

# Akış

- Makine Öğrenmesi nedir?
- Günlük Hayatımızdaki Uygulamaları
- Verilerin Sayısallaştırılması
- Özellik Belirleme
  - Özellik Seçim Metotları
    - Bilgi Kazancı (Informaiton Gain-IG)
    - Sinyal-Gürültüye Oranı: (S2N ratio)
    - Alt küme seçiciler (Wrappers)
  - Yeni Özelliklerin Çıkarımı
    - Temel Bileşen Analizi (Principal Component Analysis)
    - Doğrusal Ayırteden Analizi (Linear Discriminant Analysis)
- Sınıflandırma Metotları
  - Doğrusal Regresyon
  - Karar Ağaçları (Decision Trees)
  - Yapay Sinir Ağları
  - En Yakın K Komşu Algoritması (k - Nearest Neighbor)
- Kümeleme Algoritmaları:
  - K-Ortalama (K-Means)
  - Kendi Kendini Düzenleyen Haritalar (Self Organizing Map -SOM)



# Makine Öğrenmesi

Çok büyük mikardaki verilerin elle işlenmesi,  
analizinin yapılması mümkün değildir.

Bu tür problemlere çözüm bulmak amacıyla  
makine öğrenmesi metotları geliştirilmiştir.

Bu metotlar  
**geçmişteki verileri kullanarak**  
veriye en uygun **modeli** bulmaya çalışırlar.  
Yeni gelen verileri de bu modele göre analiz edip sonuç üretirler.

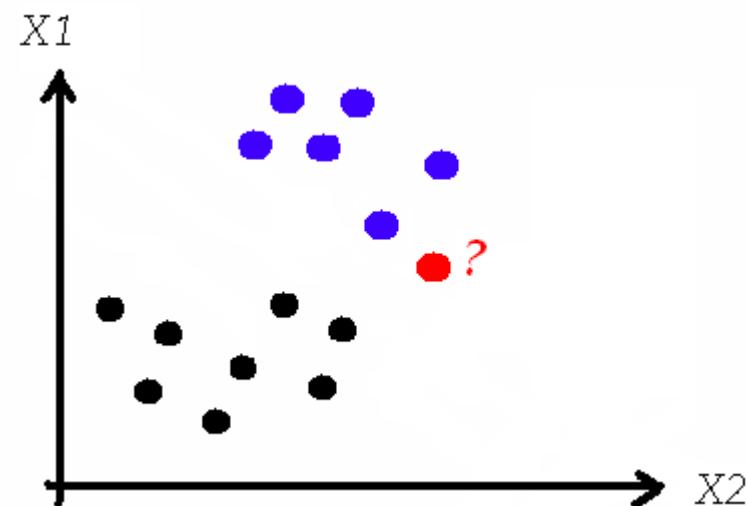


# Metot türleri

- Farklı uygulamaların analizlerden farklı beklenileri olmaktadır.
- Makine öğrenmesi metotlarını bu beklenilere göre sınıflandırmak mümkündür.



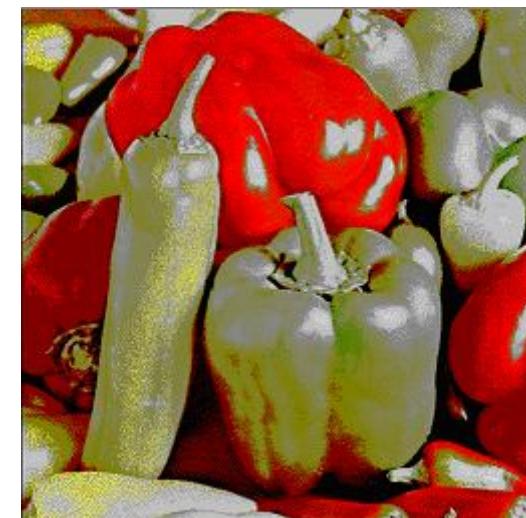
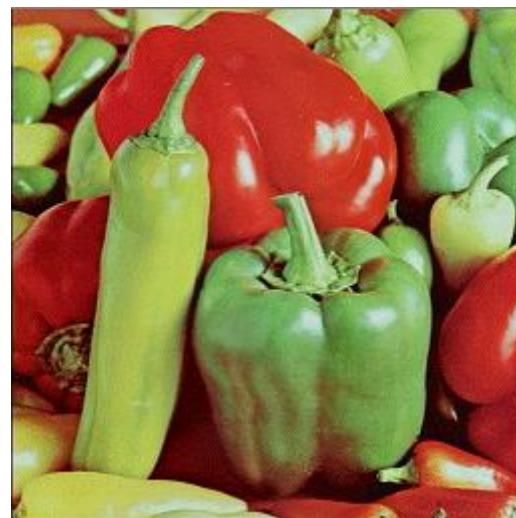
# Sınıflandırma



Kırmızı hangi sınıfın?

