



KOCAELİ ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ

Bilgisayar

Mühendisliğinde Matematik

Uygulamaları

8. Hafta

Yrd. Doç. Dr. A. Burak İNNER

Kocaeli Üniversitesi Bilgisayar Mühendisliği
Yapay Zeka ve Benzetim Sistemleri Ar-Ge Lab.
<http://yapbenzet.kocaeli.edu.tr>

∞ Regresyon

- En Küçük Kareler Yöntemi
 - Doğru Uydurma
 - Polinom Uydurma
 - Üstel Fonksiyonlara Eğri Uydurma

∞ İnterpolasyon

- Lagrange İnterpolasyonu (Polinom İnterpolasyon)
- Newton İnterpolasyonu

REGRESYON VE İNTERPOLASYON

Eđri uydurma için, veri **hatalarına bađlı** olarak birbirinden ayrılan iki yaklaşım vardır:

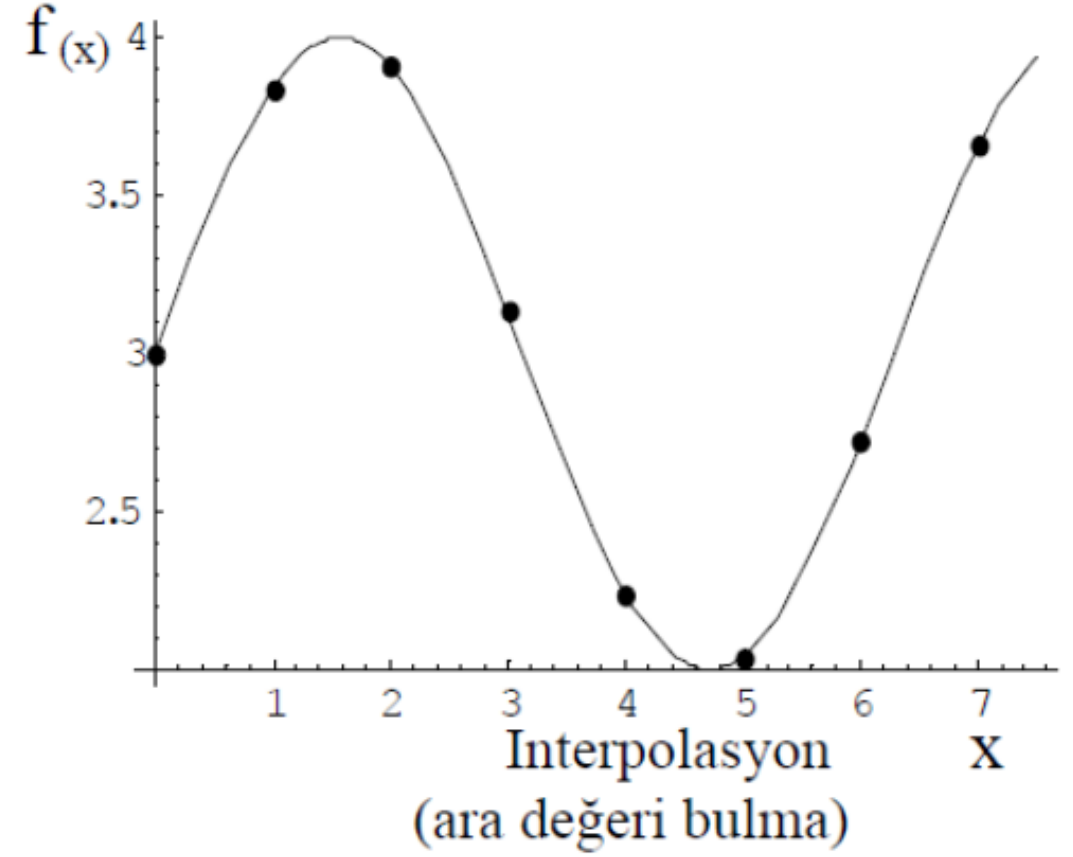
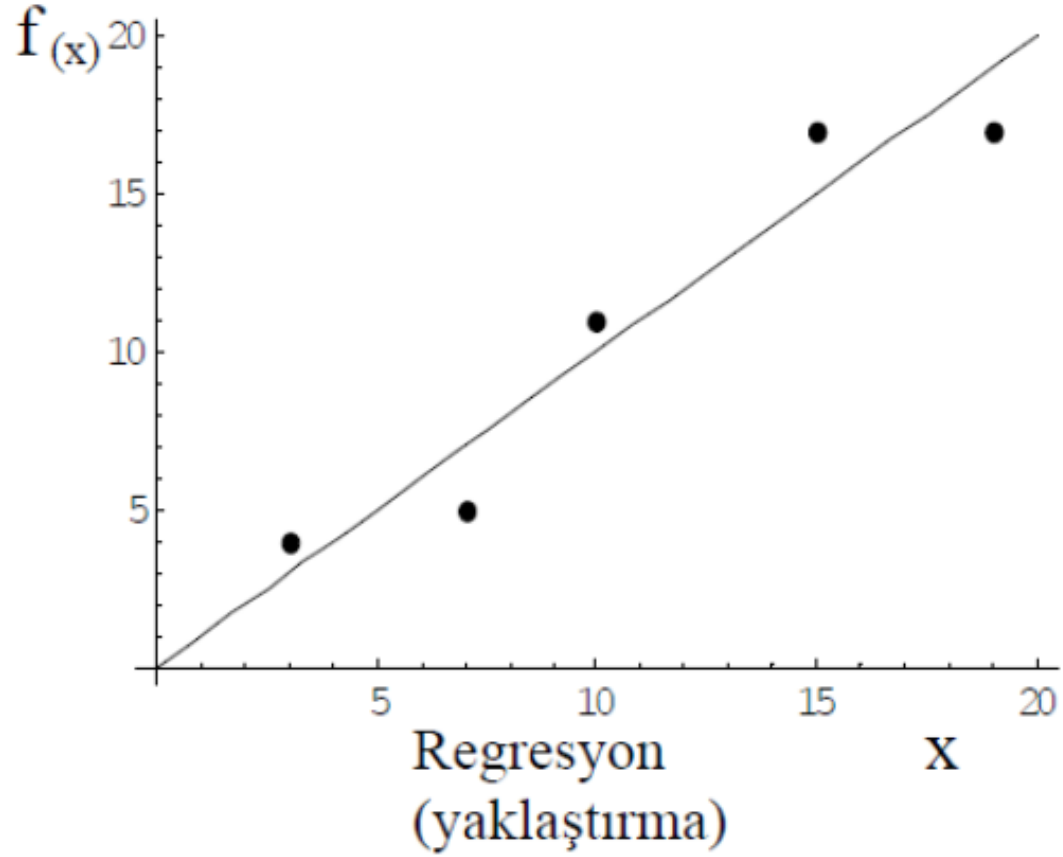
Regresyon

