

# Makine Öğrenmesi-1

## Akış

- **Makine Öğrenmesi nedir?**
- **Günlük Hayatımızdaki Uygulamaları**
- **Verilerin Sayısallaştırılması**
- **Özellik Belirleme**
  - Özellik Seçim Metotları
    - Bilgi Kazancı (Informaiton Gain-IG)
    - Sinyalin 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)
  - Öğrenmeli Vektör Kuantalama (Learning Vector Quantization)
- **Kümeleme Algoritmaları:**
  - Hiyerarşik Kümeleme
  - K-means
  - Kendi Kendini Düzenleyen Haritalar (Self Organizing Map -SOM)
  - DBscan
- **Regresyon Algoritmaları**
- **Çok Boyutlu Verilerle Çalışmak**
- **Veri Sızıntısı**
- **Pekiştirmeli Öğrenme**

# Makine Öğrenmesi

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

Bu tür problemlere çözüm bulmak amacıyla makine öğrenmesi metotları geliştirilmiştir.

Bu metotlar

**geçmişteki verileri kullanarak**

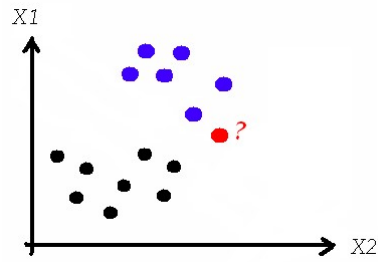
veriye en uygun **modeli** bulmaya çalışırlar.

Yeni gelen verileri de bu modele göre analiz edip sonuç üretirler.

## Metot türleri

- Farklı uygulamaların analizlerden farklı beklentileri olmaktadır.
- Makine öğrenmesi metotlarını bu beklentilere göre sınıflandırmak mümkündür.

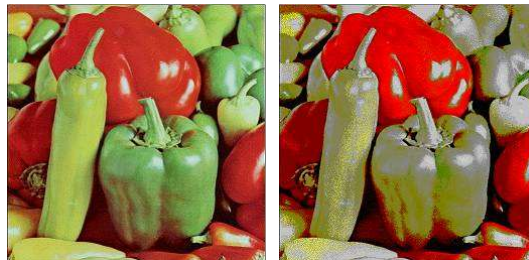
## Sınıflandırma



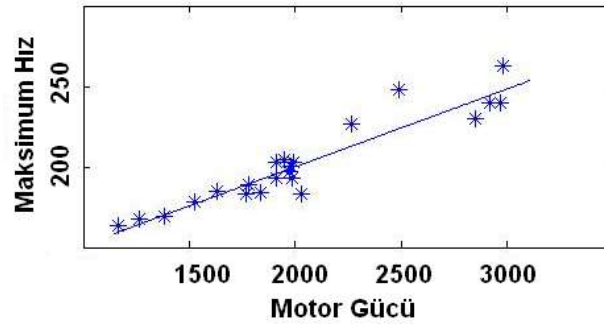
Kırmızı hangi sınıftan?

## Kümeleme

- 256 rengi 16 renge nasıl indiririz?



## Regresyon Eğri Uydurma



## Pekiştirmeli Öğrenme

- Modele çıkışının ne olması gerektiği söylenmez sadece doğru ya da yanlış yaptığı söylenir. ML3'te göreceğiz.

## Birliktelik Kuralları Keşfi

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

## Makine Öğrenmesinin

## Günlük Hayatımızdaki Uygulamaları

## El yazısı / Kitap Yazısı Tanıma HCR /OCR



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

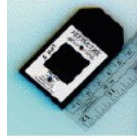
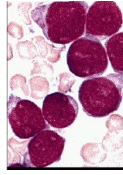
konularına

yazarlarına

birlikte satılışlarına

göre kümelemek?