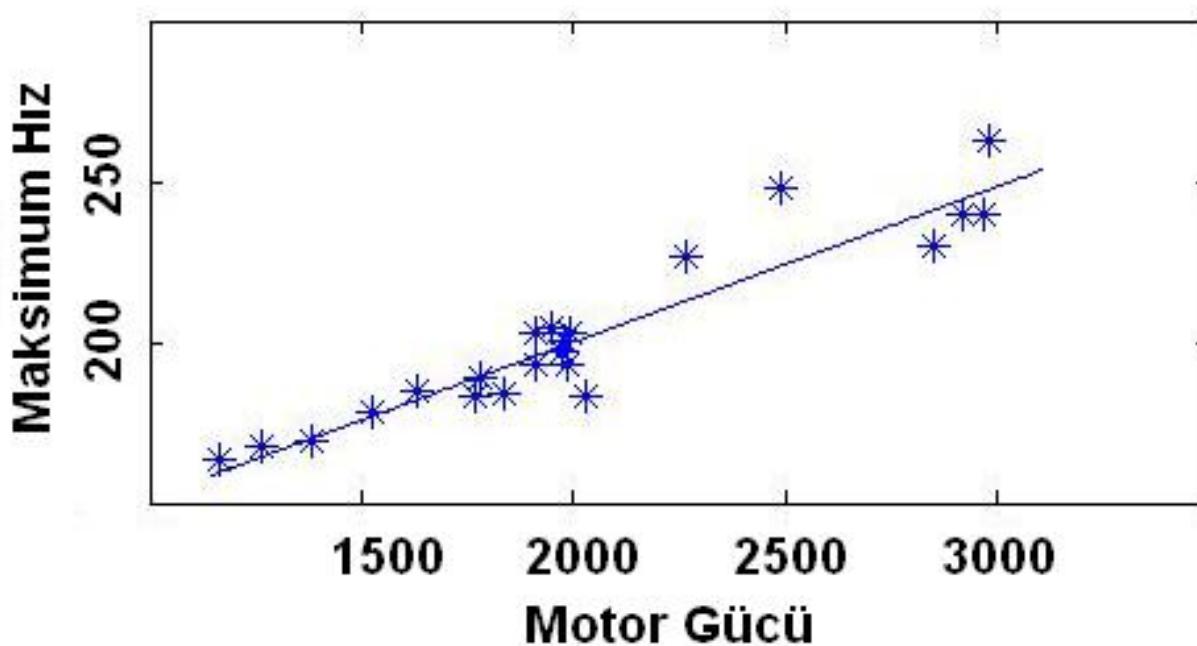
# Kümeleme

- 256 rengi 16 renge nasıl indiririz?



# Regresyon

## Eğri Uydurma



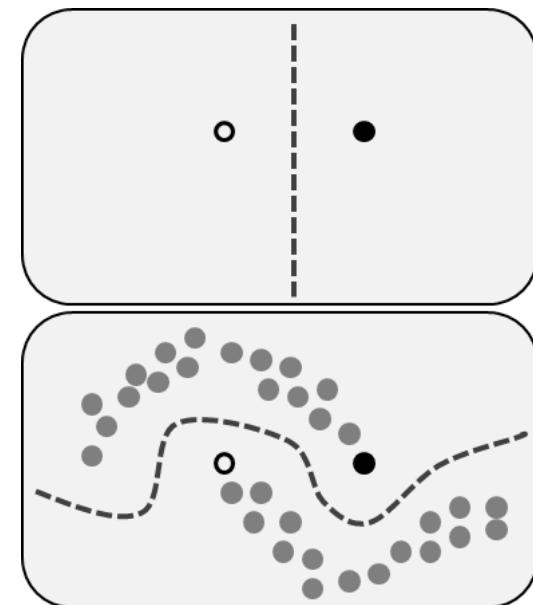
# Birlikte Satış Kuralları Keşfi

- Sepet analizi
  - hangi ürünler birlikte satılıyor?
- Raf düzenlemesi
  - hangi ürünler yan yana konmalıdır?
- Promosyonlar
  - neyin yanında ne verilmeli?