Hata oranı büyük deđerlerin, her bir veri noktasından **geçmeyen**, verilerin genel eğiliminin **tek bir eđri** ile gösterimidir. (Deneysel veriler için)

İnterpolasyon

Hata oranı küçük deđerlerin, iyi bilinen ayırık noktaların **her birinden geçecek** şekilde **eđri veya eđri uydurularak** gösterimidir.

REGRESYON VE İNTERPOLASYON



REGRESYON

- ∞ Regresyon analizi yaparken en çok kullanılan yöntemlerden biri **en küçük kareler yöntemi**dir.
- ∞ En Küçük Kareler Yöntemi gerçek yaşamın çeşitli alanlarında herhangi bir uygulama ile toplanan veriler tablo şekline getirilerek incelenir ve toplanan veriyi modelleyen bir fonksiyon bulunmaya çalışılır.

REGRESYON

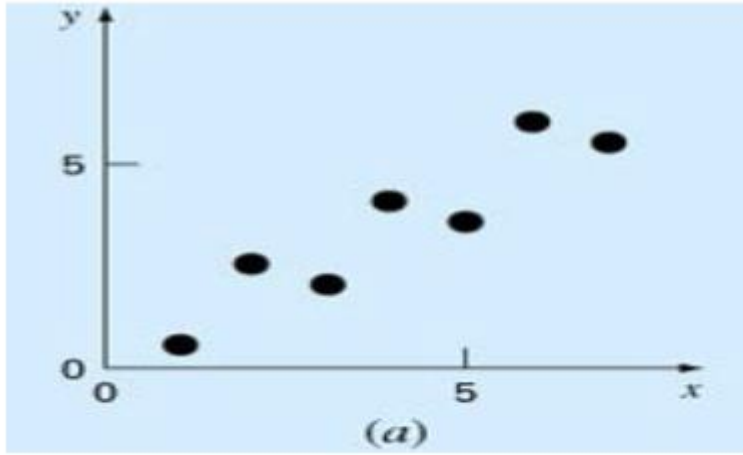
- ☞ Çoğu zaman bu veri tablosuna tam olarak uyan bir fonksiyon bulmak mümkün olmaz.
- ☞ Veri tablosuna en iyi uyan fonksiyon belirlenmeye çalışılır. Bir veri tablosuna en iyi uyan fonksiyonu bulma sürecine **regresyon analizi** denir.

