



Sayısal Analiz

Ders İçeriği

7.
Hafta

- ❖ EN KÜÇÜK KARELER YÖNTEMİ
- ❖ Örnekler

Sayısal Analiz

- **En küçük kareler yöntemi**, birbirine bağlı olarak değişen iki fiziksel büyülüük arasındaki matematiksel bağlantıyı, mümkün olduğunca gerçeğe uygun bir denklem olarak yazmak için kullanılan, standart bir regresyon yöntemidir.
- Yöntem, ölçüm sonucu elde edilmiş veri noktalarına "**mümkün olduğu kadar yakın**" geçecek bir fonksiyon eğrisi bulmaya yarar.

Sayısal Analiz

- Gerçek yaşamın çeşitli alanlarında herhangi bir uygulama ile toplanan veriler tablo şekline getirilerek incelenir ve toplanan veriyi modelleyen bir fonksiyon bulunmaya çalışılır. 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.

YÖNTEMİN AMACI :

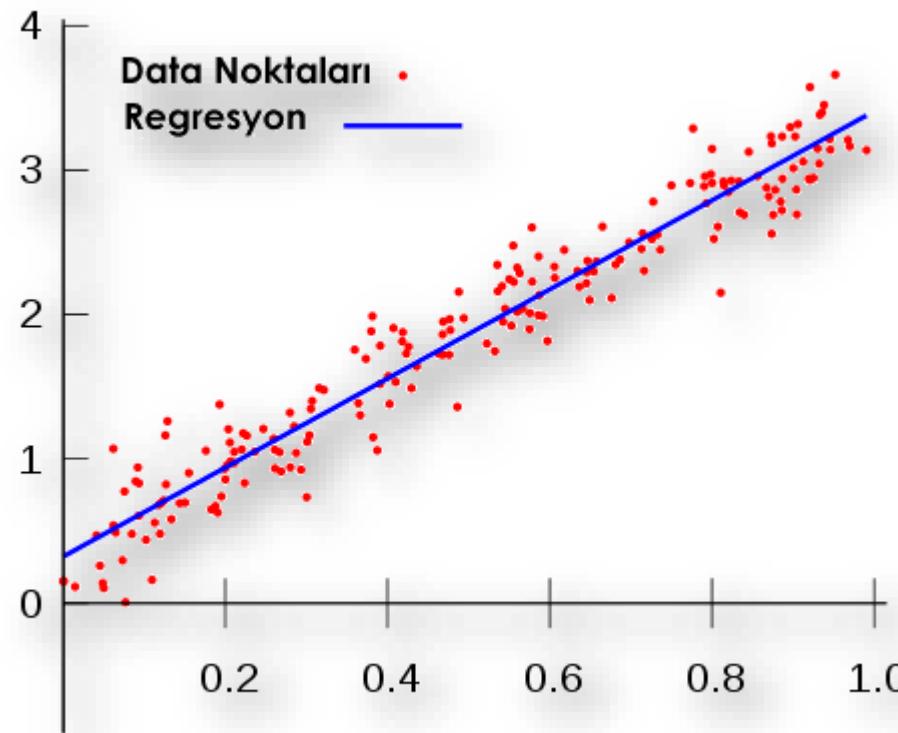
- Aralarında doğrusal (lineer) bir bağlantı olan, **x** ve **y** adında iki fiziksel büyülük olduğunu varsayılm.

İki büyülük arasındaki bağlantı doğrusal olduğunu göre,
denklem

$$y = a x + b$$

dir.

- Aranılan durum, bu denklemdeki a ve b sayıları için mümkün olan en doğru değerlerdir.
- Değerleri belirlemek için bir dizi ölçüm yaptığımızı düşünelim.



- Belli ölçümler sonucunda $i = 1, 2, \dots, n$ için (x_i, y_i) verileri elde edilmiş olsun.
- Burada, her bir y_i 'nin x_i 'ye bağlı olarak değiştiği varsayılmaktadır.
- (x_i, y_i) düzlemde noktalar olarak düşünüldüğünde, pratikte bu noktalar düzgün bir eğri üzerinde, (başka bir deyimle, bilinen bir fonksiyonun grafiği üzerinde) bulunmazlar.
- Bazı durumlarda, (x_i, y_i) 'ler arasında ne tür bir bağıntı bulunduğu dahi bilinmeyebilir.