# Yeni problemler-1

- Aktif Öğrenme: eğitim örneklerinin etiketlerinin başta belli olmadığı, algoritmanın etiketini istediği örnekleri seçtiği problemlerdir.
- Yarı eğiticili öğrenme: modelin oluşturulmasında etiketli verilerin yanında etiketsiz verilerinde kullanıldığı problemlerdir.



# Yeni problemler-2

- Partial labels: Örneklerin birden fazla etiketi var. Sadece 1 i doğru. İçinde birden fazla kişinin olduğu resimler ve isimleri var. Hangisi hangisinin ismi belli değil.
- Multi-label learning: Bir örneğin birden fazla doğru etiketinin olduğu durumlar. Bir kişi hem baba, hem öğretmen, hem şışman olabilir.



# Yeni problemler-3

- Learning from crowds: Örneklerin çok sayıda öğretmenin etiketlediği, hangisinin doğru olduğunu bilinmediği durumlar. 100 kişiye elinizdeki verileri etiketletiyorsunuz.
- Budgeted learning: Elinde bir bütçe var. Bununla elindeki eğitim kümесinin örneklerini, örneklerin özelliklerini satın alıyorsun. Amaç eldeki bütçe (minimum data) en iyi test performansını elde etmek



# Akış

- Makine Öğrenmesi nedir?
- **Günlük Hayatımızdaki Uygulamaları**
- Verilerin Sayısallaştırılması
- Özellik Belirleme
  - Özellik Seçim Metotları
    - Bilgi Kazancı (Information Gain-IG)
    - Sinyal-Gürültüye Oranı: (S/N ratio)
    - Alt kümeye seçiciler (Wrappers)
  - Yeni Özelliklerin Çıkarımı
    - Temel Bileşen Analizi (Principal Component Analysis)
    - Doğrusal Ayırteden Analizi (Linear Discriminant Analysis)
- Sınıflandırma Metotları
  - Doğrusal Regresyon
  - Karar Ağaçları (Decision Trees)
  - Yapay Sinir Ağları
  - En Yakın K Komşu Algoritması (k - Nearest Neighbor)
  - Öğrenmeli Vektör Kuantalama (Learning Vector Quantization)
- Kümeleme Algoritmaları:
  - K-Ortalama (K-Means)
  - Kendi Kendini Düzenleyen Haritalar (Self Organizing Map -SOM)



# Makine Öğrenmesinin Günlük Hayatımızdaki Uygulamaları

YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



# El yazısı / Kitap Yazısı Tanıma HCR /OCR

Thanks a million for your support  
of the Ellen McArthur Trust  
I hope our biggest reward  
will be the wife and daughter  
of the kids out in the water!  
See you for it!

1997 - 1998 - 1999 - 2000  
and we welcome our children elsewhere. If you're not yet there  
please consider giving us a donation or buy "Boat People" the  
book to help.  
It's happening right now because everyone needs to work  
together - and I am working.  
2000 - In November we'll be making another trip to the Maldives  
and I expect to meet all the local guides, because  
we're not here to compete but to try to understand understand.  
2001 - I'm back in Asia again, this time in Thailand  
working with the local guides, and I am a product  
of Thailand's guides, guides, and guides are really  
people at heart. So I believe that they are the ones  
who are most important in terms of their own development  
and so I think that they are very important to help them  
achieve their goals, and the I think that, this programme  
is very special.

2002 - That is just the present.  
2003 - I am going to continue to promote tourism guides  
in my new capacity, and start more projects like  
the one I'm working on right now, promoting the local culture  
and the local guides, or tourism guides for the local  
culture, or culture guides for the tourism.

2004 - And so on...  
2005 - I will be an entrepreneur, opening more schools, maybe  
and trying to do what other guides can't do, which  
is to promote more diversity in guides and provide guides  
a better place, because this is a much harder area  
than just tourism, and it's more guides to be able  
to work and earn money, and to provide them, and  
make more opportunities for them, maybe to change  
their lives, if you know what I mean. I am a  
traveler myself and I travel a lot, so I understand  
what it means to travel and to experience different things,  
and I hope to continue to inspire people around me to do the  
same, so I guess I like to think that my guides  
are the ones who have the most potential to make a difference.

İşlem: Şekillerin hangi harf olduğunu tahmin etme



# Kredi Taleplerini Değerlendirme

- Birisi bankadan borç ister.
- Banka borcu versin/vermesin.
- Nasıl?



# e-ticaret

- Birisi Amazon.com dan bir kitap yada ürün alıyor.

## Görev ne olabilir?



# e-ticaret

- Birisi Amazon.com dan bir kitap yada ürün alıyor.

## Görev ne olabilir?

Müşteriye alması muhtemel kitapları önerelim.

Ama nasıl?