REGRESYON

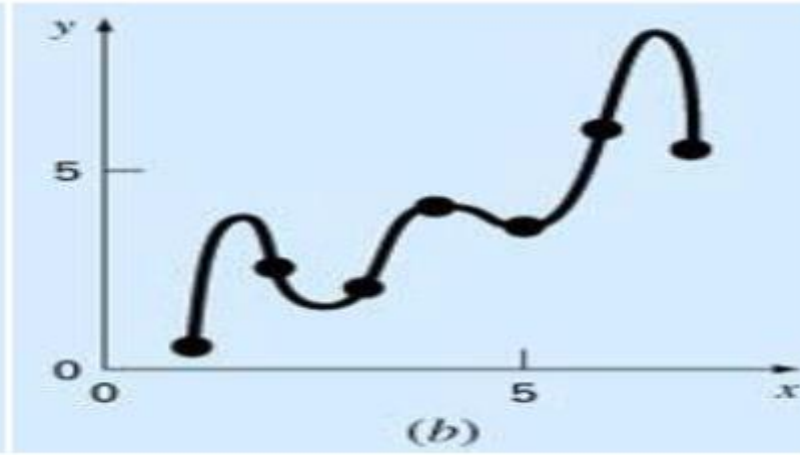
Verilerde **önemli hatalar** olduğunda, interpolasyon uygun değildir ve ara değerleri tahmin etmek için kullanıldığında tatmin edici sonuçlar vermez. Genellikle **deneysel veriler** bu tiptedir ve en küçük kareler yöntemi ile gösterimi daha iyi sonuçlar verir.

REGRESYON

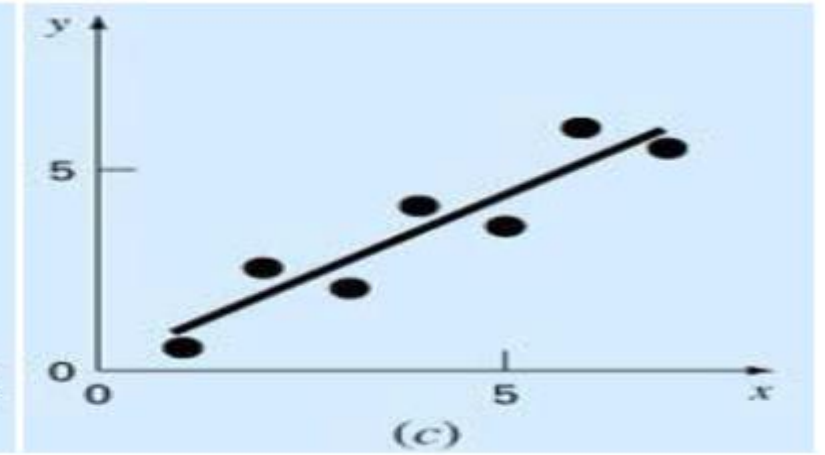
- ✓ Verilerinin belirli bir doğruyla gösterilebilmesi durumunda doğru uydurma yöntemine başvurulur.



Veriler



İnterpolasyon



En küçük kareler yöntemi

REGRESYON

∞ Bu yöntemde doğruya yaklaşımdaki hataların karelerinin toplamını minimum yapacak doğru denklemi araştırılır.

$$y = a_0 + a_1x + \varepsilon \quad \rightarrow \quad \varepsilon = y_{gerçek} - y_{yaklaşık} = y - (a_0 + a_1x)$$

$$S_r = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y - a_0 - a_1x_i)^2$$

a_0 : kesme noktası

a_1 : eğim

ε : hata veya artık

S_r : hataların kareleri toplamı

REGRESYON

∞ Hataların kareleri toplamını minimum yapacak a_0 ve a_1 değerleri, bu değerlerin türevlerinin sıfıra eşitlenmesiyle bulunur. Son durumda a_0 ve a_1 aşağıdaki gibi olur.