ALL



AML



## Gen Mikrodizilimleri

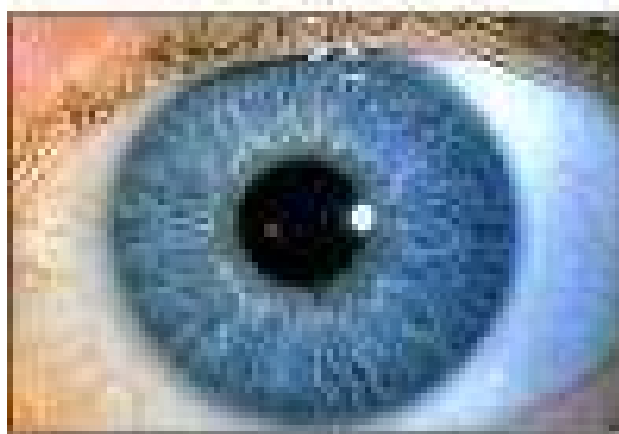
100 kişinin (hasta/sağlam) elimizde gen dizilimleri var.

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

En iyi tedaviyi önerebilir miyiz?

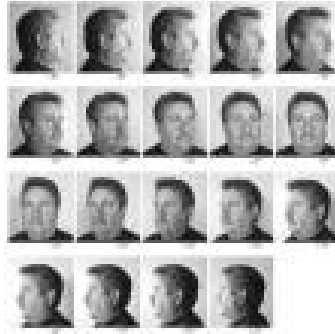
Nasıl? Elimizde hangi bilgiler olmalı?

## Bu adam kim? İçeri girsin mi?

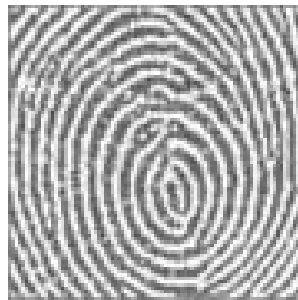




Bu adam kim?  
Bu adam havaalanında mı?



Bu parmak izi kimin?  
Bu adamı tutuklayalım mı?



Bu ses kimin?  
Bu ses ne diyor?



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 spamciler kullanıyor.

## Olağan dışı bir durum var mı? Güvenlik kamerası kayıtları



## Kamera kaydındaki kiři ne anlatıyor?

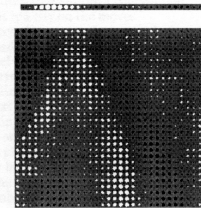
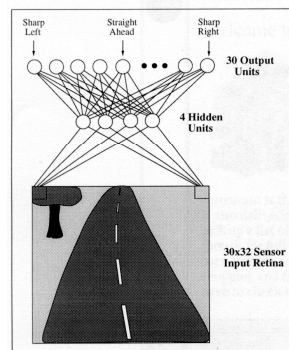
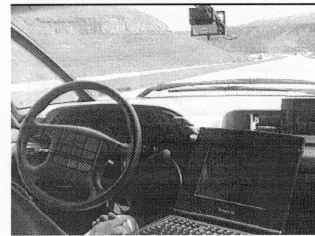


## ALVIN

**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?

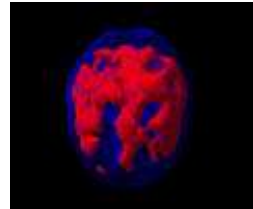
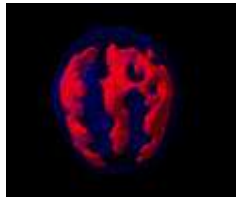
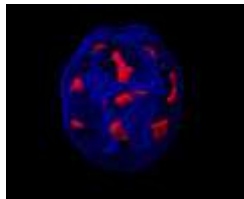