Kitapları

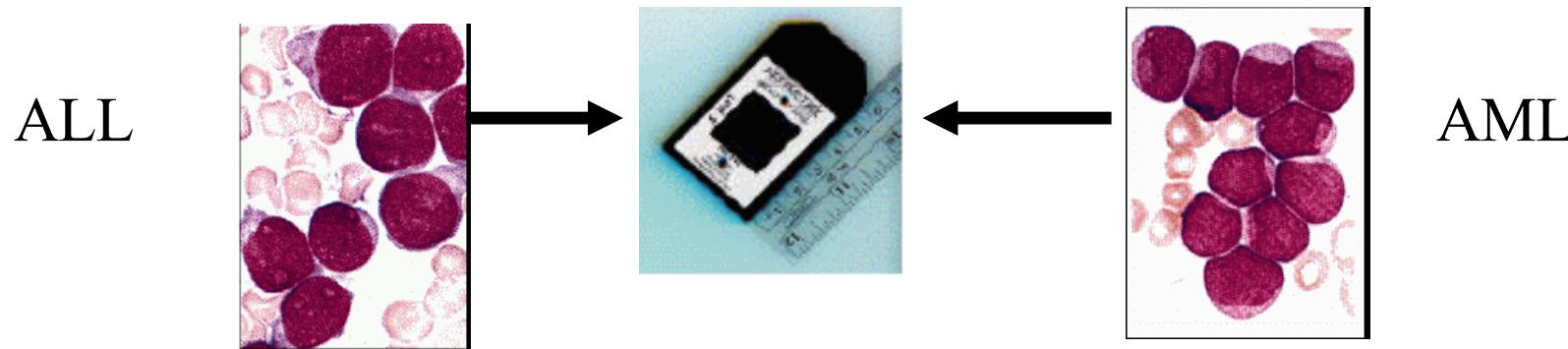
konularına

yazarlarına

birlikte satışlarına

göre kümelemek?





## Gen Mikrodizilimleri

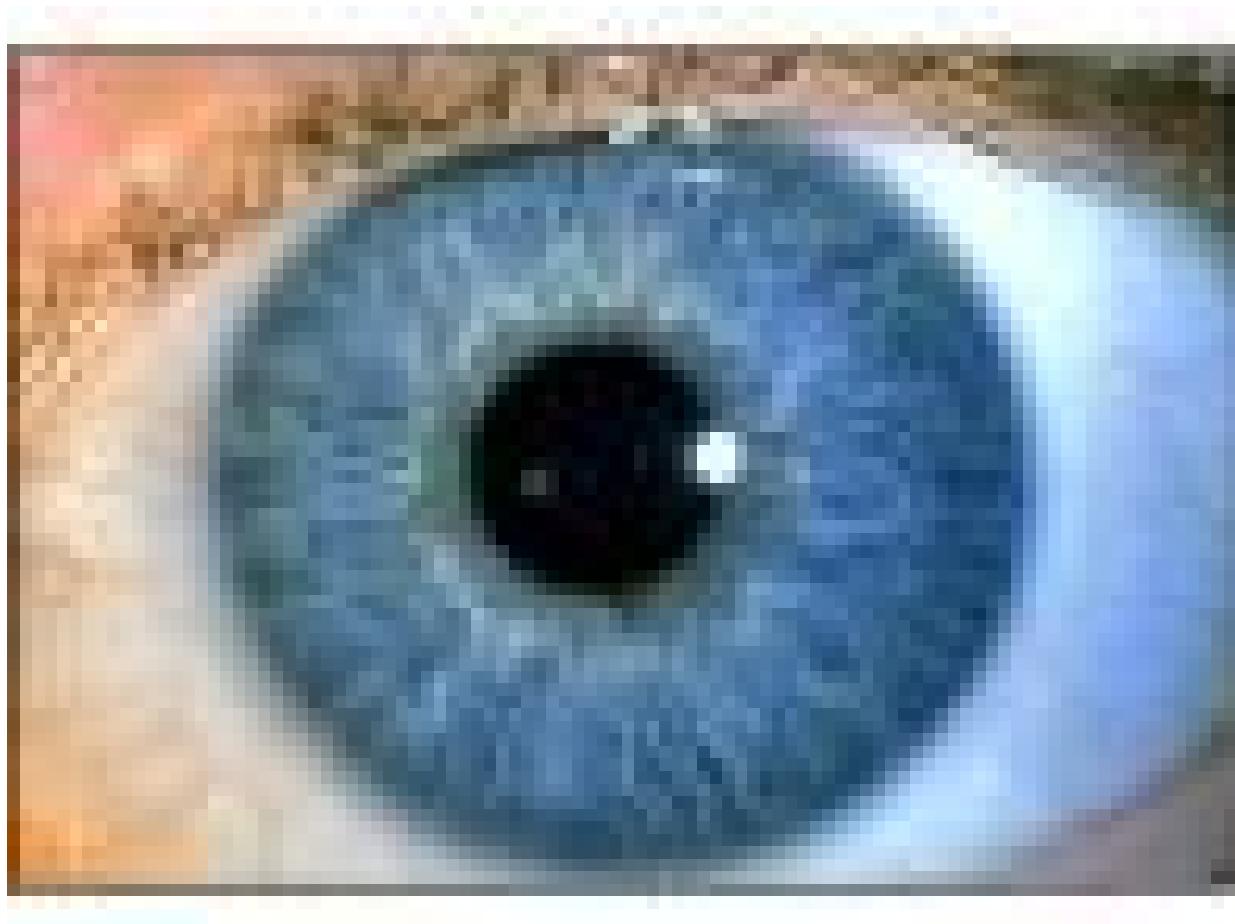
100 kişinin (hasta/sağlam) elimizde gen dizilimleri var.