$$(n)a_0 + (\sum x_i)a_1 = \sum y_i$$

$$(\sum x_i)a_0 + (\sum x_i^2)a_1 = \sum x_i y_i$$

REGRESYON

∞ Aşağıdaki denklem takımı çözülerek a_0 ve a_1 değerleri bulunur.

$$a_1 = \bar{y} - a_1 \bar{x} = \frac{\sum y_i}{n} - a_1 \frac{\sum x_i}{n}$$

$$a_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

Not : $\sum x_i y_i$ ile $\sum x_i \sum y_i$ ve $\sum x_i^2$ ile $(\sum x_i)^2$ birbirinden farklı ifadelerdir.

REGRESYON

- ∞ Eğri uydurmanın uyumluluğunu belirlemek için korelasyon katsayısı (r) ile belirlenir. Doğrusal regresyon için korelasyon katsayısı aşağıdaki eşitlik ile hesaplanır:

$$r = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}$$

REGRESYON

Tablodaki deęerler için düz bir doğru uydurun.

x_i	y_i
1	0.5
2	2.5
3	2.0
4	4.0
5	3.5
6	6.0
7	5.5

REGRESYON

$$a_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} = \frac{7 \times 119.5 - 28 \times 24}{7 \times 140 - 28^2} = 0.839$$

$$a_1 = \bar{y} - a_1 \bar{x} = 3.429 - 0.839 \times 4 = 0.073$$

$$y = A + Bx = 0.073 + 0.839x$$

$$r = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}} = \frac{7 \times 119.5 - 28 \times 24}{\sqrt{7 \times 140 - 28^2} \sqrt{7 \times 105 - 24^2}} = \frac{164.5}{176.53} = 0.932$$

REGRESYON

Tablo değerlerinin girilmesi ; **MODE**→**3(STAT)** →**2(A+Bx)**

Tablo değerleri girildikten sonra **AC** tuşuna basarak veriler kaydedilir.

$\sum x^2$ =**SHIFT**→**1(STAT)** →**3(SUM)** →**1**

$\sum x$ =**SHIFT**→**1(STAT)** →**3(SUM)** →**2**

$\sum y^2$ =**SHIFT**→**1(STAT)** →**3(SUM)** →**3**

$\sum y$ =**SHIFT**→**1(STAT)** →**3(SUM)** →**4**

$\sum xy$ =**SHIFT**→**1(STAT)** →**3(SUM)** →**5**

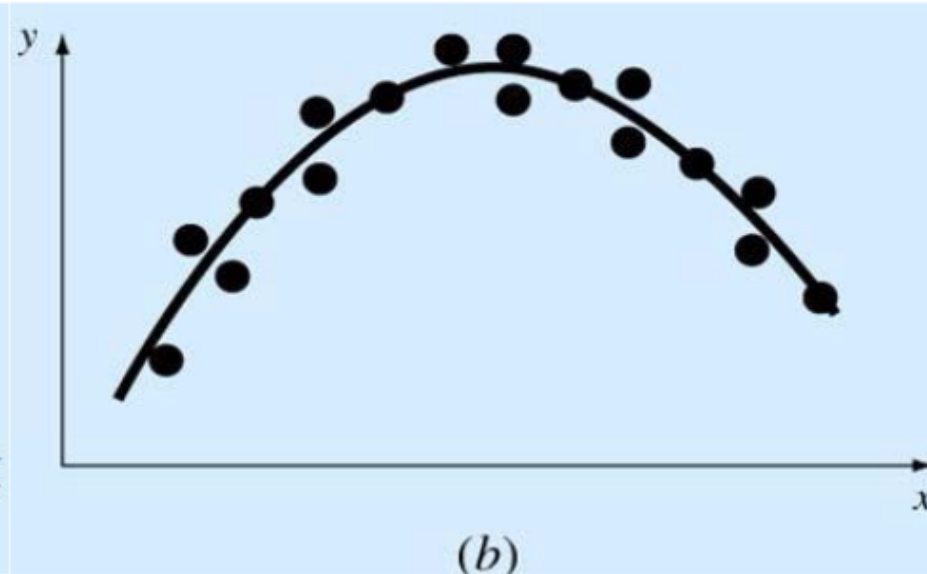
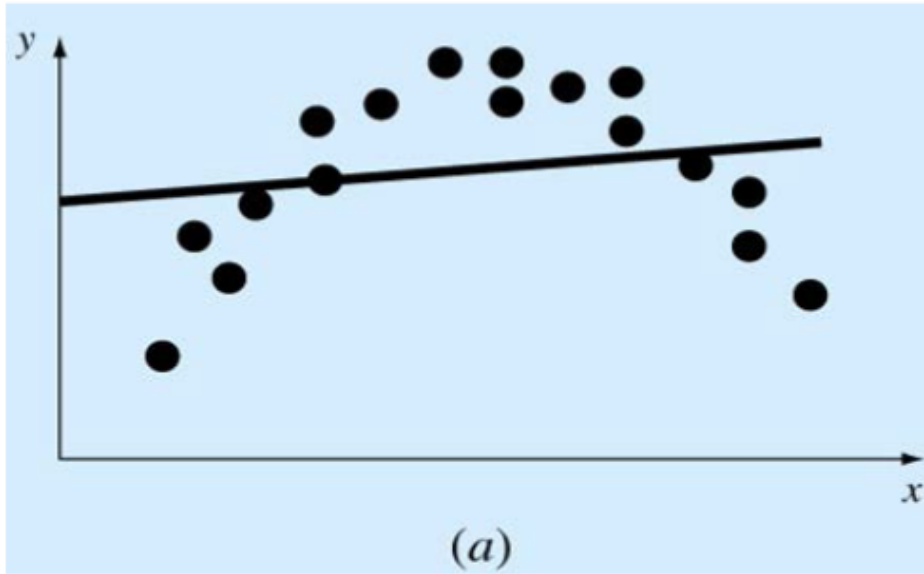
A =**SHIFT**→**1(STAT)** →**5(Reg)** →**1**

