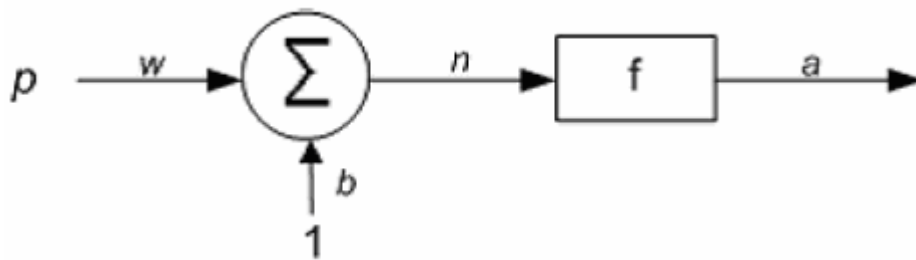
Şekil 5.5 Örnek bir Gizli Markov Model (Baygün 2006)

Ses tanıma uygulamaları bir eğitim aşaması ve bu aşamayı takip eden bir test aşamasından oluşur. Eğitim aşaması süresince temel alınacak referans modelin parametreleri kestirilir. Gizli markov modelde kestirilecek üç parametre vardır. Bunlardan biri, t anında i . durumda ve $t+1$ anında j . durumda olma geçiş olasılıklarını gösteren a_{ij} elemanlarını içeren durum geçiş olasılığı matrisi A 'dır. $O = \{o_1, o_2, \dots, o_T\}$ olan bir gözlem dizisi tanımlandığı zaman bu vektör dizisinin her bir elemanı, ses tanıma sistemlerinde kullanılan özellik vektörlerini gösterir. $B = [b_j(o_t)]$ gözlem sembol olasılık dağılımı matrisidir. $b_j(o_t)$, o_t vektörünün t anında j . durumda gözlemlenmesi olasılığını gösterir. $\pi = \{\pi_i\}$ vektörü, başlangıçta i . durumda olma olasılığını belirten başlangıç durum dağılımını gösterir. Bu üç parametre, $\lambda = \{A, B, \pi\}$ olan Gizli Markov Modeli oluşturur. Bunların dışında durum sayısı N , her bir durumdaki karma sayısı M de diğer parametrelerdir. Gözlem sembol olasılığını

göstermek için de çeşitli yöntemler vardır ama bunlardan en çok kullanılanı Sürekli Olasılık Yoğunluk Fonksiyonlarıdır.

5.3. Yapay Sinir Ağları

Yapay Sinir Ağı, insan beyinde yer alan sinir hücrelerinin yapısı ve çalışması temel alınarak oluşturulmuş bir modeldir. Bir ANN, nöronlar ve bu nöronların birbiriyle olan bağlantılarından meydana gelmektedir. Tek girişli bir nöron modeli şekil 5.6'da verilmiştir.



Şekil 5.6 Tek girişli bir nöron modeli

Şekil 5.6'da gösterilen nöron modelinde; p , bu nörona olan girişi, a da, bu nöronun çıkışı ifade etmektedir. Burada w , ağırlık (*weight*) ve b , öngerilim (*bias*) olarak tanımlanır.

Şekil 5.6'da verilen nöron modelinin çıkış fonksiyonu (yani a) eşitlik 5.4 ile ifade edilir;

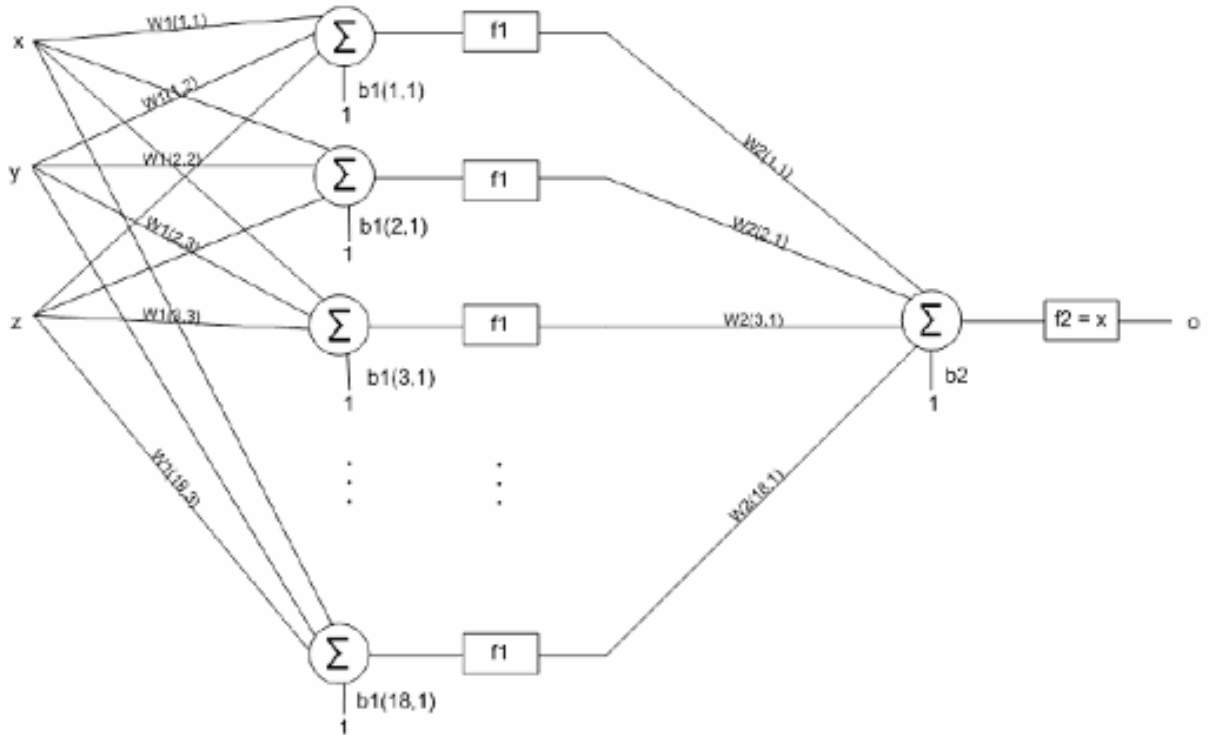
$$a = (f \cdot w + b) \quad (5.4)$$

Bu şekilde birçok nöronun birbirine bağlanarak oluşturduğu yapı, Yapay Sinir Ağlarını meydana getirir. Şekil 5.7'de çok girişli tek çıkışlı bir ANN modeli görülmektedir.

Ses tanıma sistemlerinde ANN yaklaşımına dayalı çeşitli yöntemler geliştirilmiştir. Bunlar:

- Perceptron
- Multilayer Networks
- Backward Error Propagation

- Kohonen Self-Organizing Maps
- Hopfield Nets ve Associative Memory



Şekil 5.7 Tipik çok girişli tek çıkışlı bir ANN modeli

ANN'ler ses tanıma sistemleri için kullanılabilir bir tekniktir. Ancak ANN'lerin de bazı dezavantajları bulunmaktadır. Bu teknikteki en büyük problem karmaşık sorunları çözmek için (ses tanıma gibi) çok büyük yani çok katmanlı ve çok nöron içeren sinir ağlarına ihtiyaç duyulmasıdır.

ANN modelinde eğitim ve test olmak üzere iki aşama vardır. Eğitim aşamasında ANN'deki düğümlerin birbirine bağlantı yüzdelerini gösteren ağırlık değerleri hesaplanır. Elde edilen bu ağırlık değerleri daha sonra sadece girişlerin verilip çıkışların hesaplanmasının istenildiği test aşamasında işe yarar. Test aşamasındaki algoritma eğitim aşamasındakine göre daha kolaydır. Bundan dolayı ANN'ler bu aşamada daha hızlı çalışır. ANN'ler ses tanıma sistemlerinde şu şekilde kullanılabilir. Örneğin, belirli bir kelimedenden FFT, LPC, MFCC ya da PLP gibi bir yöntemle elde edilen katsayılar, ANN'nin giriş katmanına yüklenir. Kullanılan teknik ve elde edilen katsayıların miktarı, ANN'nin başarısını ve çalışma hızını etkilemesi açısından önem taşımaktadır. Sonra çıkış katmanına bu kelimeyi temsil edecek bir kod yüklenir. İşleme devam edilir. Bir

başka kelime için bu çalışmalar tekrar edilir. Eğitim aşaması tamamlanınca test aşamasına geçilir. Bu aşamada ise konuşulan bir kelime ANN'lerin girişine verilir, eğitim aşamasında hesaplanan katsayılar (ağırlık değerleri) kullanılarak çıkışların hesaplanması için ANN'ler yeniden çalıştırılır. Elde edilen sonuçlara göre sesin tanınıp tanınmadığına karar verilir. Böylece ANN'ler ile ses tanıma işlemi gerçekleştirilmiş olur.

5.4. Bulanık-Sinirsel Sistemler

Niteliği tam anlaşılamayan, iyi seçilmeyen, açık seçik görünmeyen, net olmayan şekilde tanımlanan Bulanıklık, dereceli üyelik kavramı ile teknik bilim dünyasına taşınmıştır (Nabiyev 2005). Bulanık kümelerde dereceli üyelik kavramı ilk kez 1965 yılında California üniversitesinden Prof. Dr. Lotfi A. Zadeh tarafından ortaya atılmıştır.

Yapay Sinir Ağları ile Bulanık Mantığın karışımından oluşan Bulanık-Sinirsel sistemler (Fuzzy-Neural systems), sistem modelleme, tıbbi teşhis, örüntü tanıma, ses tanıma gibi alanlarda önemli gelişmeler gösterebilecek potansiyele sahiptir.

Bulanık-Sinirsel Sistemlerde ağ, üyelik fonksiyonlarının üretilmesini sağlayacak biçimde tasarlanmaktadır. İstenen çıkış değeri elde edildiğinde, üretilmiş üyelik fonksiyonları optimum sayılır ve Yapay Sinir Ağı devre dışı bırakılır.

5.4.1. Bulanık-sinirsel sınıflandırıcılar

5.4.1.1. Bulanık çıkarım

Geleneksel deduktif çıkarım kuralının (Modus Ponens) bulanık mantıkta çalışması şu şekilde olmaktadır. U ve V sonlu evrensel kümelerde x ve y değişkenleri tanımlanmış olsun. Bu değişkenler değerlerini A_k ve B_k bulanık kümelerinden alır. Bu bulanık kümelerin $\mu_{A_k}(u_i)$ ve $\mu_{B_k}(u_i)$ üyelik fonksiyonları verilmiş olsun. x ve y değişkenleri arasındaki ilişki,

$$\text{IF } x = A_k \text{ THEN } y = B_k \quad (k=1,2,\dots,m)$$

kuralları yardımıyla verilmektedir. Bu kurallar çerçevesinde giriş ve çıkışı bağlayan R ilişkisi eşitlik 5.5 ile hesaplanmaktadır;

$$\mu_R(u_i, v_j) = \max_k \min(\mu_{A_k}(u_i), \mu_{B_k}(v_j)) \quad (5.5)$$

Burada R ilişkisi ' $A \rightarrow B$ ' koşulunun sayısal değerlendirmesidir. R ilişkisini kullanarak, U kümesinde tanımlı bulanık kümelerin tümüne, V evrensel kümesinde tanımlı olan B bulanık kümesi karşılık düşer. B 'nin üyelik fonksiyonu eşitlik 5.6 ile hesaplanır.

$$\mu_B(v_j) = \max_i \min(\mu_{A_k}(u_i), \mu_R(u_i, v_j)) \quad (5.6)$$

Küme teorisi açısından bu çıkarım $B = A \circ R$ olarak gösterilir. *Modus Ponens*'in bulanık mantıktaki tanımı şekil 5.8'de verilmiştir.

$F \rightarrow G$

F'

G'

Şekil 5.8 Bulanık mantıkta *Modus Ponens*

Bulanık mantıktaki ifadelendirmenin klasik modus ponens'ten farkı burada F ve F' kümelerinin aynı olmamasıdır. Eğer F ile F' karşılaştırılabilir ise bu benzerlik çerçevesinde G' bulanık çıkarımı yapılabilir. G' bulanık çıkarımı, F' ile $F \rightarrow G$ ilişkisini ifade eden R bulanık kümelerinin kompozisyonu biçiminde belirlenir (Eşitlik 5.7).

$$G' = F' \circ R = \sum_{i=1}^m \vee_{u_i \in U} (\mu_{F'}(u_i) \wedge \mu_R(u_i, v_j)) / v_j \quad (5.7)$$

Burada $\{F, F'\} \subset U$ ve $\{G, G'\} \subset V$ olmaktadır.

5.4.1.2. Bulanık-sinirsel sınıflandırıcılar

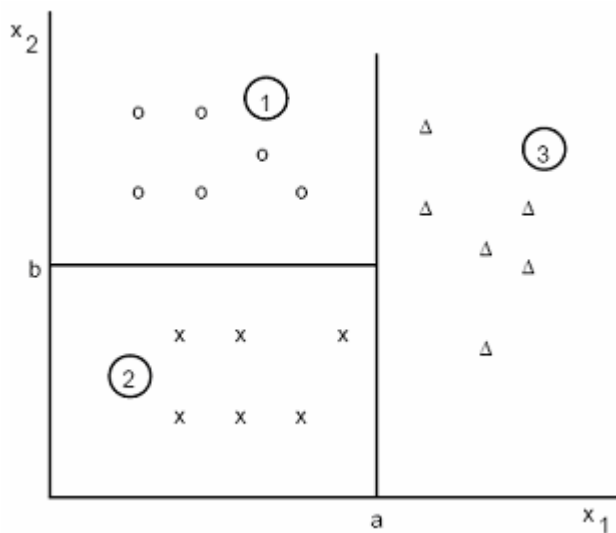
x_1 ve x_2 özellikleri ile karakterize edilen her bir örneğin bir bulanık-sinirsel sınıflandırıcı ile 3-farklı sınıfta gruplandırılması örneği şekil 5.9'da ve bu örneğe ilişkin sınıflandırma ağacı da şekil 5.10'da görülmektedir.