Bu dizilimleri analiz ederek hasta olup olmadığı bilinmeyen birisinin hasta olup olmadığını yada hastalığının türünü öğrenebilir miyiz?

En iyi tedaviyi öneremiz miyiz?

Nasıl? Elimizde hangi bilgiler olmalı?

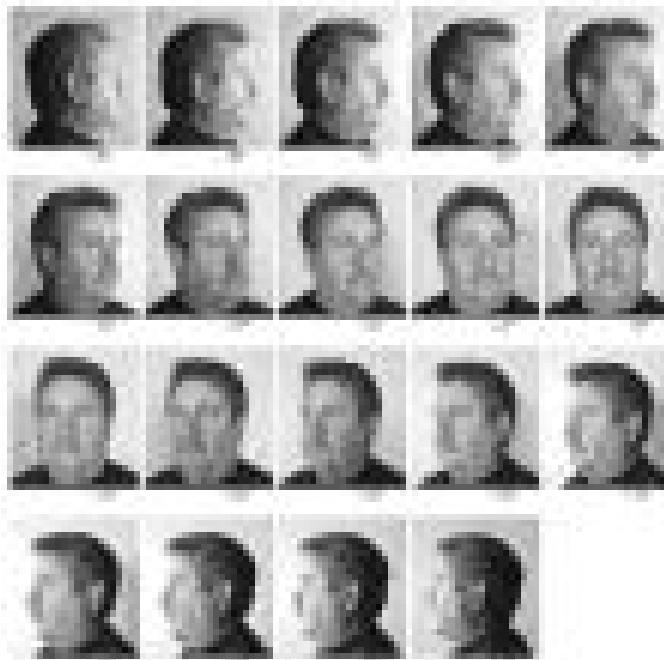
# Bu adam kim? İçeri girsin mi?



YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



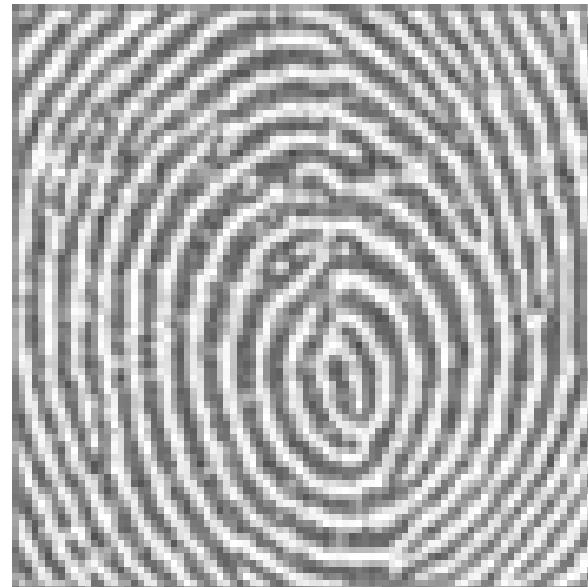
Bu adam kim?  
Bu adam havaalanında mı?



YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



Bu parmak izi kimin?  
Bu adamı tutuklayalım mı?



Bu ses kimin?  
Bu ses ne diyor?



YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



# Bu imza kimin? Yoksa taklit mi?



Taklit olup olmadığını nasıl anlarız?  
Zaman bilgisi ?

# Bu metnin konusu nedir?

## Bu mail spam mi?



Anti spam yazılımları nasıl çalışır?