B =**SHIFT**→**1(STAT)** →**5(Reg)** →**2**

r =**SHIFT**→**1(STAT)** →**5(Reg)** →**3**

REGRESYON

- ✓ Bazı mühendislik verilerinin belirli bir şekli olsa da düz bir doğruyla gösterilmesi mümkün değildir. Böyle durumlarda bir eğri, verilere daha uyumlu olabilir ve polinom uydurmak doğru daha doğru sonuçlar verir.



REGRESYON

∞ En küçük kareler yöntemi, yüksek dereceli polinomlara eğri uydurmak için kolayca genişletilebilir:

$$y = a_0 + a_1x + a_2x^2 + \varepsilon$$

$$\rightarrow \varepsilon = y_{gerçek} - y_{yaklaşık} = y - (a_0 + a_1x + a_2x^2)$$

$$S_r = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (a_0 + a_1x + a_2x^2)^2$$

REGRESYON

☞ Hatanın karelerinin toplamını minimum yapacak a_0 , a_1 ve a_2 değerleri, bu değerlerin türevlerinin sıfıra eşitlenmesiyle bulunur. Son durumda aşağıdaki eşitlikler elde edilir.

$$n(a_0) + (\sum x_i) a_1 + (\sum x_i^2) a_2 = \sum y_i$$

$$(\sum x_i) a_0 + (\sum x_i^2) a_1 + (\sum x_i^3) a_2 = \sum x_i y_i$$

$$(\sum x_i^2) a_0 + (\sum x_i^3) a_1 + (\sum x_i^4) a_2 = \sum x_i^2 y_i$$

REGRESYON



Tablodaki deęerleri 2. dereceden polinoma yaklařtırın

x_i	y_i
0	2.1
1	7.7
2	13.6
3	27.2
4	40.09
5	61.1

REGRESYON

$$n(a_0) + (\sum x_i) a_1 + (\sum x_i^2) a_2 = \sum y_i \rightarrow 6a_0 + 15a_1 + 55a_2 = 152.6$$

$$(\sum x_i) a_0 + (\sum x_i^2) a_1 + (\sum x_i^3) a_2 = \sum x_i y_i \rightarrow 15a_0 + 55a_1 + 255a_2 = 585.6$$

$$(\sum x_i^2) a_0 + (\sum x_i^3) a_1 + (\sum x_i^4) a_2 = \sum x_i^2 y_i \rightarrow 55a_0 + 225a_1 + 979a_2 = 2488.8$$