## Adalet

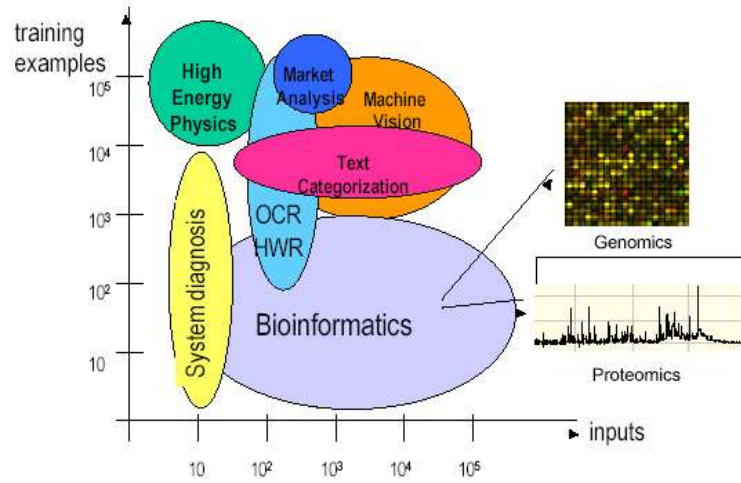
- Çin’de pilot uygulama:
  - bir şehrin mahkeme hakimleri bir bilgisayar programı
  - Amaç: Daha adil bir dünya
    - Aynı özelliklere sahip davalarda aynı kararların alınması
  - Sistemin eğitimi için neler gerekli?
    - Milyonlarca/Milyarlarca (orası Çin) davaya ait verilerin kategorilenmesi

## Beyin Aktiviteleri

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



## Uygulamalardaki boyut örnek sayıları



Mehmet Fatih AMASYALI Yapay Zeka Ders Notları

YILDIZ TEKNİK ÜNİVERSİTESİ BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

## 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 artmakta.
- Çözüm bekleyen ve büyük potansiyelleri olan birçok problem (?) var.

Mehmet Fatih AMASYALI Yapay Zeka Ders Notları

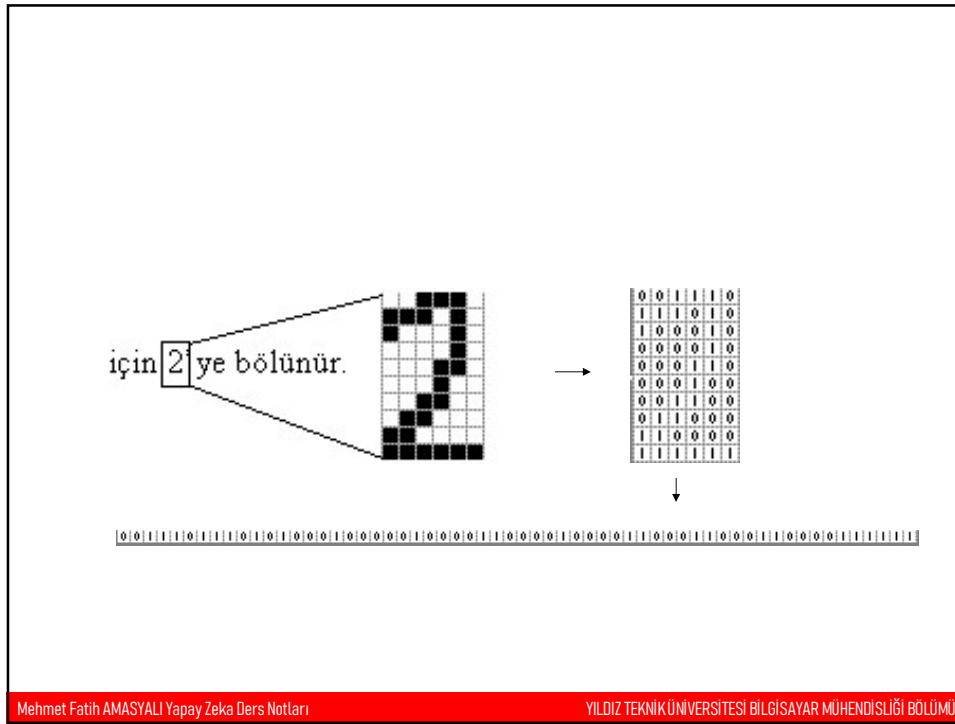
YILDIZ TEKNİK ÜNİVERSİTESİ BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

## 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 büyüklüğünde 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



- 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 oluşturmada kullanmadığı test kümesi için tahminde bulunması istenir.
- Bu tahminlerle gerçek sınıfların aynılığının ölçüsü sınıflandırıcının başarı ölçüsüdür.