Burada sınıflandırma kuralı şu şekilde oluşturulabilir ve bu örneğe ilişkin bulanık-sinirsel sınıflandırma ağacı şu şekilde olur (Şekil 5.11);

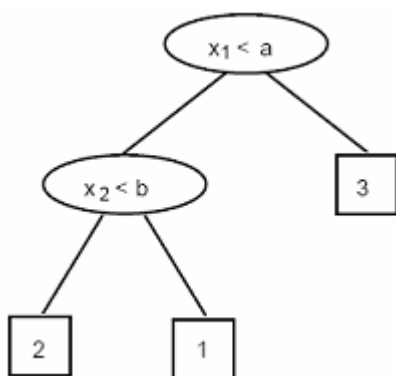
```

IF  $x_1 < a$ 
  THEN IF  $x_2 < b$ 
    THEN class 2
    ELSE class 1
  ELSE class 3

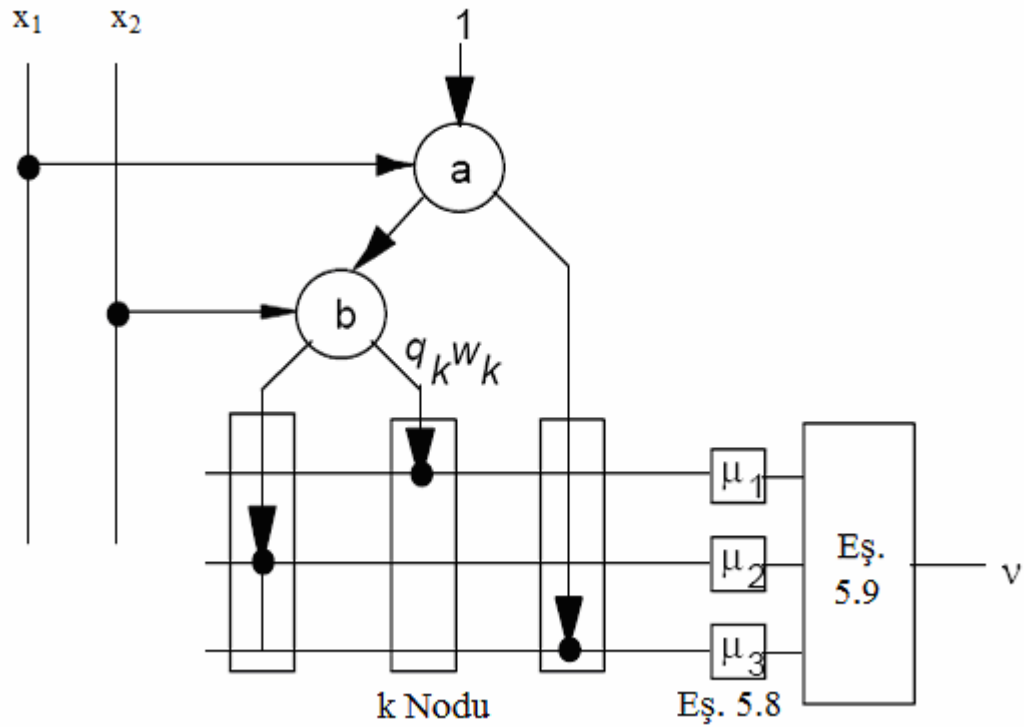
```



Şekil 5.9 Bir 3-sınıf sınıflandırma örneği



Şekil 5.10 Bir 3-sınıf sınıflandırma örneğinin ağaç olarak gösterimi



Şekil 5.11 Bir bulanık-sinirsel sınıflandırma ağacı

$$\mu_i = \frac{\sum_{k=1}^K q_k w_k \langle \mu \rangle_{k^i}}{\sum_{k=1}^K q_k w_k} \quad (5.8)$$

$$v = \frac{\sum_{i=1}^M \mu_i c_i}{\sum_{i=1}^M \mu_i} \quad (5.9)$$

5.5. Destek Vektör Makineleri

Bu bölümde öncelikle kısıtlamalı optimizasyon teorisi, ardından istatistiksel öğrenme teorisi ve son olarak destek vektör makineleri açıklanacaktır.

5.5.1. Kısıtlamalı optimizasyon

5.5.1.1. Kısıtlamalı optimizasyon teorisi

Bu bölümde optimizasyon probleminin genel biçimi açıklanacaktır. Optimizasyon problemi, verilen bazı kısıtlamalar altında bir fonksiyonun maksimumunu ya da

minimumunu bulmadır (Cristianini vd. 2000). Bir optimizasyon problemi şu şekilde tanımlanır (Eşitlik 5.10, 5.12 ve eşitsizlik 5.11);

$$\text{Minimizasyon} \quad f(w), \quad w \in \Omega \quad (5.10)$$

$$\text{Kısıtlamalar:} \quad g_i(w) \leq 0, \quad i = 1, 2, \dots, k \quad (5.11)$$

$$h_i(w) = 0, \quad i = 1, 2, \dots, m \quad (5.12)$$

Burada $f(w)$ *Amaç Fonksiyonu*, $g_i(w) \leq 0$ ifadesi *Eşitsizlik Kısıtlaması* ve $h_i(w) = 0$ ifadesi de *Eşitlik Kısıtlaması* olarak isimlendirilir. Amaç fonksiyonunun verilen bu eşitlik ve eşitsizlik kısıtlamaları altında bulunan optimal değerine de *Optimizasyon Probleminin Değeri* adı verilir.

Amaç fonksiyonunun tanımlı olduğu ve tüm kısıtlamaların sağlandığı bölgeye *Feasible Bölge* denir. Feasible bölge şu şekilde tanımlanır (Eşitlik 5.13);

$$R = \{w \in \Omega : g(w) \leq 0, h(w) = 0\} \quad (5.13)$$

Optimizasyon probleminin çözümü $w^* \in \mathfrak{R}$ gibi bir noktadır öyle ki amaç fonksiyonunun bu noktada alacağı değer en küçük olmalıdır yani $f(w) < f(w^*)$ olacak bir başka $w \in \mathfrak{R}$ noktası daha bulunamaz. Bu noktaya aynı zamanda *Global Minimum* denir. Bir $w^* \in \Omega$ noktası eğer bazı $\varepsilon > 0$ için $f(w) \geq f(w^*)$ oluyor ise ve her $w \in \Omega$ için $\|w - w^*\| < \varepsilon$ ise, bu noktaya *Yerel Minimum* adı verilir.

Bir optimizasyon probleminde amaç fonksiyonu, eşitlik ve eşitsizlik kısıtlamaları doğrusal fonksiyon ise bu probleme *Doğrusal Program* adı verilir. Eğer amaç fonksiyonu karesel ve tüm kısıtlamalar doğrusal ise bu optimizasyon problemi bir *Karesel Program* adını alır.

Bir eşitsizlik kısıtlamasına, $g_i(w) \leq 0$, *Aktif* denir, eğer çözüm, w^* ise, ve bu çözüm eşitsizlik kısıtlamasını sıfır yapıyor ise yani $g_i(w^*) = 0$ oluyor ise, aksi halde bu eşitsizlik kısıtlamasına *Aktif-Değil* denir. O halde tüm eşitlik kısıtlamaları her zaman aktiftir. Gevşek değişken olarak adlandırılan kimi değişkenler, bir eşitsizlik kısıtlamasını eşitlik kısıtlamasına dönüştürürler (Eşitlik 5.14);

$$g_i(w) \leq 0 \Leftrightarrow g_i(w) + \xi_i = 0, \quad \xi_i \geq 0. \quad (5.14)$$

Gerçek-değerli bir $f(w)$ fonksiyonu *Konvektir* eğer her $w, u \in \mathfrak{R}^n$ ve herhangi bir $\theta \in (0,1)$ için şu koşul sağlanıyorsa (Eşitsizlik 5.15);

$$f(\theta w + (1-\theta)u) \leq \theta f(w) + (1-\theta)f(u). \quad (5.15)$$

Ω kümesi içinde tanımlı bir optimizasyon problemi, amaç fonksiyonu ve tüm kısıtlamalar konveks ise konvektir denir. Tüm kısıtlamalar doğrusal, amaç fonksiyonu konveks, karesel ve $\Omega = \mathfrak{R}^n$ ise, bu optimizasyon problemi *Konveks Karesel Program* adını alır.

Optimizasyon teorisi hem optimal noktaları karakterize eden temel özelliklerin tanımlanması ve hem de çözümleri elde etme için oluşturulacak algoritmaların tasarlanması ile ilgilenir.

5.5.1.2. Lagrange teorisi

Lagrange teorisinin amacı başlangıçta eşitsizlik kısıtlamaları olmadığı zaman bir optimizasyon probleminin çözümünü karakterize etmektir. Bu teorideki ana kavramlar Lagrange çarpanları ve Lagrange fonksiyonudur. Bu yöntem 1797 yılında Lagrange tarafından mekanik problemleri için, Fermat'ın 1629 yılında ortaya attığı bir sonuç geliştirilerek, geliştirildi. 1951 yılında Kuhn ve Tucker eşitsizlik kısıtlamalarının kullanımına da izin vererek yöntemi genişletti ve bugün Kuhn-Tucker teorisi olarak bilinen biçime dönüştürdü.

w^* noktasının $f(w)$ nun bir minimumu olması için bir gerek şart, $f \in C^1$ olmak üzere, $\frac{\partial f(w^*)}{\partial w} = 0$ olmasıdır. Bu şart aynı zamanda bir yeter şarttır.

$f(w)$ amaç fonksiyonu ile verilen bir optimizasyon probleminde eşitlik kısıtlamaları $h_i(w) = 0$, $i = 1, 2, \dots, m$ ile gösterilir ise, Lagrange fonksiyonu şu şekilde tanımlanır (Eşitlik 5.16);

$$L(w, \beta) = f(w) + \sum_{i=1}^m \beta_i h_i(w) \quad (5.16)$$

Burada β_i Lagrange Çarpanı olarak isimlendirilir.

Bir normal w^* noktasının $h_i(w) = 0$, $i = 1, 2, \dots, m$ kısıtlamaları altında $f(w)$ 'nin bir minimumu olması için bir gerek şart, $f, h_i \in C^1$ olmak üzere, bazı β^* değerleri için,

$$\frac{\partial L(w^*, \beta^*)}{\partial w} = 0 \quad (5.17)$$

$$\frac{\partial L(w^*, \beta^*)}{\partial \beta} = 0 \quad (5.18)$$

olmasıdır. Bu koşullar aynı zamanda $L(w, \beta^*)$ 'nin w 'nin bir konveks fonksiyonu olmasını sağlayan yeter koşullardır (Eşitlik 5.17 ve 5.18).

$\Omega \subseteq \mathfrak{R}^n$ uzayında verilen bir optimizasyon problemi,

$$\text{Minimizasyon} \quad f(w), \quad w \in \Omega \quad (5.19)$$

$$\text{Kısıtlamalar:} \quad g_i(w) \leq 0, \quad i = 1, 2, \dots, k \quad (5.20)$$

$$h_i(w) = 0, \quad i = 1, 2, \dots, m \quad (5.21)$$

olsun (Eşitlik 5.19, 5.21 ve eşitsizlik 5.20). Buna göre genelleştirilmiş Lagrange fonksiyonu şu şekilde tanımlanır (Eşitlik 5.22):

$$L(w, \alpha, \beta) = f(w) + \sum_{i=1}^k \alpha_i g_i(w) + \sum_{i=1}^m \beta_i h_i(w) \quad (5.22)$$

Primal problemin Lagrange duali,

$$\text{Maksimizasyon} \quad \theta(\alpha, \beta), \quad (5.23)$$

$$\text{Kısıtlamalar:} \quad \alpha \geq 0 \quad (5.24)$$

olur (Eşitlik 5.23 ve eşitsizlik 5.24). Burada $\theta(\alpha, \beta) = \inf_{w \in \Omega} L(w, \alpha, \beta)$ dir. Optimal çözüm için amaç fonksiyonunun değeri, optimizasyon probleminin değeri olarak adlandırılır.

$\Omega \subseteq \mathfrak{R}^n$ uzayında verilen bir optimizasyon problemi, $f \in C^1$ konveks olmak üzere,

$$\text{Minimizasyon} \quad f(w), \quad w \in \Omega \quad (5.25)$$

$$\text{Kısıtlamalar:} \quad g_i(w) \leq 0, \quad i = 1, 2, \dots, k \quad (5.26)$$

$$h_i(w) = 0, \quad i = 1, 2, \dots, m \quad (5.27)$$

olsun (Eşitlik 5.25, 5.27 ve eşitsizlik 5.26). Bir normal w^* noktasının $f(w)$ nun bir minimumu olması için gerek ve yeter şartlar α^*, β^* nın var olmasıdır, şöyle ki (Eşitlik 5.28, 5.29, 5.30 ve eşitsizlik 5.31, 5.32);

$$\frac{\partial L(w^*, \alpha^*, \beta^*)}{\partial w} = 0, \quad (5.28)$$

$$\frac{\partial L(w^*, \alpha^*, \beta^*)}{\partial \beta} = 0, \quad (5.29)$$

$$\alpha_i^* g_i(w^*) = 0, \quad i = 1, 2, \dots, k \quad (5.30)$$

$$g_i(w^*) \leq 0, \quad i = 1, 2, \dots, k \quad (5.31)$$

$$\alpha_i^* \geq 0, \quad i = 1, 2, \dots, k \quad (5.32)$$

Burada üçüncü ifade (Eşitlik 5.30'da) belirtilen ilişki, *Karush-Kuhn-Tucker (KKT) Tamamlayıcı Koşulları* olarak bilinir. Bu ilişki, aktif kısıtlamalar için (yani $\alpha_i^* \geq 0$ için), gerçekleşir, halbuki aktif-olmayan kısıtlamalarda $\alpha_i^* = 0$ olur.

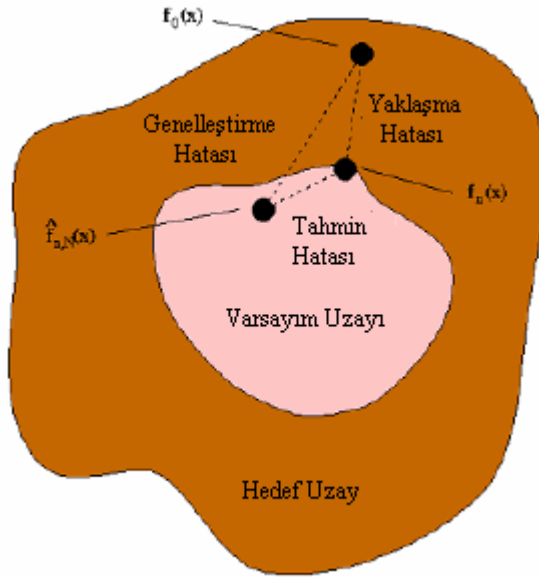
5.5.2. İstatistiksel Öğrenme Teorisi

Modellemede amaç, hedef uzay içerisindeki temel fonksiyona en yakın olan varsayım uzayından bir model seçmektir. Hatalar iki durumdan kaynaklanmaktadır.

Yaklaşma hatası, varsayım uzayının hedef uzaya göre daha küçük olmasından dolayı oluşmaktadır ve bu durumda temel fonksiyon varsayım uzayının dış tarafında yer alabilir.