REGRESYON

∞ Cramer yöntemiyle çözüm uygulanırsa;

$$1. \quad A = \begin{vmatrix} 152.6 & 15 & 55 \\ 585.6 & 55 & 225 \\ 2488.8 & 225 & 979 \end{vmatrix} \quad B = \begin{vmatrix} 6 & 152.6 & 55 \\ 15 & 585.6 & 225 \\ 55 & 2488.8 & 979 \end{vmatrix} \quad C = \begin{vmatrix} 6 & 15 & 152.6 \\ 15 & 55 & 585.6 \\ 55 & 225 & 2488.8 \end{vmatrix} \quad D = \begin{vmatrix} 6 & 15 & 55 \\ 15 & 55 & 225 \\ 55 & 225 & 979 \end{vmatrix}$$

$$2. \quad a_0 = \frac{\det A}{\det D} = \frac{979}{3920} = 2.479$$

$$a_1 = \frac{\det B}{\det D} = \frac{9248.4}{3920} = 2.359 \quad \rightarrow \quad a_0 + a_1 x + a_2 x^2 = 2.479 + 2.359x + 1.861 x^2$$

$$a_2 = \frac{\det C}{\det D} = \frac{7294}{3920} = 1.861$$

REGRESYON

Tablo değerlerinin girilmesi ; **MODE** → **3(STAT)** → **3(_+CX²)**

Tablo değerleri girildikten sonra **AC** tuşuna basarak veriler kaydedilir.

$\sum x^2 =$ **SHIFT** → **1(STAT)** → **3(SUM)** → **1**

$\sum x =$ **SHIFT** → **1(STAT)** → **3(SUM)** → **2**

$\sum y^2 =$ **SHIFT** → **1(STAT)** → **3(SUM)** → **3**

$\sum y =$ **SHIFT** → **1(STAT)** → **3(SUM)** → **4**

$\sum xy =$ **SHIFT** → **1(STAT)** → **3(SUM)** → **5**

REGRESYON

$$\sum x^3 = \text{SHIFT} \rightarrow 1(\text{STAT}) \rightarrow 3(\text{SUM}) \rightarrow 6$$

$$\sum x^2 y = \text{SHIFT} \rightarrow 1(\text{STAT}) \rightarrow 3(\text{SUM}) \rightarrow 3$$

$$\sum x^4 = \text{SHIFT} \rightarrow 1(\text{STAT}) \rightarrow 3(\text{SUM}) \rightarrow 8$$

$$A = \text{SHIFT} \rightarrow 1(\text{STAT}) \rightarrow 5(\text{Reg}) \rightarrow 1$$

$$B = \text{SHIFT} \rightarrow 1(\text{STAT}) \rightarrow 5(\text{Reg}) \rightarrow 2$$

$$r = \text{SHIFT} \rightarrow 1(\text{STAT}) \rightarrow 5(\text{Reg}) \rightarrow 3$$

REGRESYON

$y = a_0 x^{a_1}$ şeklindeki lineer olmayan denklemin, kuvvet fonksiyonunu elde etmek için her iki tarafın doğal logaritması alınır:

$$\ln y = \ln(a_0 x^{a_1}) \rightarrow \ln y = \ln(a_0) + a_1 \ln(x)$$

$$\ln y = Y \quad \ln a_0 = A \quad a_1 = B \quad \ln x = X \text{ alınırsa } Y = A + BX$$

$$\varepsilon = \sum_{i=1}^n (Y_i - A - Bx_i)^2$$

REGRESYON

↻ Hatanın karelerinin toplamını minimum yapacak A ve B değerleri, bu değerlerin türevlerinin sıfıra eşitlenmesiyle bulunur. Son durumda aşağıdaki eşitlikler elde edilir :

$$(n)A + (\sum x_i)B = \sum Y_i$$

$$(\sum x_i)A + (\sum x_i^2)B = \sum x_i y_i$$

REGRESYON

- ✓ Regresyon ve Makine Öğrenmesi (Machine Learning)
- ☞ Çok büyük miktarlardaki verinin elle işlenmesi ve analizinin yapılması mümkün değildir.
- ☞ Amaç geçmişteki verileri kullanarak gelecek için tahminlerde bulunmaktır.
- ☞ Bu problemleri çözmek için Makine Öğrenmesi (machine learning) yöntemleri geliştirilmiştir.