Ancak, yapılan ölçümlerin doğası gereği,

her $i = 1, 2, \dots, n$ için $y_i = f(x_i)$

olacak biçimde bir fonksiyonun var olduğu, ölçümlerde yapılan hata nedeniyle bu eşitlıkların bazıları veya hepsinin sağlanmadığı kabul edilebilir.

- Ölçülen y_i değeri $f(x_i)$ için yaklaşık değer kabul edilerek bu yaklaşımındaki hatanın minimum olduğu f fonksiyonu belirlenmeye çalışılır.
- Bu amacı gerçekleştirmek için f fonksiyonunun bir takım parametrelere bağlı bir ifadesi bulunduğu varsayılıp eldeki veriler yardımıyla bu parametreler belirlenmeye çalışılır.

Örneğin ;

f fonksiyonu

$$y = f(x) = mx + b$$

ifadesinde olduğu gibi bir doğrusal fonksiyon veya

$$y = f(x) = ax^2 + bx + c$$

ifadesinde olduğu gibi bir karesel fonksiyon olabilir ki bu durumda belirlenmesi gereken parametreler a, b, c, m dir.

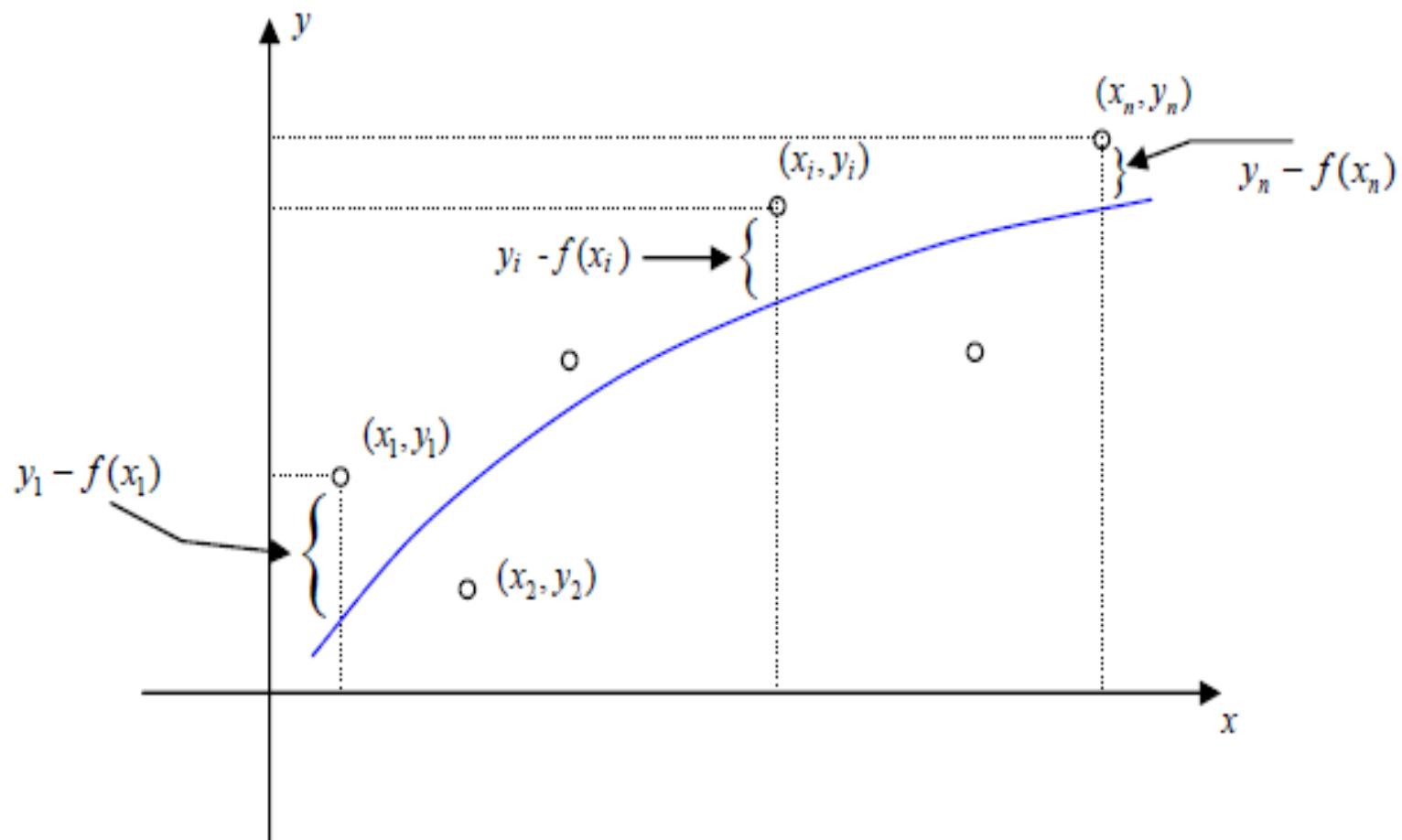
y_i değeri $f(x_i)$ için yaklaşık değer,