Tahmin hatası, öğrenme prosedüründen dolayı oluşan hatadır.

Bu hatalar birlikte genelleştirme hatalarını oluşturmaktadır. Modelleme hataları şekil 5.12'de görülmektedir.



Şekil 5.12 Modelleme hataları

Risk'i minimize eden bulunmak istenen fonksiyona f denilirse, Risk,

$$R[f] = \int_{X \times Y} L(y, f(x)) P(x, y) dx dy \quad (5.33)$$

olur (Eşitlik 5.33). Burada $P(x, y)$ bilinmemektedir. Buna göre deneysel risk minimizasyonu prensibine göre bir yaklaşım bulmak mümkündür.

5.5.2.1. VC-boyutu

VC-boyutu bir fonksiyonlar kümesinin kapasitesini ölçen skaler bir değerdir.

Eğer $\{x^i\}_{i=1}^p$, bir noktalar kümesi olmak üzere, bu noktalar kümesi tüm olası 2^p konfigürasyon içinde ayrılabilir ise ve $q > p$ olan bir $\{x^i\}_{i=1}^q$ noktalar kümesi de bulunmuyor ise bu fonksiyonlar kümesinin VC-boyutu p 'dir denir.

5.5.2.2. Yapısal risk minimizasyonu

VC-boyutu h olan bir varsayım uzayı S_h için, $S_1 \subset S_2 \subset \dots \subset S_\infty$ olmak üzere, yapısal risk minimizasyonu şu problemin çözümlerini içerir (Eşitlik 5.34);

$$\min_{S_h} R_{emp}[f] + \sqrt{\frac{h \ln\left(\frac{2l}{h} + 1\right) - \ln\left(\frac{\delta}{4}\right)}{l}} \quad (5.34)$$

5.5.3. Destek Vektör Makinelerine giriş

Bu bölümde Destek Vektör Makinelerine (SVM) ilişkin genel kavramlar ve teorik temeller açıklanacaktır. Bu bölüm ilk olarak SVM sınıflandırıcılarının açıklanması ile başlayacak, ardından SVM sınıflandırıcıları, *doğrusal SVM sınıflandırıcıları* ve *doğrusal olmayan SVM sınıflandırıcıları* olarak iki kategori halinde sunulacaktır. Son olarak, yapılan uygulamada kullanılan SVM sınıflandırıcısı çeşitlerinden *SM SVM sınıflandırıcıları* ile *LS SVM sınıflandırıcıları* açıklanacaktır.

SVM, istatistiksel öğrenme teorisi ve yapısal riski en aza indirme ilkesine dayanan, sınıflandırma ve regresyon problemlerinin çözümü amacıyla Vapnik tarafından ortaya atılmış bir öğrenme yöntemidir (Vapnik 1995, 1998). SVM, herhangi bir sınıflandırma ya da regresyon problemini, bir karesel programlama problemine dönüştürerek yerel çözümlere takılmadan çözerler. Yerel çözümlere takılmama özelliği, SVM'nin diğer tekniklere göre sahip olduğu avantajlardan biridir. Ayrıca SVM, oldukça yüksek genelleme yapabilme yeteneğine sahiptir. SVM, çok değişik alanlarda uygulanmıştır (Cristianini vd. 2000, Schölkopf vd. 1999). SVM'nin uygulama alanlarına örnek olarak, el yazısı tanıma, yüz tanıma, 3-boyutlu nesne tanıma, ses tanıma, konuşmacı tanıma, metin sınıflandırma verilebilir. SVM sınıflandırıcıları, margin'i (aralığı) maximum yapan bir en uygun (optimal) ayırıcı aşırıdüzlemi oluşturmaya çalışır. Burada bahsedilen margin kavramı, ayırıcı aşırıdüzlemden, en yakın veri noktasına olan minimum uzaklığı tanımlamaktadır.

5.5.3.1. Destek Vektör Makinesi sınıflandırıcıları

SVM sınıflandırıcılarında karşılaşılabilecek iki durum, verilerin doğrusal olarak ayrılabilen bir yapıda olması ya da doğrusal olarak ayrılamayan bir yapıda olmasıdır. Ancak, gerçek yaşam problemlerinin büyük çoğunluğu birçok farklı bileşenden oluşan problemlerdir ve doğrusal olarak ayrılmış bir yapı şeklinde olmayabilirler. Doğrusal olarak ayrılabilen veriler arasında maksimum margin'in bulunması işlemi kolaydır, ancak doğrusal olarak ayrılamayan veriler üzerinde sınıflandırma yaparken, bu veriler

öncelikle doğrusal olarak ayrılacakları farklı bir uzaya aktarılmalıdırlar. Ardından bu yeni uzayda sınıflandırılırlar.

SVM sınıflandırıcıları, doğrusal SVM sınıflandırıcıları ve doğrusal olmayan SVM sınıflandırıcıları olmak üzere iki bölümde incelenecektir.

5.5.3.1.1. Doğrusal Destek Vektör Makinesi sınıflandırıcıları

Doğrusal SVM sınıflandırıcıları doğrusal olarak ayrılabilen ve doğrusal olarak ayrılamayan veriler üzerinde işlem yapan SVM sınıflandırıcıları olarak iki grupta incelenecektir.

5.5.3.1.1.1. Doğrusal olarak ayrılabilme durumu

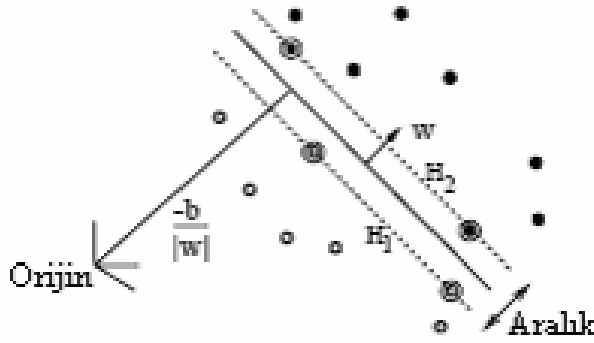
l elemandan oluşan eğitim veri kümesinin $\mathcal{X} = \{x_i, y_i\}, i = 1, 2, \dots, l$ olduğu varsayalım. Burada $y_i \in \{-1, 1\}$ etiket değerleri ve $x_i \in \mathcal{R}^d$ özellikler vektörüdür.

Doğrusal olarak ayrılabilme durumunda, bu iki değerli veriler bir aşırıdüzlem ile ayrılacaktır. Bu aşırıdüzleme *Ayırıcı Aşırıdüzlem* adı verilir. SVM'nin amacı, bu aşırıdüzlemin iki örnek grubuna eş uzaklıkta olmasını sağlamaktır. Doğrusal olarak ayrılabilme durumu için, ayırıcı aşırıdüzlemi şu şekilde tanımlanabilir (Eşitlik 5.35):

$$f(x) = w^T \cdot x + b = \sum_{i=1}^l w_i x_i + b \quad (5.35)$$

Burada w bir ağırlık vektörünü (veya aşırıdüzlemin normalini), b skaleri de bir sapma değerini ifade etmektedir ve bu parametrelerin alacağı değerler ayırıcı aşırıdüzlemin pozisyonunu belirler.

Doğrusal ayrılabilme durumunda optimal ayırıcı aşırıdüzlem şekil 5.13'de görülmektedir.



Şekil 5.13 Doğrusal ayrılabilme durumunda optimal ayırıcı aşırıdüzlem

Aşırıdüzlem üzerindeki herhangi bir x noktası, w aşırıdüzlemin normali ve $|b|/\|w\|$ aşırıdüzlemin orijine dik uzaklığı olmak üzere,

$$w^T \cdot x + b = 0 \quad (5.36)$$

koşulunu sağlar (Eşitlik 5.36). SVM yönteminde araştırılan, örnekler doğrusal olarak ayrılabilirse, $y_i = +1$ ve $y_i = -1$ etiketli örneklere eşit uzaklıkta olan en uygun ayırıcı aşırıdüzlemin bulunmasıdır. Bunun için eğitim veri kümesinin aşağıdaki yapıları sağladığı varsayılırsa:

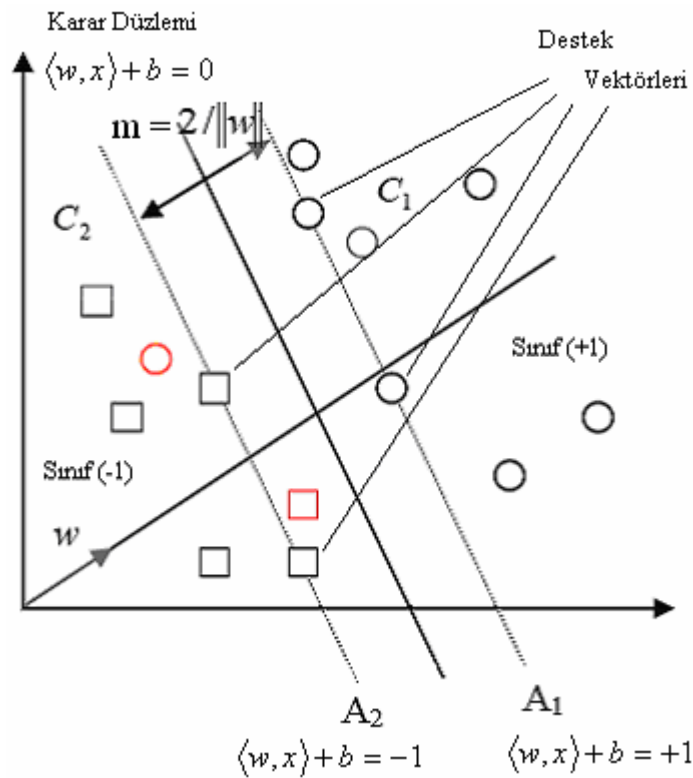
$$y_i = +1 \text{ için, } w^T \cdot x_i + b \geq +1 \quad (5.37)$$

$$y_i = -1 \text{ için, } w^T \cdot x_i + b \leq -1 \quad (5.38)$$

5.37 ve 5.38 eşitsizlikleri bir arada ifade edilecek olursa, $i = 1, 2, \dots, l$ için,

$$y_i \cdot (w^T \cdot x_i - b) \geq 1 \quad (5.39)$$

koşulu elde edilir. 5.39 eşitsizliğini (koşulunu) sağlayan aşırıdüzlemin iki tarafındaki en yakın örneklere olan dik uzaklıkları toplamı *Margin (Aralık, Sınır)* olarak adlandırılır. En uygun (optimal) ayırıcı aşırıdüzlem, margin'i maksimum yapan aşırıdüzlemdir (Şekil 5.14). En uygun ayırıcı aşırıdüzlemi bulmak için, w ve b değerleri hesaplanacaktır.



Şekil 5.14 Doğrusal ayrılabilir durumda optimal ayırıcı aşırıdüzlem

5.37 eşitsizliğinin şekil 5.14'deki C_1 sınıfını ayıran A_1 aşırıdüzlemini oluşturan eşitsizlik, 5.38 eşitsizliğinin ise şekil 5.14'deki C_2 sınıfını ayıran A_2 aşırıdüzlemini oluşturan eşitsizlik olduğu varsayılırsa, bu durumda A_1 aşırıdüzleminin orijine dik uzaklığı $|1-b|/\|w\|$ ve A_2 aşırıdüzleminin orijine dik uzaklığı $|-1-b|/\|w\|$ olur. Bu iki aşırıdüzlemin optimal aşırıdüzleme uzaklıkları ise $1/\|w\|$ olur ve iki örnek kümesi arasındaki uzaklık A_1 ve A_2 aşırıdüzlemlerinin birbirlerine paralel olmalarından dolayı $2/\|w\|$ olur. A_1 ve A_2 aşırıdüzlemleri arasındaki maksimum uzaklık ise ağırlık vektörü w 'nin Euclid normunun (yani $\|w\|$ değerinin) minimize edilmesiyle bulunur. Dolayısıyla margin'i maksimize etme problemi, verilen kısıtlamalar altında $\|w\|$ değerini minimize etme problemi haline dönüşmüş olur. Artık bu optimizasyon problemi bir konveks karesel programlama (QP) problemidir ve Lagrange optimizasyon tekniği bu optimizasyon probleminin çözümü için etkili bir yöntemdir. SVM yöntemiyle yapılmaya çalışılan, bu iki aşırıdüzlemin arasındaki uzaklığın (margin'in) maksimum olmasını sağlamaktır, dolayısıyla $\|w\|$ 'nin minimum olmasını sağlamaktır. Aralarındaki

margin maksimum olan aşırıdüzlemlerin üzerinde yer alan noktalara ise *Destek Vektörleri* adı verilmektedir.

Maksimum margin'in bulunması işlemi;

$$\text{Minimizasyon} \quad f = \|w\|^2 / 2 \quad (5.40)$$

$$\text{Kısıtlama:} \quad g_i = -(y_i (w^T \cdot x_i + b) - 1) \leq 0, \quad i = 1, 2, \dots, l \quad (5.41)$$

ile ifade edilir. Ağırlık vektörü w 'nun Euclid normu $\|w\| = \sqrt{w^T w}$ olarak tanımlanır. Buna göre yukarıdaki ifadeler yeniden yazılır ise,

$$\text{Minimizasyon} \quad f = w^T w / 2 \quad (5.42)$$

$$\text{Kısıtlama:} \quad g_i = -(y_i (w^T \cdot x_i + b) - 1) \leq 0, \quad i = 1, 2, \dots, l \quad (5.43)$$

olur. Burada 5.42 eşitliği, çözülecek olan problem ve 5.43 eşitsizliği de, bu problemin çözümü sırasında kullanılan koşuldur (kısıtlamadır) ve bu ifadeler ikinci dereceden bir optimizasyon problemidir. Bu problem standart karesel programlama yöntemleri ile çözülebilir. Ancak, problemin çözümü için Lagrange yöntemini kullanmak, problemi doğrusal olarak ayrılamayan durum için çözümede kolaylık sağlamaktadır. Problemin çözümü için, problemin Lagrange fonksiyonu bulunur. Problemin Lagrange fonksiyonu,

$$L_p(w, b, \alpha) = f(x) + \sum_{i=1}^l \alpha_i g_i(x) \quad (5.44)$$

$$L_p(w, b, \alpha) = \frac{w^T \cdot w}{2} - \sum_{i=1}^l \alpha_i (y_i (w^T \cdot x_i + b) - 1) \quad (5.45)$$

$$L_p(w, b, \alpha) = \frac{w^T \cdot w}{2} - \sum_{i=1}^l \alpha_i y_i w^T \cdot x_i - \sum_{i=1}^l \alpha_i y_i b + \sum_{i=1}^l \alpha_i \quad (5.46)$$

olur. Burada $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_l)$, eğitim örneklerine ait *Lagrange Çarpanları* kümesidir.