Spamciler nasıl çalışıyor?

**Yeni nesil spam mailler:** Mesaj resimde,  
*metinde ise anti spamlardan kaçmak* için gereken kelimeler var.

Makine öğrenmesi metodlarını hem spamciler hem anti spmciler kullanıyor.

# Olağan dışı bir durum var mı? Güvenlik kamerası kayıtları



YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ



# Kamera kaydındaki kişi ne anlatıyor?



YILDIZ TEKNİK ÜNİVERSİTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

Sonuç: İletişimin artması

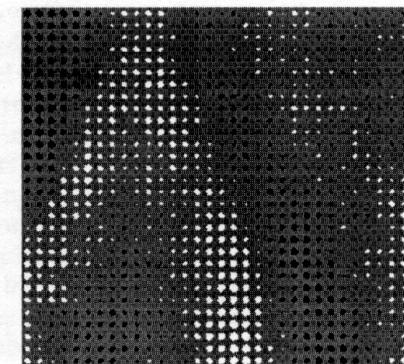
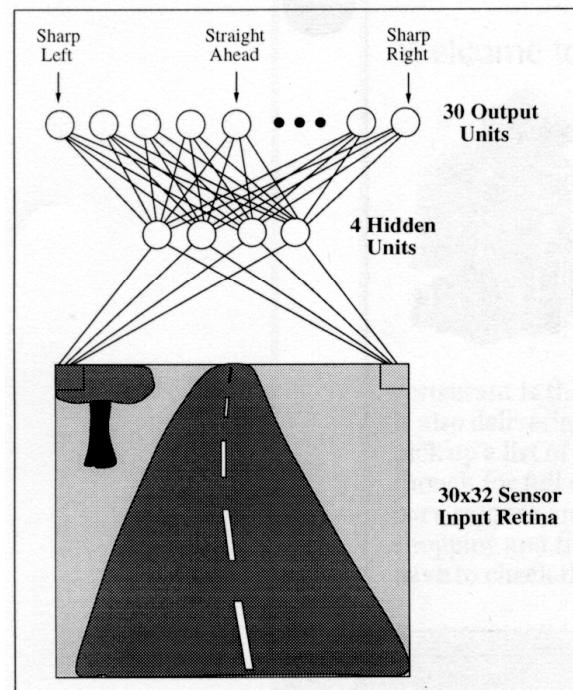
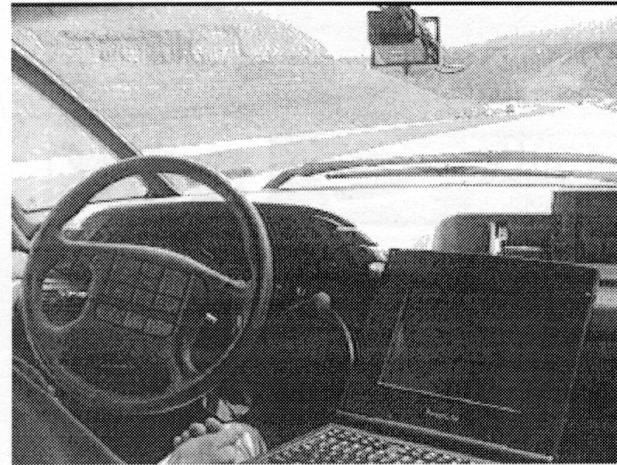


# ALVIN

Otobanda saatte 70 mil hızla sürücüsüz gidebilen bir otomobil

Bütün denemeler  
trafiğe kapalı alanlarda  
gerçekleştirilmiştir ☺

Neden şehiriçi değil?  
Neden otoban?  
Neden diğer arabalar yok?  
Araba birine çarparsa suçlu kim?