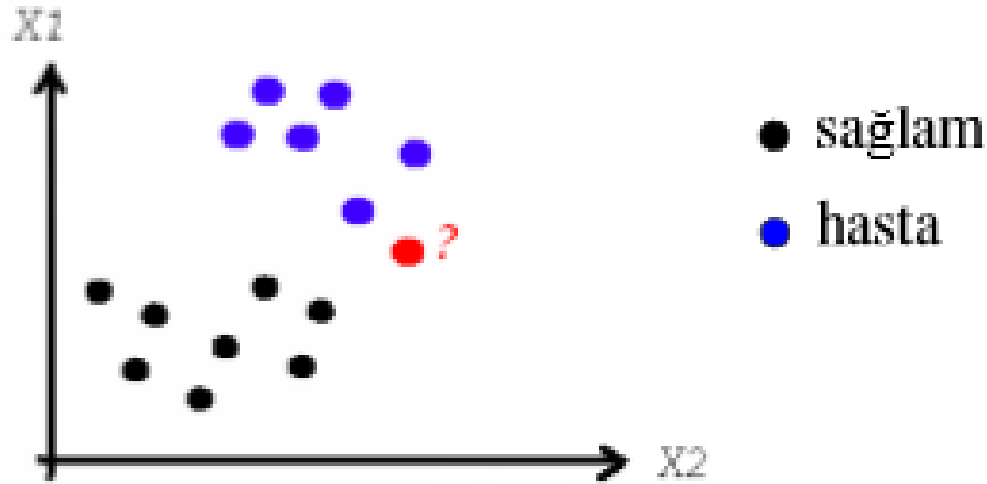


☞ Makine öğrenmesinde metodlar;

- Makine öğrenmesi yöntemleri, geçmişteki veriyi kullanarak yeni veri için en uygun modeli bulmaya çalışır.
- Verinin incelenip, içerisinden işe yarayan bilginin çıkarılmasına da Veri Madenciliği (data mining) adı verilir.
- Farklı uygulamaların, analizlerden farklı beklentileri olmaktadır.
- Makine öğrenmesi metodlarını bu beklentilere göre sınıflandırmak mümkündür.

SINIFLANDIRMA

- Geçmiş bilgileri hangi sınıftan olduğu biliniyorsa, yeni gelen verinin hangi sınıfa dahil olacağını bulunmasıdır.



Kırmızı hangi sınıfa dahildir ?