Eşitlik 5.46'da ifade edilen Lagrange fonksiyonunun çözümü için Karush-Kuhn-Tucker (KKT) koşulları kullanılarak primal problem, dual problem haline dönüştürülür. Bu problem için KKT koşulları şunlardır;

Gradyent koşulu:

$$\frac{\partial L_p}{\partial w} = w^T - \sum_{i=0}^l \alpha_i y_i x_i = 0, \quad \frac{\partial L_p}{\partial w} = \left(\frac{\partial L_p}{\partial w_1}, \frac{\partial L_p}{\partial w_2}, \dots, \frac{\partial L_p}{\partial w_d} \right) \quad (5.47)$$

$$\frac{\partial L_p}{\partial b} = \sum_{i=0}^l \alpha_i y_i = 0, \quad (5.48)$$

$$\frac{\partial L_p}{\partial \alpha_i} = g_i(x) = -(y_i (w^T \cdot x_i + b) - 1) = 0. \quad (5.49)$$

Orthogonality koşulu:

$$\alpha_i g_i = -\alpha_i (y_i (w^T \cdot x_i + b) - 1) = 0, \quad i = 1, 2, \dots, l \quad (5.50)$$

Feasibility koşulu:

$$-y_i \cdot (w^T \cdot x_i + b) + 1 \leq 0, \quad i = 1, 2, \dots, l \quad (5.51)$$

Negatif-olmama koşulu:

$$\alpha_i \geq 0, \quad i = 1, 2, \dots, l \quad (5.52)$$

5.47, 5.48, 5.49, 5.50 eşitlikleri ve 5.51, 5.52 eşitsizlikleri KKT koşulları olarak bilinmektedir. Eşitlik 5.47 ve 5.48 Lagrange fonksiyonunda (Eşitlik 5.46'da) eşitliğin sağ tarafına yazılacak olursa , optimizasyon problemi dual biçime dönüşür. Elde edilen dual problem şu şekildedir (Eşitlik 5.53, 5.54 ve eşitsizlik 5.55):

$$\text{Maksimizasyon} \quad L_D = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j x_i^T \cdot x_j \quad (5.53)$$

$$\text{Kısıtlama:} \quad \sum_{i=1}^l \alpha_i y_i = 0 \quad (5.54)$$

$$\alpha_i \geq 0, \quad i = 1, 2, \dots, l \quad (5.55)$$

Bu dual problemin verilen kısıtlamalar altında çözümünü ikinci dereceden optimizasyon problemi ile gerçekleştirilir. Eşitlik 5.53 karesel programlama ile çözümlenerek α değerleri bulunur. w ise eşitlik 5.47'den elde edilir.

Dikkat edilirse burada her eğitim örneği için bir tane Lagrange çarpanının olduğu görülmektedir. Çözümde elde edilen Lagrange çarpanlarının büyük çoğunluğunun değeri sıfır olacaktır. Lagrange çarpanı $\alpha_i \geq 0$ olan x_i örnekleri *Destek Vektörleri* olarak isimlendirilir ve bu noktalarda kısıtlamalar aktiftir. Destek vektörleri A_1 veya A_2 aşırıdüzlemlerinin üzerinde yer alırlar. Lagrange çarpanı $\alpha_i = 0$ olan x_i örnekleri destek vektörü değildir ve bu noktalarda kısıtlamalar aktif değildir. Lagrange çarpanı sıfır olan bu örnekler ise A_1 veya A_2 aşırıdüzlemlerinin arka taraflarında kalan örneklerdir. b değeri de eşitlik 5.50'nin çözümü sonucu bulunur:

5.5.3.1.1.2. Doğrusal olarak ayrılama durumu

Örnekler doğrusal olarak tamamen ayrılabilir durumda değilse problemin çözümü için pozitif zayıflık değişkenleri, ξ , $i = 1, 2, \dots, l$ kullanılır. 5.37 ve 5.38 eşitsizlikleri (koşulları) bu zayıflık değişkenleri ile yeniden tanımlanacak olursa, yeni ifadeler 5.56 ve 5.57 eşitsizlikleri olacaktır.

$$y_i = +1 \text{ için, } w^T \cdot x_i + b \geq +1 - \xi_i \quad \xi_i \geq 0 \quad (5.56)$$

$$y_i = -1 \text{ için, } w^T \cdot x_i + b \leq -1 + \xi_i \quad \xi_i \geq 0 \quad (5.57)$$

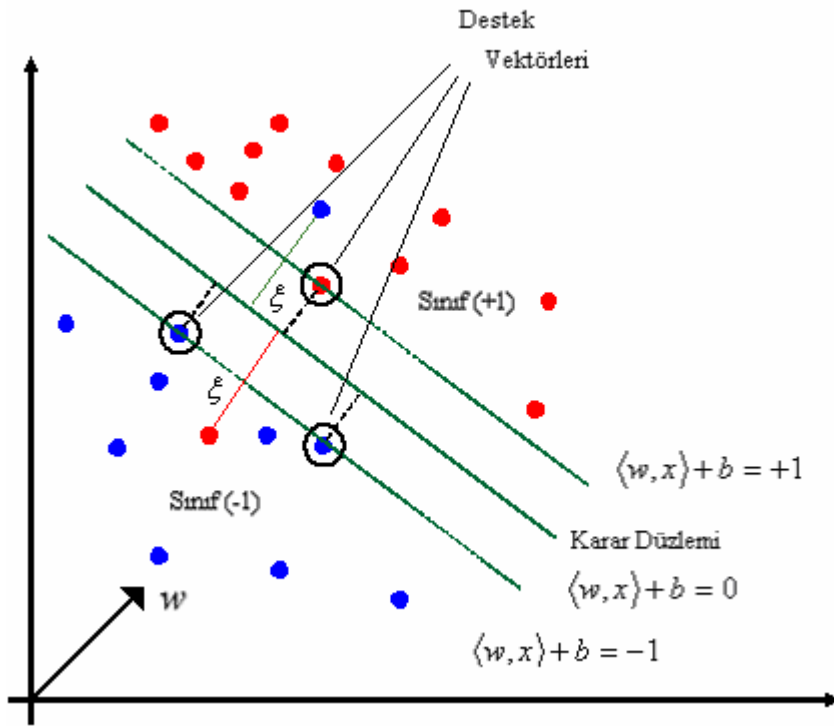
Buna göre problem şu şekilde yeniden tanımlanır ise,

$$\text{Minimizasyon} \quad f(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i \quad (5.58)$$

$$\text{Kısıtlamalar:} \quad y_i (w^T \cdot x_i + b) \geq +1 - \xi_i, \quad i = 1, 2, \dots, l \quad (5.59)$$

$$\xi_i \geq 0, \quad i = 1, 2, \dots, l \quad (5.60)$$

olur.



Şekil 5.15 Doğrusal ayırlamama durumunda optimal ayırıcı aşırıdüzlem

Doğrusal ayırlamama durumunda optimal ayırıcı aşırıdüzlem şekil 5.15’de görülmektedir. Doğrusal olarak ayırlamama durumunda sistemin ezberlemesini önlemek, yani sistemin eğitim verisi içinde olası her durum için bir çözüm üretmemesini sağlamak için, sisteme bir C üst sınırı eklenir. C parametresi, eğitim hataları için bir cezalandırma faktörü ve margin için de bir düzenleme parametresi olarak düşünülebilir. Bu üst sınır Lagrange çarpanlarının alabilecekleri maksimum değeri göstermektedir. Bu şekilde Lagrange çarpanlarının $0 \leq \alpha_i \leq C$ aralığında kalması sağlanmaktadır. Bu bilgilere göre Lagrange fonksiyonu yeniden şu şekilde ifade edilebilir;

$$L_p = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i - \sum_{i=1}^l \alpha_i \{y_i(x_i \cdot w + b) - 1 + \xi_i\} - \sum_{i=1}^l \mu_i \xi_i \quad (5.61)$$

Eşitlik 5.61’de, μ_i , ξ_i ’nin pozitif olmasını sağlamak için kullanılmış olan Lagrange parametreleridir. Problemin çözümü için KKT koşulları yeniden ele alınırsa;

Gradyent koşulu:

$$\frac{\partial L_p}{\partial w} = w^T - \sum_{i=0}^l \alpha_i y_i x_i = 0, \quad \frac{\partial L_p}{\partial w} = \left(\frac{\partial L_p}{\partial w_1}, \frac{\partial L_p}{\partial w_2}, \dots, \frac{\partial L_p}{\partial w_d} \right) \quad (5.62)$$

$$\frac{\partial L_p}{\partial b} = \sum_{i=0}^l \alpha_i y_i = 0, \quad (5.63)$$

$$\frac{\partial L_p}{\partial \xi_i} = C - \alpha_i - \mu_i = 0 \quad (5.64)$$

$$\frac{\partial L_p}{\partial \alpha_i} = g_i(x) = -(y_i(w^T \cdot x_i + b) - 1 + \xi_i) = 0. \quad (5.65)$$

Orthogonality koşulu:

$$\alpha_i g_i = -\alpha_i (y_i(w^T \cdot x_i + b) - 1 + \xi_i) = 0, \quad i = 1, 2, \dots, l \quad (5.66)$$

Feasibility koşulu:

$$-(y_i \cdot (w^T \cdot x_i + b) - 1 + \xi_i) \leq 0, \quad i = 1, 2, \dots, l \quad (5.67)$$

Negatif-olmama koşulu:

$$\alpha_i \geq 0, \quad i = 1, 2, \dots, l \quad (5.68)$$

$$\xi_i \geq 0, \quad i = 1, 2, \dots, l \quad (5.69)$$

$$\mu_i \geq 0, \quad i = 1, 2, \dots, l \quad (5.70)$$

$$\mu_i \xi_i = 0, \quad i = 1, 2, \dots, l \quad (5.71)$$

olur. 5.62, 5.63, 5.64, 5.65, 5.66, 5.71 eşitlikleri ve 5.67, 5.68, 5.69, 5.70 eşitsizlikleri KKT koşulları olarak bilinmektedir. KKT koşulları sonucu elde edilen 5.62, 5.63 ve 5.64 eşitlikleri Lagrange fonksiyonunda (Eşitlik 5.61'de) eşitliğin sağ tarafına yazılacak olursa , optimizasyon problemi dual biçime dönüşür. Elde edilen dual problem şu şekildedir (Eşitlik 5.72);

$$\text{Maksimizasyon} \quad L_D = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j x_i^T \cdot x_j \quad (5.72)$$

$$\text{Kısıtlama:} \quad 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, l \quad (5.73)$$

$$\sum_{i=1}^l \alpha_i y_i = 0, \quad i = 1, 2, \dots, l \quad (5.74)$$

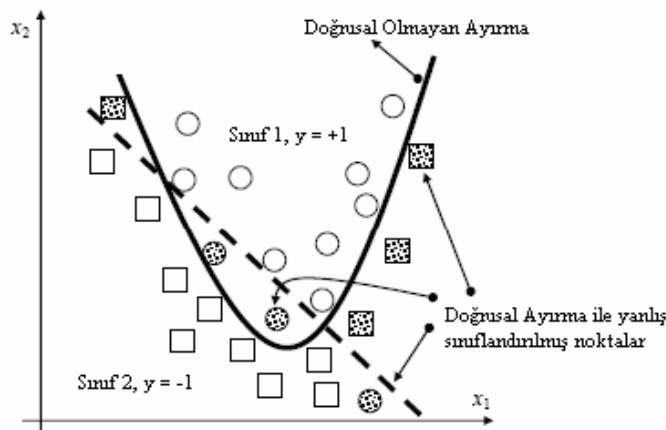
Eşitlik 5.74, dual problemin çözümünde kullanılacak eşitlik kısıtlamasıdır. Bu problemin çözümünde $0 \leq \alpha_i \leq C$ aralığında yer alan Lagrange çarpanlarına karşılık gelen x_i değerleri *Destek Vektörleridir*.

5.5.3.1.2. Doğrusal olmayan Destek Vektör Makinesi sınıflandırıcıları

Gerçek-dünya problemlerinin büyük çoğunluğu birçok farklı bileşenden oluşan problemlerdir ve bu problemlerde veriler doğrusal olmayan karar yüzeyleri ile modellenebilir. Doğrusal olmayan problemlerin çözümünü bulmanın yolu, çekirdek fonksiyonları ile örneklerin öncelikle daha yüksek boyutlu ve doğrusal olarak ayrılabilirleri bir uzaya taşıyıp, ardından çözümün bu yeni uzayda aranması ile mümkün olmaktadır.

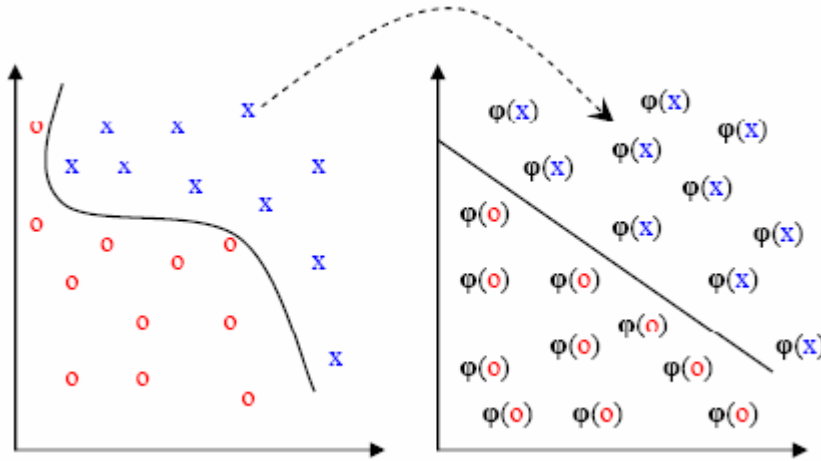
Giriş uzayındaki eğitim örneklerini bir H Euclid uzayına taşıyan Φ fonksiyonu göz önüne alınırsa,

$\Phi : \mathcal{R}^d \mapsto H$ olur.