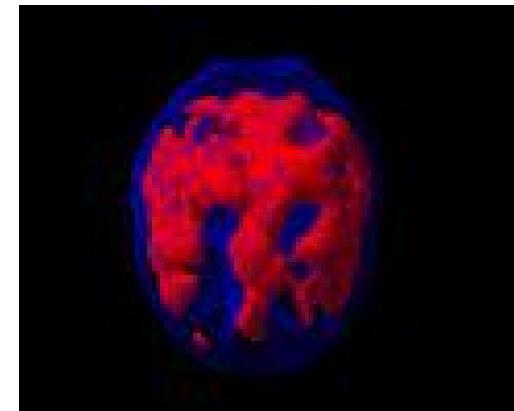
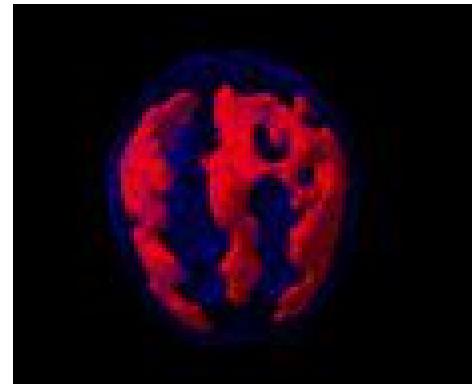
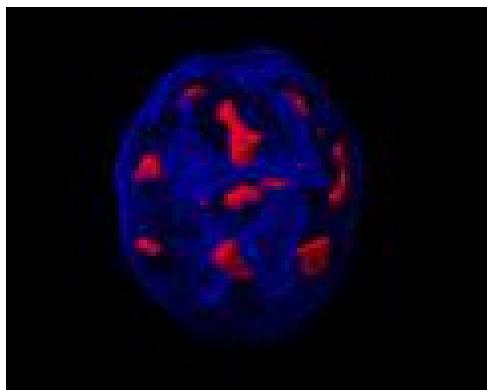
# Adalet

- Çin'de pilot uygulama:
  - bir şehrin mahkeme hakimleri bir bilgisayar programı
  - Amaç: Daha adil bir dünya
    - Aynı özelliklere sahip davalarda aynı kararların alınması
  - Sistemin eğitimi için neler gerekli?
    - Milyonlarca/Milyarlarca (orası Çin) davaya ait verilerin kategorilenmesi