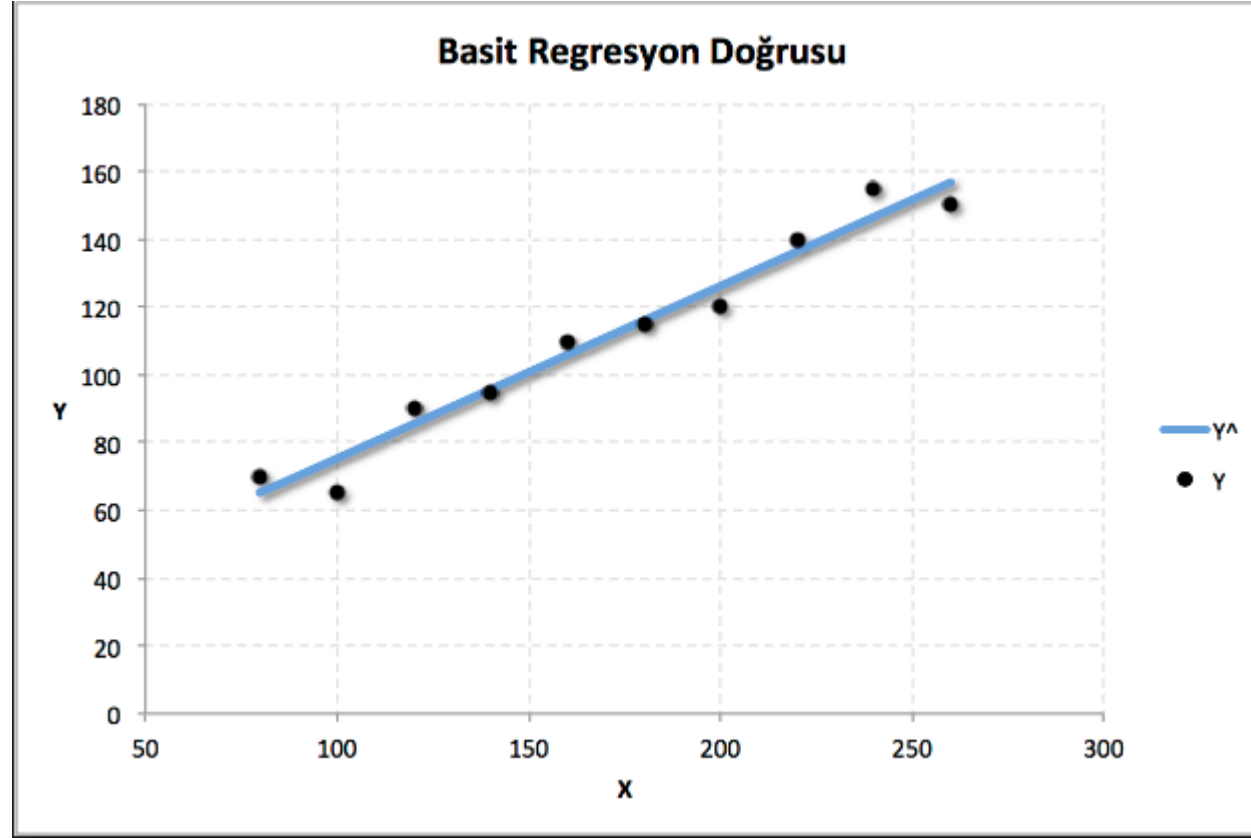
KÜMELEME

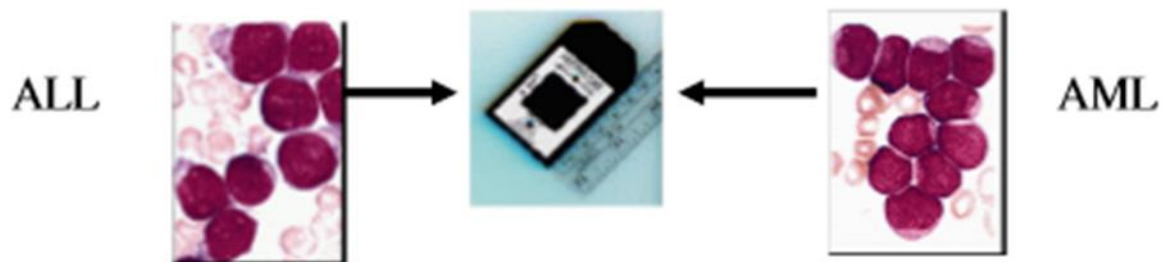
- Geçmişteki verilerin sınıfları/etiketleri verilmediği/bilinmediği durumlarda verilerin birbirlerine yakın benzerliklerinin yer aldığı kümelerin bulunmasıdır.
- 256 rengi 16 renge nasıl indiririz ?



REGRESYON-EĐRİ UYDURMA

Geçmiş bilgilere ait sınıflar yerine sürekli bilginin yer aldığı problemlerdir.





Gen Mikrodizilimleri

100 kişinin (hasta/sağlam) elimizde gen dizilimleri var.
Bu dizilimleri analiz ederek hasta olup olmadığı bilinmeyen birisinin hasta olup olmadığını ya da hastalığının türünü öğrenebilir miyiz ?

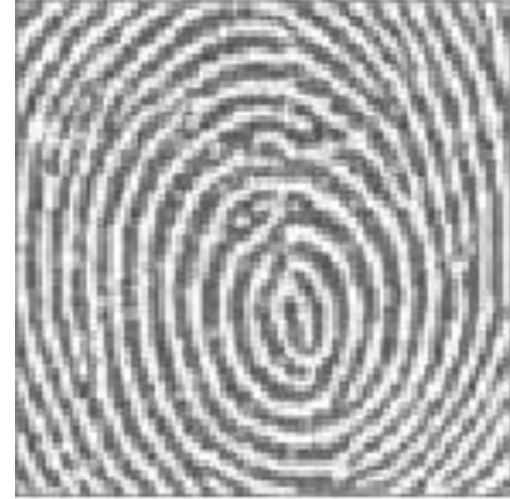
En iyi tedaviyi önerebilir miyiz ?

Nasıl ? Elimizde hangi bilgiler olmalı ?

Bu adam kim? İçeri girsin mi?



Parmak izi kime ait?



Bu imza kimin ? Yoksa taklit mi?

Bu ses kimin ?
Bu ses ne diyor ?



Teşekkürler.



Dersin Sonu

Kocaeli Üniversitesi Bilgisayar Mühendisliği
Yapay Zeka ve Benzetim Sistemleri Ar-Ge Lab.
<http://yapbenzet.kocaeli.edu.tr/>