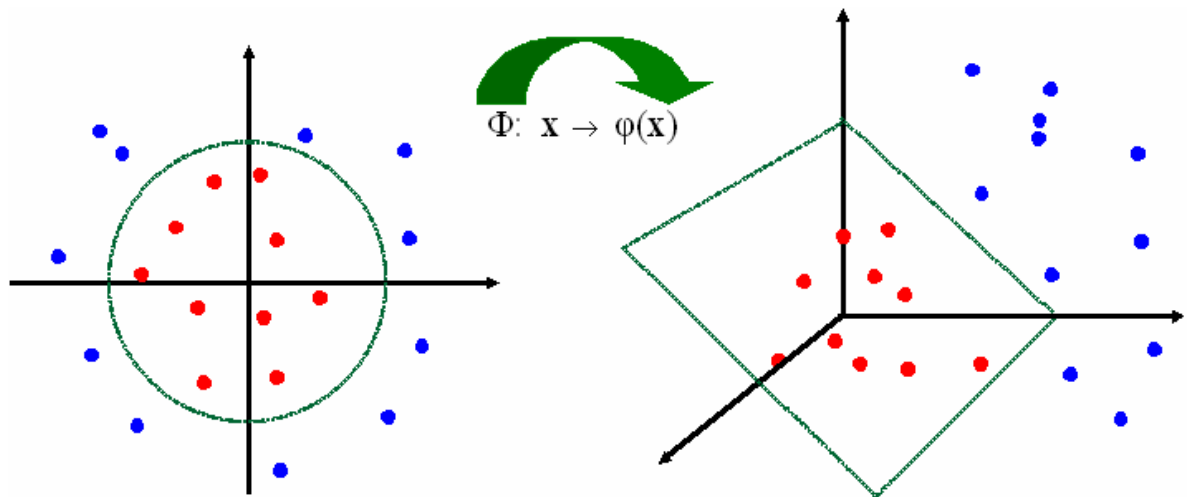
Şekil 5.16 Bir doğrusal olmayan Destek Vektör Makinesi

Şekil 5.16'da bir doğrusal olmayan SVM; şekil 5.17 ve 5.18'de ise bir giriş uzayını özellik uzayına eşleme yani bir giriş uzayını verilerin doğrusal olarak ayrılabilceği daha yüksek boyutlu bir özellik uzayına taşıma görülmektedir.



Şekil 5.17 Giriş uzayını özellik uzayına eşleme

Destek Vektör Makinelerini doğrusal olmayan problemlere uygularken kullanılan yöntemde temel fikir, doğrusal olarak ayrılamayan veriyi doğrusal olarak ayrılabilceği bir yüksek boyutlu özellik uzayına taşımaktır. Böylece en uygun ayırıcı aşırıdüzlem bu özellik uzayında bulunabilir. Giriş uzayındaki eğitim verisi vektörleri çekirdek fonksiyonları kullanılarak özellik uzayına aktarılır.



Şekil 5.18 Giriş uzayını özellik uzayına eşleme

Artık, Destek Vektör Makinelerinin eğitim aşaması, H uzayındaki verilerin $\Phi(x_i) \cdot \Phi(x_j)$ iç çarpımlarına bağlı olacaktır. Bu iç çarpım K ile gösterilir ise,

$$K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j) \quad (5.75)$$

olarak ifade edilir (Eşitlik 5.75). Burada K fonksiyonu *Çekirdek Fonksiyonu (Kernel Function)* olarak adlandırılmaktadır. Sonuç olarak karar fonksiyonu,

$$f(x) = \sum_{i=1}^{ls} \alpha_i y_i \Phi(x_i) \cdot \Phi(x) + b = \sum_{i=1}^{ls} \alpha_i y_i K(x_i, x) + b \quad (5.76)$$

fonksiyonunun (Eşitlik 5.76'nın) işareti ile belirlenir. Yani karar fonksiyonu yeniden yazılır ise,

$$\text{Karar fonksiyonu} = \text{sign}(f(x)) = \text{sign}\left(\sum_{i=1}^{ls} \alpha_i y_i K(x_i, x) + b\right) \quad (5.77)$$

olur (Eşitlik 5.77). Bu fonksiyonda ls , destek vektörlerinin sayısı, x_i de destek vektörleridir.

Çekirdek fonksiyonu olarak çeşitli fonksiyonlar kullanılabilir. Bunlardan birkaçı şunlardır:

Radyal tabanlı çekirdek (kernel) fonksiyonu (RBF):

Radyal tabanlı çekirdek fonksiyonu (RBF) eşitlik 5.78'de görülmektedir.

$$K(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (5.78)$$

Polinom çekirdek (kernel) fonksiyonu:

Polinom çekirdek fonksiyonu eşitlik 5.79'da görülmektedir.

$$K(x_i, x_j) = (x_i^T \cdot x_j)^d \quad (5.79)$$

Doğrusal çekirdek (kernel) fonksiyonu:

Doğrusal çekirdek fonksiyonu eşitlik 5.80'de görülmektedir.

$$K(x_i, x_j) = (x_i^T \cdot x_j) \quad (5.80)$$

Mercer Teoremi:

Bir fonksiyonun çekirdek fonksiyonu olarak tanımlanabilmesi için sağlaması gereken bazı koşullar vardır:

- $K(x, y) = K(y, x)$ (Çekirdek fonksiyonu simetrik olmalıdır)
- $\iint K(x, y) f(x) f(y), dx dy \geq 0$ 'dır.

5.5.4. Soft Margin Destek Vektör Makinesi sınıflandırıcıları

Örnekler doğrusal olarak tamamen ayrılabilir durumda değilse problemin çözümü için pozitif zayıflık değişkenleri, ξ , $i = 1, 2, \dots, l$ kullanılır. 5.37 ve 5.38 eşitsizliklerindeki koşullar pozitif zayıflık değişkenleri ile yeniden tanımlanacak olursa, yeni ifadeler,

$$y_i = +1 \text{ için, } w^T \cdot x_i + b \geq +1 - \xi_i \quad \xi_i \geq 0 \quad (5.81)$$

$$y_i = -1 \text{ için, } w^T \cdot x_i + b \leq -1 + \xi_i \quad \xi_i \geq 0 \quad (5.82)$$

şeklinde olacaktır (5.81 ve 5.82 eşitsizlikleri). Problem şu şekilde yeniden tanımlanır ise,

$$\text{Minimizasyon} \quad f(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i \quad (5.83)$$

$$\text{Kısıtlamalar:} \quad y_i (w^T \cdot x_i + b) \geq +1 - \xi_i, \quad i = 1, 2, \dots, l \quad (5.84)$$

$$\xi_i \geq 0, \quad i = 1, 2, \dots, l \quad (5.85)$$

olur (Eşitlik 5.83 ve eşitsizlik 5.84, 5.85).

Doğrusal ayrılamama durumunda optimal ayırıcı aşırıdüzlem şekil 5.15'de görülmektedir. Doğrusal olarak ayrılamama durumunda sistemin ezberlemesini önlemek, yani sistemin eğitim verisi içinde olası her durum için bir çözüm

üretmemesini sağlamak için, sisteme bir C üst sınırı eklenir. C parametresi, eğitim hataları için bir cezalandırma faktörü ve margin için de bir düzenleme parametresi olarak düşünülebilir. Bu üst sınır Lagrange çarpanlarının alabilecekleri maksimum değeri göstermektedir. Bu şekilde Lagrange çarpanlarının $0 \leq \alpha_i \leq C$ aralığında kalması sağlanmaktadır. Bu bilgilere göre Lagrange fonksiyonu yeniden şu şekilde ifade edilebilir (Eşitlik 5.86):

$$L_p = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i - \sum_{i=1}^l \alpha_i \{y_i(x_i \cdot w + b) - 1 + \xi_i\} - \sum_{i=1}^l \mu_i \xi_i \quad (5.86)$$

Yukarıdaki ifadede, μ_i , ξ_i 'nin pozitif olmasını sağlamak için kullanılmış olan Lagrange parametreleridir. Problemin çözümü için KKT koşulları yeniden ele alınırsa;

Gradyent koşulu:

$$\frac{\partial L_p}{\partial w} = w^T - \sum_{i=0}^l \alpha_i y_i x_i = 0, \quad \frac{\partial L_p}{\partial w} = \left(\frac{\partial L_p}{\partial w_1}, \frac{\partial L_p}{\partial w_2}, \dots, \frac{\partial L_p}{\partial w_d} \right) \quad (5.87)$$

$$\frac{\partial L_p}{\partial b} = \sum_{i=0}^l \alpha_i y_i = 0, \quad (5.88)$$

$$\frac{\partial L_p}{\partial \xi_i} = C - \alpha_i - \mu_i = 0 \quad (5.89)$$

$$\frac{\partial L_p}{\partial \alpha_i} = g_i(x) = -(y_i(w^T \cdot x_i + b) - 1 + \xi_i) = 0. \quad (5.90)$$

Orthogonality koşulu:

$$\alpha_i g_i = -\alpha_i (y_i (w^T \cdot x_i + b) - 1 + \xi_i) = 0, \quad i = 1, 2, \dots, l \quad (5.91)$$

Feasibility koşulu:

$$-(y_i \cdot (w^T \cdot x_i + b) - 1 + \xi_i) \leq 0, \quad i = 1, 2, \dots, l \quad (5.92)$$

Negatif-olmama koşulu:

$$\alpha_i \geq 0, \quad i = 1, 2, \dots, l \quad (5.93)$$

$$\xi_i \geq 0, \quad i = 1, 2, \dots, l \quad (5.94)$$

$$\mu_i \geq 0, \quad i = 1, 2, \dots, l \quad (5.95)$$

$$\mu_i \xi_i = 0, \quad i = 1, 2, \dots, l \quad (5.96)$$

olur. 5.87, 5.88, 5.89, 5.90, 5.91, 5.96 eşitlikleri ve 5.92, 5.93, 5.94, 5.95 eşitsizlikleri KKT koşulları olarak bilinmektedir. 5.87, 5.88 ve 5.89 eşitlikleri Lagrange fonksiyonunda (Eşitlik 5.86'da) eşitliğin sağ tarafına yazılacak olursa, optimizasyon problemi dual biçime dönüşür. Elde edilen dual problem şu şekildedir (Eşitlik 5.97, 5.99 ve eşitsizlik 5.98);

$$\text{Maximizasyon} \quad L_D = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j x_i^T \cdot x_j \quad (5.97)$$

$$\text{Kısıtlama:} \quad 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, l \quad (5.98)$$

$$\sum_{i=1}^l \alpha_i y_i = 0, \quad i = 1, 2, \dots, l \quad (5.99)$$

Bu problemin çözümünde $0 \leq \alpha_i \leq C$ aralığında yer alan Lagrange çarpanlarına karşılık gelen x_i değerleri *Destek Vektörleridir*.

5.5.5. En Küçük Kareler Destek Vektör Makinesi sınıflandırıcıları

Destek Vektör Makineleri, veri sınıflandırma ve fonksiyon kestirimi için yeni güçlü araçlardır. SVM, giriş verisini doğrusal bir şekilde ayrılabilirliği bir yüksek boyutlu özellik uzayına taşırlar. Son zamanlarda SVM biyomühendislik ve örüntü tanıma gibi birçok çeşitli alanda başarıyla uygulanmıştır. Bunun bir sebebi, destek vektör makinelerinin sıklıkla önceki yöntemlere göre daha iyi sonuç üretmesidir. Çünkü SVM, yapısal riski minimize etmek için tasarlanmıştır, oysa ki önceki tekniklerin çoğu deneysel riski minimize etme temeli üzerinde kuruludurlar.

Destek Vektör Makinelerindeki eğitim problemi bir konveks karesel programlama probleminin çözümüne indirgenebilir. Ancak bunun birçok avantajına rağmen ortaya çıkan bir problem de şudur ki, karesel programlama problemindeki matrisin boyutu eğitim noktalarının sayısı ile orantılıdır. Öyle ki, hayli büyük veri kümeleri için karesel programlama paketi kullanılamaz hale gelebilmektedir.

Bu sebeple Suykens ve Vandewalle tarafından (1999) sınıflandırma için destek vektör makinelerinin bir değiştirilmiş versiyonu sunuldu ve bu versiyona *En Küçük Kareler (LS) SVM* ismi verildi. LS SVM, çözümü, karesel programlama problemine karşılık bir doğrusal denklem kümesinde aramaktaydı. N sayıda veri noktası için bir LS SVM sınıflandırıcısı bir $(N+1) \times (N+1)$ boyutunda bir kare matrisin tersini alarak işlem yapmaktaydı.

5.5.5.1. İkili Sınıf Tahmini için Standart Destek Vektör Makineleri

$\{x_i, y_i\}_{i=1}^N$, verilen bir eğitim veri kümesi olsun. Burada $x_i \in \mathcal{R}^n$ i.giriş vektörünü ve $y_i \in \{-1, +1\}$ bilinen ikili hedefleri ifade eden kavramlardır. Buna göre, standart SVM sınıflandırıcısı eşitsizlik 5.100 ve 5.101'de verilen şartları sağlar.

$$w^T \Phi(x_i) + b \geq +1 \quad y_i = +1 \quad \text{için,} \quad (5.100)$$

$$w^T \Phi(x_i) + b \leq -1 \quad y_i = -1 \quad \text{için.} \quad (5.101)$$

Yani diğer ifadeyle,

$$y_i (w^T \Phi(x_i) + b) \geq +1 \quad i = 1, 2, \dots, N \quad (5.102)$$

olur (5.102 eşitsizliği).

$$K(x_i, x_j) = \Phi(x_i)^T \cdot \Phi(x_j) \quad (5.103)$$

5.103 eşitliği çekirdek (kernel) fonksiyonudur.

Burada $\Phi : \mathcal{R}^n \mapsto \mathcal{R}^m$ olan bir dönüşüm olup giriş uzayını doğrusal bir şekilde ayrılabilir özellik uzayına eşler. Böylece sınıflandırma fonksiyonu şu şekilde verilebilir (Eşitlik 5.104):

$$y(x) = \text{sign}\{w^T \Phi(x) + b\} \quad (5.104)$$

5.102 eşitsizliğine gevşek değişkenler ilave edilirse yeni ifadeler,

$$y_i (w^T \Phi(x_i) + b) \geq +1 - \xi_i, \quad i = 1, 2, \dots, N \quad (5.105)$$

$$\xi_i \geq 0, \quad i = 1, 2, \dots, N \quad (5.106)$$

olur (Eşitsizlik 5.105 ve 5.106). Böylece minimizasyon problemi,

$$\text{Minimizasyon} \quad f(w, \xi) = \frac{1}{2} w^T w + C \sum_{i=1}^N \xi_i \quad (5.107)$$

$$\text{Kısıtlamalar:} \quad y_i (w^T \cdot \Phi(x_i) + b) \geq +1 - \xi_i, \quad i = 1, 2, \dots, N \quad (5.108)$$

$$\xi_i \geq 0, \quad i = 1, 2, \dots, N \quad (5.109)$$

olur (5.107 eşitliği ve 5.108, 5.109 eşitsizliği). Burada C , eğitim hataları ile margin arasındaki ölçüyü kontrol eden bir pozitif sabit parametreyi ve f de, verilen kısıtlamalar altında minimize edilmeye çalışılan amaç fonksiyonunu ifade etmektedir. Sistemin bu şekline *Primal Form* denir ve sistemi *Dual forma* dönüştürerek çözmek daha kolay olduğu için sistemin dual formuna geçiş yapılır. Sistemin dual formu Lagrange yöntemi ile KKT koşulları kullanılarak elde edilir. Böylece problemin çözümü elde edilmiş olur.