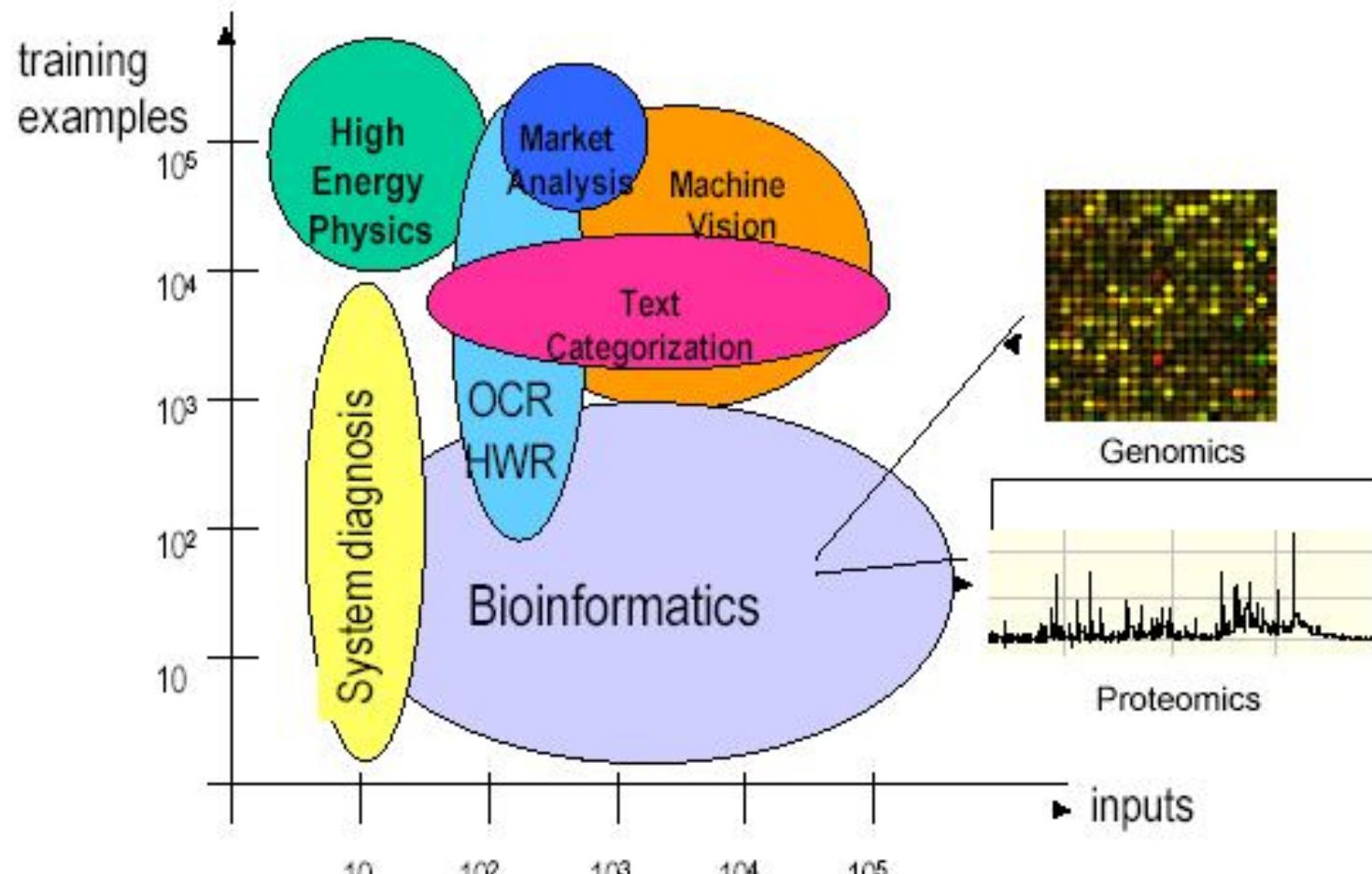


# Beyin Aktiviteleri

- İnsanların
  - değişik şeyler düşünürkenki,
  - değişik duygulara sahipkenki,
  - problem çözerken ki
  - beyin aktiviteleri kaydedilir.
- Görev?



# Uygulamalardaki boyut örnek sayıları



# Öğrenmenin geleceği

- Öğrenme modülleri birçok uygulamaya girmiş durumda.
- Ticari olarak satılan (gerçek dünya için tasarlanmış) birçok ürün var.
- Bu tür ürün ve modüllerin sayısı giderek artmakte.
- Çözüm bekleyen ve büyük potansiyelleri olan birçok problem (?) var.



# Verilerin Sayısallaştırılması

Resim	Resmin her bir pikselinin renkli resimlerde R,G,B değerleri, siyah-beyaz resimlerde 1–255 arası gri seviyesi kullanılarak sayılara çevrilir. Renkli resimler 3 adet, siyah beyazlar 1 adet en*boy büyüğünde matrisle ifade edilir.
Metin	Metindeki harfler, heceler ve kelimeler genelde frekanslarına göre kodlanarak sayılara çevrilir.
Hareketli görüntü	Resim bilgisine ek olarak resmin hangi resimden sonra geldiğini gösteren zaman bilgisini de içerir. Bu ek bilgi haricinde yapılan işlem resim ile aynıdır.
Ses	Ses, genlik ve frekansın zaman içinde değişimiyle kodlanır.

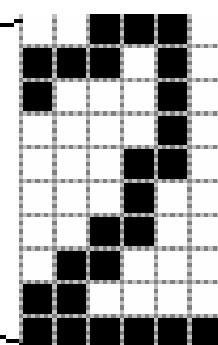


# Örnek sınıflandırma uygulaması

- Sistem: bir kitap fotokopisinin içindeki yazılarının metne dönüştürülmesi
- Öncelikle metindeki satırlar bulunur.
- Her bir satırdaki harfler bulunur. Her harfe ait onlarca örnek resimden etiketlenmiş bir veritabanı oluşturulur. Her bir resim için



için 2 ye bölünür.



0	0	1	1	1	0
1	1	1	0	1	0
1	0	0	0	1	0
0	0	0	0	1	0
0	0	0	1	1	0
0	0	0	1	0	0
0	0	1	1	0	0
0	1	1	0	0	0
1	1	0	0	0	0
1	1	1	1	1	1



0 0 1 1 1 0 1 1 1 0 1 0 0 0 1 0 0 0 0 1 0 0 0 0 1 1 0 0 0 1 1 0 0 0 1 1 0 0 0 0 1 1 1 1 1 1

- Bu şekilde tanınmak istenen harf için çeşitli fontlarla yazılmış birçok örneği temsil eden 60 boyutlu vektörler elde edilir.
- Bu uygulamamız için özellik sayımız 60'tır. Diğer bir deyişle örneklerimiz 60 boyutlu bir uzayda temsil edilmektedirler.
- Elimizde 10 rakama ait farklı fontlarla yazılmış 10'ar resim olursa veri kümemiz 100 örnek\* 60 boyutluk bir matris olacaktır.
- Elimizde her örneğin hangi harf olduğunu gösteren sınıf bilgiside bulunmaktadır.
- Bu matris eğitim ve test kümesi oluşturmak için 2'ye bölünür.
- Eğitim kümesi bir sınıflandırıcıya verilir.
- Modellenir.
- Modelin başarısını ölçmek için sınıflandırıcının daha önce görmediği, modelini oluşturmakta kullanmadığı test kümesi için tahminde bulunması istenir.
- Bu tahminlerle gerçek sınıfların ayınlığının ölçüsü sınıflandırıcının başarı ölçüsüdür.