5.5.5.2. En Küçük Kareler Destek Vektör Makinesi sınıflandırıcıları

Suykens ve Vandewalle, (1999) Vapnik'in standart SVM sınıflandırıcısını *En Küçük Kareler Destek Vektör Makinesi sınıflandırıcısı* biçimine dönüştürdüler;

$$\text{Minimizasyon} \quad J(w, e) = w^T w + \sum_{i=1}^N e_i^2 \quad (5.110)$$

$$\text{Kısıtlamalar:} \quad y_i (w^T \cdot \Phi(x_i) + b) = +1 - \xi_i, \quad i = 1, 2, \dots, N \quad (5.111)$$

Eşitlik 5.110'daki problemin çözümü için Lagrange yöntemi kullanılarak,

$$L(w, b, e; \alpha) = J(w, e) - \sum_{i=1}^N \alpha_i \{y_i (w^T \Phi(x_i) + b) - 1 + e_i\} \quad (5.112)$$

eşitliği elde edilir. Burada α_i Lagrange çarpanlarıdır.

Eşitlik 5.112'nin çözümü için optimallik koşulları kullanılarak aşağıdaki $(N+1) \times (N+1)$ boyutundaki doğrusal denklem sistemi elde edilir (Eşitlik 5.113).

$$\begin{bmatrix} 0 & Y^T \\ Y & ZZ^T + \gamma^{-1}I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (5.113)$$

Burada,

$$Z^T = [y_1 \Phi(x_1), \dots, y_N \Phi(x_N)] \quad (5.114)$$

$$Y^T = [y_1, y_2, \dots, y_N] \quad (5.115)$$

$$1^T = [1, \dots, 1] \quad (5.116)$$

$$\alpha^T = [\alpha_1, \alpha_2, \dots, \alpha_N] \quad (5.117)$$

$$ZZ^T = y_i y_j \Phi(x_i) \Phi(x_j) \quad (5.118)$$

olur.

Eşitlik 5.113'deki doğrusal denklem sisteminin çözülmesi ile,

$$bY + (\gamma^{-1}I + ZZ^T)\alpha = 1 \quad (5.119)$$

5.119 eşitliği elde edilir. $A = \gamma^{-1}I + ZZ^T$ olmak üzere, $bY + A\alpha = 1$ olur. Sonuç olarak,

$$\alpha = A^{-1}(1 - bY) \quad (5.120)$$

$$b = \frac{Y^T A^{-1} 1}{Y^T A^{-1} Y} \quad (5.121)$$

5.120 ve 5.121 eşitlikleri elde edilir. Görüldüğü gibi En Küçük Kareler SVM sınıflandırıcılarında α ve b parametreleri, doğrusal denklem sisteminin çözümünden elde edilmektedir. Çekirdek fonksiyonu kullanılarak karar fonksiyonu yani dual model şu şekilde yazılabilir;

$$f(x_{yeni}) = \text{sign} \left(\sum_{i=1}^{DVS} \alpha_i y_i K(x_i^{DV}, x_{yeni}) + b \right) \quad (5.122)$$

Eşitlik 5.122'deki DVS , destek vektör sayısını ifade etmektedir. Buradaki çekirdek fonksiyonu,

$$K(x_1, x_2) = \langle \Phi(x_1), \Phi(x_2) \rangle \quad (5.123)$$

5.123 eşitliği ile ifade edilir. Görüldüğü gibi, En Küçük Kareler SVM sınıflandırıcılarında da yerel çözümlere takılma problemi bulunmamaktadır ve En Küçük Kareler SVM sınıflandırıcılarında çözüm, karesel programlama yöntemi ile değil de doğrusal denklem sisteminin çözümlenmesi ile elde edildiği için, uzun eğitim zamanı gerektirmemektedir. Ancak En Küçük Kareler SVM sınıflandırıcılarının standart SVM sınıflandırıcılarına göre zayıflığı da şu olmaktadır ki, bu yöntemde destek vektörlerinin sayısı tüm eğitim verisidir.

6. UYGULAMA SONUÇLARI

Gerçekleştirilen uygulama ses kodlama ve ses tanıma bölümlerinden oluşmaktadır.

Gerçekleştirilen uygulama 20 adet ses ifadesini (sözcüğünü) içermektedir. Bu sözcüklerin her birinden 10'ar adet kaydedilerek, önce sistemin eğitim aşaması için toplam 200 örüntü oluşturulmuştur. Ardından yine aynı şekilde bu sözcüklerin her birinden 10'ar adet kaydedilerek, sistemin test aşaması için de 200 örüntü oluşturulmuştur.

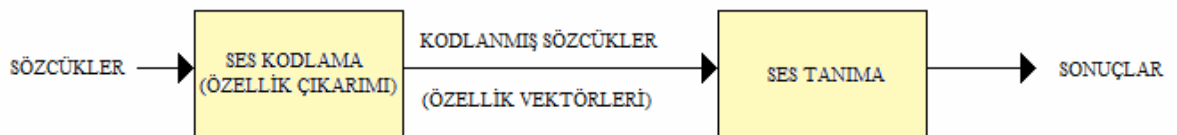
Sözcüklerin kaydedilmesi işleminde düşük gürültülü bir standart mikrofon kullanılmıştır. Bilgisayar ortamındaki kayıt aracı, Windows XP işletim sisteminde yer alan ses kaydedicisidir. Ses kaydedicisi ile sözcüklerin bilgisayar ortamına kaydı yapılırken, örnekleme frekansı 8000 Hz (Hertz) yani 8 KHz seçilmiş ve kayıt işlemi mono olarak yapılmıştır. Bu sözcükler sabit diskte belirlenen bir klasör içerisinde wav uzantılı birer ses dosyası olarak kaydedilmiştir. Tüm kayıtlar iki saniye uzunluğundadır, yani sistemin hem eğitim ve hem de test aşaması için kullanılacak olan bu kayıtların (kaydedilen sözcüklerin) tümü iki saniye uzunluğundadır. Dolayısıyla her bir kaydedilen sözcük 16000 örnekten (veriden) oluşmaktadır. Diğer ifadeyle her bir kaydedilen sözcük 16000x1 boyutunda 16000 örnek (veri) içeren bir sütun vektörüdür. Kaydedilen sözcükler tablo 6.1'de görülmektedir.

Gerçekleştirilen uygulamanın ses kodlama bölümünde kodlama yöntemi olarak Doğrusal Önkestirim Kodlama (LPC) ve ses tanıma bölümünde ise tanıma yöntemi olarak Destek Vektör Makineleri sınıflandırıcılarından olan SM SVM sınıflandırıcısı ve LS SVM sınıflandırıcısı kullanılmıştır. Uygulamanın genel yapısı şekil 6.1'de görülmektedir.

Tablo 6.1 Bilgisayar ortamına mikrofon ile kaydedilen sözcükler

ÖRÜNTÜ NUMARASI	SÖZCÜK
1	Aç
2	Araçlar
3	Ayrıntılar
4	Çıkış
5	Döşeme
6	Dosya
7	Düzen
8	Gönder
9	Görünüm
10	Kes
11	Kopyala
12	Küçük
13	Listele
14	Seç
15	Simge
16	Yapıştır
17	Yardım
18	Yazdır
19	Yeni
20	Yenile

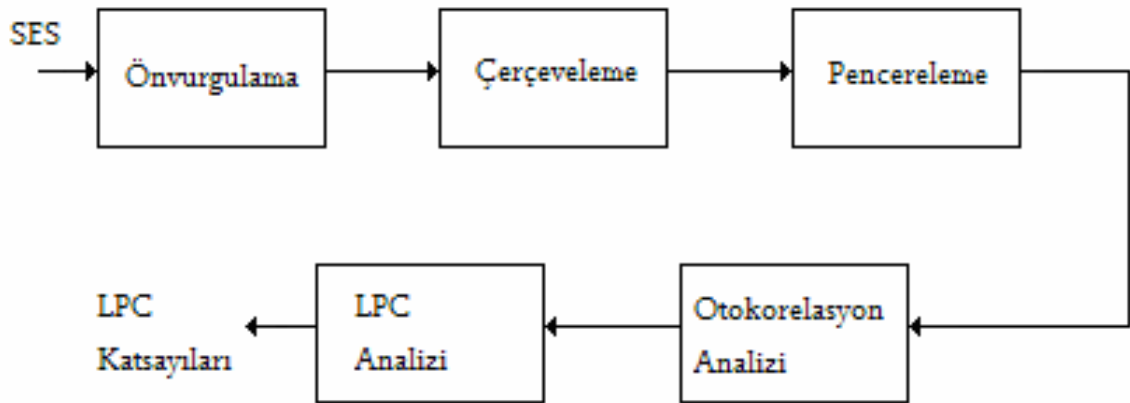
Tablo 6.1’de verilen sözcükler, bilgisayar işlemlerinde sıklıkla kullanılan komutlar olduğu için, bu sözcüklerin kullanılmasına karar verildi.

**Şekil 6.1** Uygulamanın genel yapısı

Uygulamada, kaydedilen sözcüklerin çok miktarda veri içermesinden dolayı (her bir sözcüğün 16000 veriden oluştuğu daha önce ifade edilmişti), öncelikle bu sözcüklerin kodlanarak daha az sayıda veri içeren ve sesin özelliklerini oluşturan özellik vektörlerine dönüştürülmesi gerekir. Burada ses kodlama modülüne aynı zamanda özellik çıkarımı modülü de denilebilir. Özellik çıkarımı modülünde yöntem olarak Doğrusal Önkestirim Kodlama (LPC, Linear Predictive Coding) yöntemi kullanıldı. Bu yöntem sesli ifadeye (yani sözcüğe) ait özelliklerin çıkarılması işlemi oldukça kullanışlı ve başarılı sonuçlar veren bir yöntem olarak günümüzde de ses kodlama

işlemlerinde sıkça kullanılmaktadır. Yöntemin teorik yapısı 4.Bölümde anlatıldığı için tekrar bu bölümde bu konuya değinilmeyecektir. Özellik çıkarımı bölümünde gerçekleştirilen işleme LPC analizi de denilebilir, çünkü bu bölümde ses sinyali aslında analiz edilerek, ses sinyalini temsil eden bir parametrik ifade oluşturulmaya çalışılmaktadır.

LPC analizindeki aşamalar şekil 6.2’de verilmiştir.



Şekil 6.2 LPC analizindeki aşamalar

Uygulama Matlab 5.3 yazılımı ile oluşturulmuştur. Bunun sebebi, uygulamanın ses tanıma bölümünün yazılması aşamasında Matlab 5.3’ün karesel programlama fonksiyonu olan qp fonksiyonunun kullanım kolaylığıdır.

Uygulamanın ses kodlama bölümünde LPC analizini gerçekleştiren Matlab fonksiyonu için giriş parametreleri şunlardır:

- Ses vektörü,
- Örnekleme frekansı,
- LPC analizinin derecesi,
- Çerçeve uzunluğu,
- Çerçeve yenilenme zamanı,
- Önvurgulama’dır.

Ses vektörü, mikrofon ile bilgisayar ortamına wav uzantılı bir ses dosyası olarak kaydedilen 16000x1 boyutundaki 16000 örnek (veri) içeren iki saniye uzunluğundaki sütun vektörüdür.

Örnekleme frekansı, sözcükleri Windows XP işletim sistemindeki Ses Kaydedicisi ile kaydederken seçilen öznitelik değeridir ve algoritmadaki değeri Hertz cinsinden 8000'dir.

LPC analizinin derecesi, uygulamada 8, 10, 12, 13 ve 16 seçilerek tüm sözcükler için ayrı ayrı LPC analizi gerçekleştirilmiş, her seçilen değer için uygulamanın başarı durumu gözlenmiştir.

Çerçeve uzunluğu, milisaniye cinsinden 30 olarak belirlenmiş ve her bir 30 milisaniye uzunluğundaki çerçeve için bir dizi LPC parametre kümesi hesaplanmıştır. Kullanılan pencere fonksiyonu ise Hamming olmuştur.

Çerçeve yenilenme zamanı, LPC analizi her bir çerçeve için yapıldığı için, ses sinyali boyunca her bir çerçevenin yeniden oluşumu için geçen süredir, algoritmada milisaniye cinsinden 20 ve 25 değerleri seçilerek tüm sözcükler için ayrı ayrı LPC analizi gerçekleştirilmiş, her seçilen değer için uygulamanın başarı durumu gözlenmiştir.

Ön vurgulama, bir sayısal bir-sıfır filtresi içindeki epsilon değişkenidir. Ses sinyaline ön vurgulamayı gerçekleştirmeyi sağlayan bu değişkenin varsayılan değeri 0.9378 olarak belirlenmiştir.

Uygulamanın ses kodlama bölümünde LPC analizini gerçekleştiren Matlab fonksiyonu için çıkış parametreleri, LPC katsayılarıdır.

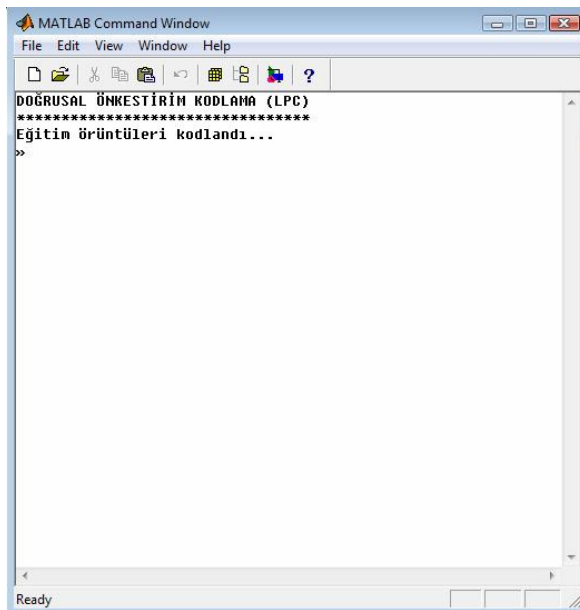
LPC katsayıları, LPC analizi sonuçlarıdır, yani a_i lerdir. Ses sinyalinin her bir çerçevesi için bir dizi LPC katsayısı hesaplanmaktadır.

Uygulamanın ses kodlama bölümünde LPC analizi yapılırken, ses sinyali analiz edilerek yani ses sinyalini temsil eden LPC katsayıları hesaplanarak ses sinyaline ilişkin özellik çıkarımı işlevi gerçekleştirilmiş olur. Böylece ses sinyali, yani her bir kaydedilen

sözcük LPC yöntemi ile kodlanmış olur. Artık ses tanıma modülü için giriş veri kümeleri hazır hale gelmiş demektir. Bundan sonraki aşama ses tanıma aşamasıdır.

Uygulamanın ses kodlama aşamasında kodlama (özellik çıkarımı) için LPC, LPCC, MFCC, PLP yöntemlerinin herhangi biri kullanılabilir.

Eğitim ve test örüntülerinin LPC yöntemi ile kodlanması işlemi, sırasıyla şekil 6.3 ve şekil 6.4’de görülmektedir.

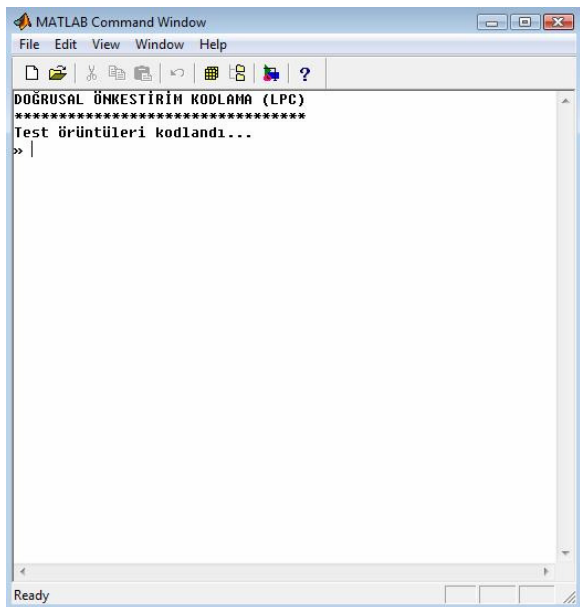


```

MATLAB Command Window
File Edit View Window Help
DOĞRUSAL ÖNKESTİRİM KODLAMA (LPC)
*****
Eğitim örüntüleri kodlandı...
>>
Ready

```

Şekil 6.3 Eğitim örüntülerinin LPC yöntemi ile kodlanması



```

MATLAB Command Window
File Edit View Window Help
DOĞRUSAL ÖNKESTİRİM KODLAMA (LPC)
*****
Test örüntüleri kodlandı...
>> |
Ready

```

Şekil 6.4 Test örüntülerinin LPC yöntemi ile kodlanması

Gerçekleştirilen uygulamanın ses tanıma bölümü iki modülden oluşmaktadır. Bunlar, *Eğitim* ve *Test* modülleridir.

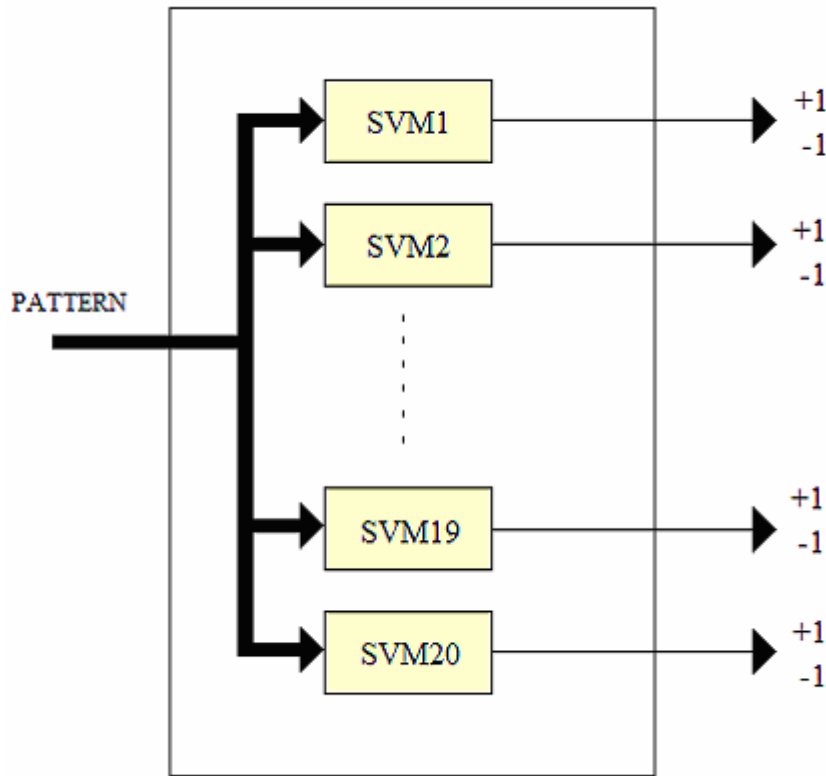
20 adet ses ifadesinin (sözcüğünün) her birinden 10'ar adet kaydedilerek, önce sistemin eğitim aşaması için toplam 200 örüntü oluşturulmuştur. Bu örüntüler önceki ses kodlama aşamasında kodlanarak her bir örüntüye ilişkin özellik vektörleri çıkarılmıştır. Bu özellik vektörleri eğitim aşaması için giriş veri kümesi olarak kullanılacaktır. Ardından yine aynı şekilde 20 adet ses ifadesinin (sözcüğünün) her birinden 10'ar adet kaydedilerek, sistemin test aşaması için toplam 200 örüntü oluşturulmuştur. Bu örüntüler önceki ses kodlama aşamasında kodlanarak her bir örüntüye ilişkin özellik vektörleri çıkarılmıştır. Bu özellik vektörleri de test aşaması için giriş veri kümesi olarak kullanılacaktır.

Uygulamanın ses tanıma bölümünde kullanılan yöntem, Destek Vektör Makinesi sınıflandırıcılarından SM SVM sınıflandırıcısı ve LS SVM sınıflandırıcısıdır. Sistem, hem SM SVM ve hem de LS SVM sınıflandırıcısı için 2 ayrı şekilde tasarlanmıştır. Her iki tasarım eğitilerek ve test edilerek sistemin başarı durumu gözlenmiştir.

Kodlanmış olan örüntüler her iki yöntem ile oluşturulan SVM sınıflandırıcılarına eğitim aşamasında giriş veri kümesi olarak verilir. SVM sınıflandırıcıları, sistemin modellemesini yapmak üzere α katsayılarını hesaplar. Bu hesaplama, SM SVM sınıflandırıcısında QP (karesel programlama); LS SVM sınıflandırıcısında doğrusal denklem sisteminin çözümü ile yapılır. Sistemin modellemesi yapıldıktan sonra eğitim hataları bulunur. Her iki SVM sınıflandırıcısında da çekirdek fonksiyonu olarak radyal tabanlı sigma (σ) genişlikli fonksiyon (RBF) kullanılmaktadır. σ parametrelerinin aldığı değerlere göre sistemin eğitim ve test hataları değişmektedir. Yapılan parametre tarama işlemleri sonucunda en iyi olabilecek sigma (σ) değerleri bulunmuştur. En iyi olma durumu eğitim hatalarının sıfır olduğu ve test hatalarının da minimum (sıfır) olduğu durumlardır. Burada amaçlanan, hem eğitim ve hem de test hatalarını sıfır yapabilmektedir. Böylece optimum (en iyi) ses tanıma sistemi elde edilmiş olacaktır.

Sistemin tasarımında her örüntü için bir adet destek vektör makinesi kullanıldı, dolayısıyla sistem, 20 örüntü için 20 adet destek vektör makinesinden oluşmaktadır. Sistemin çalışması şu şekildedir; sistem önce eğitilir, ardından eğitilmiş sistemi test etmek için örneğin, sisteme 3 numaralı örüntü giriş veri kümesi olarak verildiğinde, bu

örüntüye ilişkin 3 numaralı destek vektör makinesi +1 çıkışı üretecek, diğer destek vektör makineleri -1 çıkışı üretecektir, böylece bu örüntü sistem tarafından tanınmış olacaktır. Bu durum dışında oluşan durumların tümü örüntünün sistem tarafından tanınmadığını gösterir. Eğitilmiş sistemin test aşamasındaki işleyişi şekil 6.5’de görülmektedir.



Şekil 6.5 Test aşamasında sistemin çalışmasının şematik olarak gösterimi

Yapılan çalışmalarda sistem önce eğitilmiş ve ardından test edilmiştir. Sistemin eğitimi aşamasında 200 adet eğitim örüntüsü kullanılmıştır. Aynı şekilde sistemin test edilmesi aşamasında da 200 adet test örüntüsünün her biri sisteme teker teker verilmiş ve sistemin bu örüntüleri tanıyıp tanımadığı test edilmiştir. Bulunan en iyi parametreler ile sistem, test aşamasında SM SVM sınıflandırıcısı ile yapılan çalışmada %91’lik bir başarı ile örüntüleri tanımış; LS SVM sınıflandırıcısı ile yapılan çalışmada %71’lik bir başarı ile örüntüleri tanımıştır.

SM SVM ve LS SVM sınıflandırıcılarının eğitilmesi işlemi, sırasıyla şekil 6.6 ve şekil 6.7’de görülmektedir.


```

MATLAB Command Window
File Edit View Window Help
SM SVM SINIFLANDIRICISI ile SES TANIMA
Eğitim
Örüntü sayısı: 200
*****
1.SVM
Hatalı örnek sayısı: 0
Destek vektor sayısı: 200
-----
2.SVM
Hatalı örnek sayısı: 0
Destek vektor sayısı: 200
-----
3.SVM
Hatalı örnek sayısı: 0
Destek vektor sayısı: 200
-----
4.SVM
Hatalı örnek sayısı: 0
Destek vektor sayısı: 200
-----
5.SVM
Hatalı örnek sayısı: 0
Destek vektor sayısı: 200
*****
6.SVM
Hatalı örnek sayısı: 0
Destek vektor sayısı: 200
-----
Ready

```

Şekil 6.6 SM SVM sınıflandırıcısının eğitim aşaması

```

MATLAB Command Window
File Edit View Window Help
LS SVM SINIFLANDIRICISI ile SES TANIMA
Eğitim
Örüntü sayısı: 200
*****
1.SVM
Destek vektoru sayısı: 200
Hatalı örnek sayısı (Eğitim Hatası): 0
-----
2.SVM
Destek vektoru sayısı: 200
Hatalı örnek sayısı (Eğitim Hatası): 0
-----
3.SVM
Destek vektoru sayısı: 200
Hatalı örnek sayısı (Eğitim Hatası): 0
-----
4.SVM
Destek vektoru sayısı: 200
Hatalı örnek sayısı (Eğitim Hatası): 0
-----
5.SVM
Destek vektoru sayısı: 200
Hatalı örnek sayısı (Eğitim Hatası): 0
*****
6.SVM
Destek vektoru sayısı: 200
Hatalı örnek sayısı (Eğitim Hatası): 0
-----
Ready

```

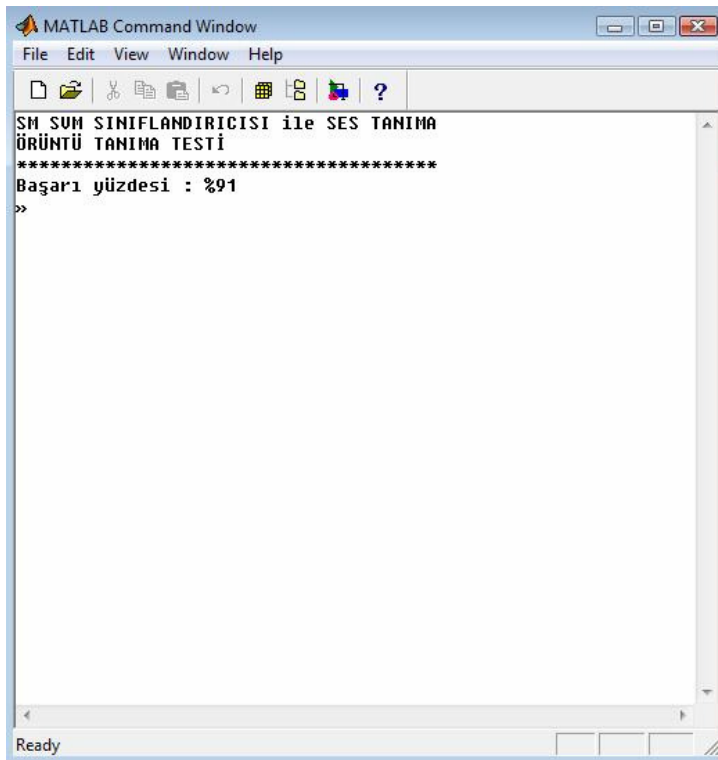
Şekil 6.7 LS SVM sınıflandırıcısının eğitim aşaması

Sistemin başarı durumunu etkileyen parametreler, hem örüntülerin kodlanması aşamasında ve hem de örüntülerin tanınması aşamasında bulunmaktadır. Bu

parametreler, ses kodlama aşaması için LPC algoritması içerisinde yer alan, LPC analizinin derecesi ve çerçeve yenilenme zamanı; ses tanıma aşaması için ise, SM SVM sınıflandırıcısında radyal tabanlı sigma genişlikli çekirdek fonksiyonu içerisinde yer alan sigma (σ) ve SM SVM sınıflandırıcısının kendi içerisinde yer alan C parametresi; LS SVM sınıflandırıcısında da radyal tabanlı sigma genişlikli çekirdek fonksiyonu içerisinde yer alan sigma (σ) parametresi ve LS SVM sınıflandırıcısının kendi içerisinde yer alan gama (γ) parametresidir.

Ses tanıma aşamasındaki SM SVM sınıflandırıcısı için, sigma (σ) parametrelerinin tümü 0.1 ve C parametresi 100 olarak belirlenmiştir.

En iyi örüntü tanıma başarısı sağlayan parametre değerleri kullanılarak gerçekleştirilen, SM SVM ve LS SVM sınıflandırıcılarının test edilmesi işlemi, sırasıyla şekil 6.8 ve şekil 6.9’da görülmektedir. Bu parametre değerleri tablo 6.2 ve tablo 6.3’de verilmiştir.

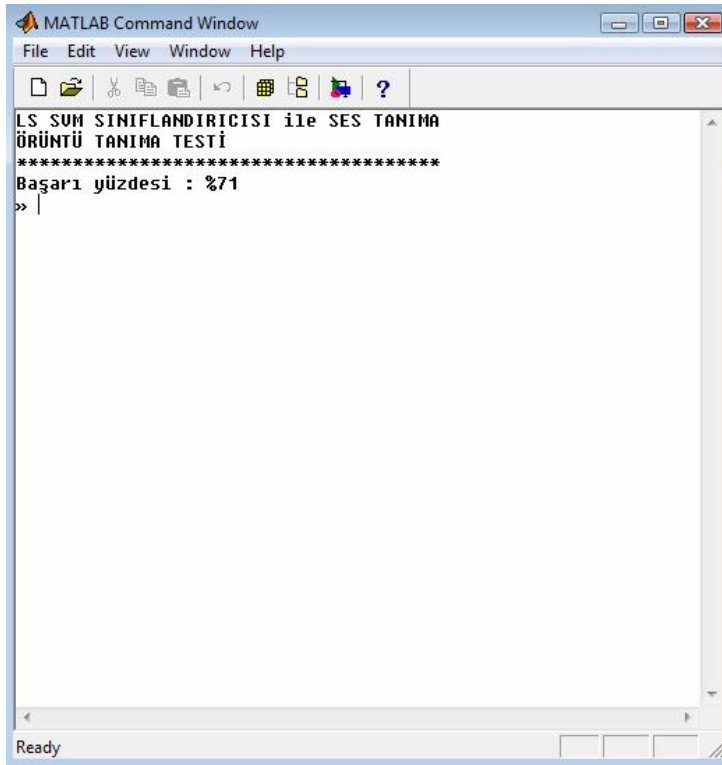


```

MATLAB Command Window
File Edit View Window Help
SM SVM SINIFLANDIRICISI İLE SES TANIMA
ÖRÜNTÜ TANIMA TESTİ
*****
Başarı yüzdesi : %91
>>
Ready

```

Şekil 6.8 SM SVM sınıflandırıcısının test aşaması



Şekil 6.9 LS SVM sınıflandırıcısının test aşaması

Ses kodlama aşamasındaki LPC algoritması içerisinde yer alan, LPC analizinin derecesi (L) ve çerçeve yenilenme zamanı (fr) parametrelerinin ve ses tanıma aşamasındaki radyal tabanlı sigma genişlikli çekirdek fonksiyonu içerisinde yer alan sigma (σ) ve SM SVM sınıflandırıcısının kendi içerisinde yer alan C parametresinin aldığı değerlere göre, SM SVM sınıflandırıcısının test aşamasındaki örüntü tanıma başarısı tablo 6.2’de verilmiştir.

Tablo 6.2 SM SVM sınıflandırıcısının örüntü tanıma başarısı

L	fr (milisaniye)	σ ve C değerleri	Tanıma Başarısı
8	20	$\sigma_1, \sigma_2, \dots, \sigma_{20} = 0.1$ $C = 100$	% 90
8	25		% 89.5
10	25		% 88
12	20		% 91
12	25		% 89.5
13	20		% 91
13	25		% 90
16	20		% 91
16	25		% 88.5

Ses kodlama aşamasındaki LPC algoritması içerisinde yer alan, LPC analizinin derecesi (L) ve çerçeve yenilenme zamanı (fr) parametrelerinin ve ses tanıma aşamasındaki radyal tabanlı sigma genişlikli çekirdek fonksiyonu içerisinde yer alan sigma (σ) parametresinin aldığı değerlere göre, LS SVM sınıflandırıcısının test aşamasındaki örüntü tanıma başarısı tablo 6.3'te verilmiştir.

Tablo 6.3 LS SVM sınıflandırıcısının örüntü tanıma başarısı

L	fr (milisaniye)	σ değerleri	Tanıma Başarısı
8	20	$\sigma_1, \sigma_2, \dots, \sigma_{20} = 0.1$	% 67.5
8	25		% 71
10	25		% 69
12	20		% 64.5
12	25		% 70
13	20		% 62.5
13	25		% 70
16	20		% 60.5
16	25		% 68.5

Uygulama sonuçları değerlendirildiği zaman, bulunan en iyi parametreler ile sistemin, test aşamasında SM SVM sınıflandırıcısı ile yapılan çalışmada %91'lik bir başarı ile örüntüleri tanımış olduğu; LS SVM sınıflandırıcısı ile yapılan çalışmada da %71'lik bir başarı ile örüntüleri tanımış olduğu gözlenmiştir.

7. SONUÇ VE ÖNERİLER

Gerçekleştirilen uygulamada elde edilen örüntü tanıma başarısı SM SVM sınıflandırıcısı için %91; LS SVM sınıflandırıcısı için de %71 olmuştur.

Tez kapsamında yapılan araştırmalarda kullanılan veriler, uygulanan yöntemler, elde edilen sonuçlar hakkında yapılan değerlendirmeler şunlardır:

Gerçekleştirilen uygulamanın ses kodlama modülünde kullanılan yöntem Doğrusal Önkestirim Kodlama; ses tanıma modülünde kullanılan yöntemler de SM SVM sınıflandırıcısı ve LS SVM sınıflandırıcısı olmuştur.

Başlangıçta giriş veri kümelerini elde etmek üzere, 200 adet örüntü sistemin eğitim aşaması için, 200 adet örüntü de sistemin test aşaması için kaydedilmiştir. Yani toplam 400 adet örüntü (sözcük) kaydı yapılmıştır.

Her bir örüntü ses kodlama modülünde Doğrusal Önkestirim Kodlama yöntemi ile kodlanmış ve ardından elde edilen veriler ses tanıma modülüne giriş veri kümeleri olarak verilmiştir. Ses tanıma modülünde SM SVM sınıflandırıcısı ve LS SVM sınıflandırıcısıyla gerçekleştirilen çalışmalar ile de sonuçlar alınmış ve değerlendirilmiştir.

Sistemin başarı durumunu etkileyen çeşitli faktörler vardır. Ortamda gürültü olması, mikrofonun konumu, ses kayıtlarını yaparken seste meydana gelen değişiklikler buna örnek olarak verilebilir. Özellikle mikrofonun konuşurken ağızdan uzaklaştırılması veya ağza yaklaştırılması sistemin başarı durumunu etkileyen faktörlerden birisi olmuştur. Ayrıca mikrofonun düşük gürültülü bir mikrofon olması ve işletim sistemindeki ayarlamaların da bu ölçüde yapılması sistemin başarı durumunu değiştirecektir. Gerçekleştirilen uygulamada düşük gürültülü bir mikrofon kullanılmıştır. Parametre taraması yapılarak sistemin örüntüleri tanımada en başarılı olduğu parametreleri seçip bu parametrelere göre uygulama gerçekleştirildiği takdirde de sistemin örüntü tanıma

başarısı değişecektir. Parametre tarama işlemi yapılan uygulamada gerçekleştirilmiş ve en iyi parametreler bulunmaya çalışılmıştır. Ayrıca örüntülerin kaydedilmesi aşamasında örnekleme frekansının yüksek olarak seçilmesi (örneğin 44.1 KHz gibi), her bir örüntüden alınan örnek sayısını arttıracak için, uygulamanın çalıştırılması esnasında belirgin bir yavaşlamaya ve süre kaybına neden olacaktır, çünkü her bir örüntüye ilişkin on binlerce veri hafızada oldukça fazla miktarda yer işgal edecektir. Bu da uygulamanın gerçekleştirilmesinde karşılaşılan problemlerden birisi olmuştur. Yapılan uygulamada örüntülerin kaydedilmesi aşamasında örnekleme frekansı 8 KHz (yani 8000Hz) olarak seçilmiştir.

Sistemin donanımsal yapısı açısından daha iyi ve performansı daha yüksek aygıtlar ile çalışılması zaman kaybının düşmesini sağlayabilecektir. Böylece parametre tarama işlemleri daha geniş bir ölçekte yapılabilecek ve en iyi parametrelerin bulunması esnasında daha iyi sonuçlara ulaşılabilecektir. Bu da sistemin başarı durumunu arttırabilecektir.

KAYNAKLAR

- Al-Haddad, S. A. R., Samad, S. A., Hussain, A. and Ishak, K. A. (2008) Isolated Malay Digit Recognition Using Pattern Recognition Fusion of Dynamic Time Warping and Hidden Markov Models, *American Journal of Applied Sciences*, 5 (6): 714-720.
- Baygün, M. K., (2006) Türkçe Komutları Tanıyan Ses Tanıma Sistemi Geliştirilmesi, Yüksek Lisans Tezi, *Pamukkale Üniversitesi Fen Bilimleri Enstitüsü*, Denizli, 69s.
- Chen, J. K. and Soong, F. K. (1994) An N-Best Candidates-Based Discriminative Training for Speech Recognition Applications, *IEEE Transactions on Speech and Audio Processing*, 2 (1): 206-216.
- Cristianini, N. and Taylor, J. S. (2000) An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods, *Cambridge University Press*, s.172.
- Davis, K. H., Biddulph, R. and Balashek, S. (1952) Automatic Recognition of Spoken Digits, *Journal of the Acoustic Society of America*, 24 (6): 637-642.
- Debyeche, M., Haton, J. P. and Houacine, A. (2006) A New Vector Quantization Front-End Process for Discrete HMM Speech Recognition System, *International Journal of Signal Processing*, 3 (1): 46-51.
- Demirci, M. D. (2005) Bilgisayar Destekli Ses Tanıma Sistemi Tasarımı, Yüksek Lisans Tezi, *İstanbul Üniversitesi Fen Bilimleri Enstitüsü*, İstanbul, 45s.
- El-Ramly, S. H., Abdel-Kader, N. S. and El-Adawi, R. (2002) Neural Networks Used for Speech Recognition, *Nineteenth National Radio Science Conference*, Alexandria, s.200-207.
- Forgie, J. W. and Forgie, C. D. (1959) Results Obtained From a Vowel Recognition Computer Program, *Journal of the Acoustic Society of America*, 31 (11): 1480-1489.
- Fry, D. B. (1959) Theoretical Aspects of Mechanical Speech Recognition, *Journal of the British Institution Radio Engineers*, 19 (4): 211-229.
- Hagiwara, R. (2008) Monthly Mystery Spectrogram Webzone, <http://home.cc.umanitoba.ca/~robh/index.html>, (02.07.2008).
- Hermansky, H. (1990) Perceptual Linear Predictive Coding Analysis of Speech, *Journal of the Acoustic Society of America*, 87 (4): 1738-1752.
- Itakura, F. (1975) Minimum Prediction Residual Applied to Speech Recognition, *IEEE Transactions on Acoustics, Speech, Signal Processing*, ASSP-23 (1): 67-72.

- Krüger, S. E., Schafföner, M., Katz, M., Andelic, E. and Wendemuth, A. (2006) Mixture of Support Vector Machines for HMM based Speech Recognition, *The 18th International Conference on Pattern Recognition*, Magdeburg, s.326-329.
- Lam, H. K. and Leung, F. H. F. (2007) Design and Training for Combinational Neural-Logic Systems, *IEEE Transactions on Industrial Electronics*, 54 (1): 612-619.
- Lin, Y. L. and Wei, G. (2005) Speech Emotion Recognition Based on HMM and SVM, *Proceedings of the Fourth International Conference on Machine Learning and Cybernetics*, Guangzhou, s.4898-4901
- Nabiyev, V. V. (2005) Yapay Zeka, *Seçkin Yayınları*, Ankara, 764s.
- Olson, K. H. and Belar, H. (1956) Phonetic Typewriter, *Journal of the Acoustic Society of America*, 28 (6): 1072-1081.
- Rabiner, L. and Juang, B. H. (1993) Fundamentals of Speech Recognition, *Prentice Hall Publishing*, New Jersey, s.507.
- Reddy, D. R. (1967) Computer Recognition of Connected Speech, *Journal of the Acoustic Society of America*, 42: 329-347.
- Sakai, T. and Doshita, S. (1962) The Phonetic Typewriter, Information Processing, *Proceedings of IFIP Congress*, Munich, s.445-450.
- Sakoe, H. and Chiba, S. (1978) Dynamic Programming Algorithm Optimization for Spoken Word Recognition, *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-26 (1): 43-49.
- Schölkopf, B., Burges, C. J. C. and Smola, A. J. (1999) Advances in Kernel Methods: Support Vector Learning, *The MIT Press*, Cambridge, s.327-352.
- Suykens, J. A. K. and Vandewalle, J. (1999) Least Squares Support Vector Machine Classifiers, *Neural Processing Letters*, 9 (3): 293-300.
- Suzuki, J. and Nakata, K. (1961) Recognition of Japanese Vowels-Preliminary to the Recognition of the Speech, *Journal of Radio Research Lab.*, 37 (8): 193-212.
- Uzunçarşılı, M. (2005) Vektör Nicemleme Tekniklerine Dayalı Konuşmacı Tanıma Algoritmalarının İncelenmesi, Yüksek Lisans Tezi, *Ankara Üniversitesi Fen Bilimleri Enstitüsü*, Ankara, 68s.
- Vapnik, V. (1995) The Nature of Statistical Learning Theory, *Springer-Verlag Publishing*, New York, s.187.
- Vapnik, V. (1998a) Statistical Learning Theory, *John Wiley Publishing*, New York, s.768.

- Vapnik, V. (1998b) The Support Vector Method of Function Estimation, In *Nonlinear Modelling Advanced Black Box Techniques*, **Kluwer Academic Publishers**, Boston, s55-85.
- Velichko, V. M. and Zagoruyko, N. G. (1970) Automatic Recognition of 200 Words, *International Journal of Man-Machine Studies*, 2: 223.
- Vintsyuk, T. K. (1968) Speech Discrimination by Dynamic Programming, *Kibernetika*, 4 (2): 81-88.
- Vogt, R. and Sridharan, S. (2008) Explicit Modelling of Session Variability for Speaker Verification, *Computer Speech and Language*, 22 (1): 17-38.
- Wachter, M. D., Matton, M., Demuynck, K. and Wambacq, P. (2007) Template-Based Continuous Speech Recognition, *IEEE Transactions on Audio, Speech and Language Processing*, 15 (4): 1377-1390.
- Yaniv, R. and Burshtein, D. (2003) An Enhanced Dynamic Time Warping Model for Improved Estimation of DTW Parameters, *IEEE Transactions on Speech and Audio Processing*, 11 (3): 216-228.

ÖZGEÇMİŞ

Osman ERAY, 29 Nisan 1979 tarihinde Burdur'un Gölhisar ilçesinde doğmuştur. İlkokulu Gölhisar'da; ortaokulu ve liseyi Antalya'da okumuştur. 2002 yılında Kocaeli Üniversitesi Elektronik ve Haberleşme Mühendisliği bölümünden mezun olmuştur. 2002-2003 yılları arasında askerliğini 18. Zırhlı Tugay Hava Savunma Batarya Komutanlığı'nda yedek subay olarak yapmıştır. 2003-2006 yılları arasında Gölhisar Meslek Yüksekokulu'nda çalışmıştır. 2007 yılından beri Akdeniz Üniversitesi Korkuteli Meslek Yüksekokulu Bilgisayar Teknolojileri ve Programlama bölümünde Öğretim Görevlisi olarak çalışmaktadır. Evli ve bir çocuk babasıdır.