$$f(x_i) \approx y_i,$$

kabul edilince yapılan hata

$$y_i - f(x_i)$$

dir ve amaç, bu hatalar minimum olacak şekilde bir f fonksiyonu bulmaktır.



$y_i - f(x_i)$ farklarından her birine bir artık denir.

En küçük kareler yönteminde aranan fonksiyon, ya da onun parametreleri tüm artıkların kareleri toplamı olan

$$\sum_{i=1}^n (y_i - f(x_i))^2 = (y_1 - f(x_1))^2 + \dots + (y_n - f(x_n))^2$$

ifadesini minimum yapacak şekilde belirlenir.

Bu, yönteme neden **en küçük kareler yöntemi** dendiğini açıklar.

Sözü edilen kareler toplamının minimum olması için her bir hatanın küçük olması gereğine dikkat ediniz.

Bir veri tablosuna en iyi uyan doğrusal fonksiyonun grafiği olan doğruya **regresyon doğrusu** veya **en küçük kareler doğrusu** denir.

$$y=f(x)=mx + b$$

fonksiyonunun belirlediği doğruya yani regresyon doğrusunu bulalım.

Önce $y_i - f(x_i)$

artık değerleri bulunarak bunların karelerinin toplamı olan

$$F(m, b) = \sum_{i=1}^n (y_i - mx_i - b)^2 = (y_1 - mx_1 - b_1)^2 + \dots + (y_n - mx_n - b_n)^2$$

fonksiyonu oluşturulur.

Her m ve b için $f(m,b) \geq 0$ olduğundan ve $y = f(x) = mx + b$ doğrusu verilen noktalardan uzaklaştıkça $f(m,b)$ sonsuza iraksayacağından, $f(m,b)$ 'nin bir mutlak minimum değeri vardır ve bu minimum değer $f(m,b)$ fonksiyonunun bir kritik noktasında ortaya çıkar.

Örnek :

Bir üretici, ürettiği ürünün çeşitli üretim seviyeleri için maliyetini belirliyor ve aşağıdaki tabloyu oluşturuyor:

Ürün sayısı (x yüz)	2	5	6	9
Maliyet (y bin YTL)	4	6	7	8

Bu üretici için gider fonksiyonunu yukarıdaki tabloya en iyi uyan doğrusal fonksiyon olarak belirleyelim.

Elimizdeki veri tablosu, düzlemede şu (x,y) noktalarını verir:

$$(2,4) , (5,6) , (6,7) \text{ ve } (9,8).$$

Bu noktaların hepsini üzerinde bulunduran bir doğru yoktur.

Amacımız, bu noktalara uyan doğruya, yani regresyon doğrusunu bulmaktır.

Regresyon doğrusunun denklemi

$$y=c(x)=mx+b$$

m ve b belirlenerek bulunacaktır.

Artıklar hesaplanarak ve veri tablosunu genişletilerek ;

Ürün sayısı (x yüz)	2	5	6	9
Maliyet (y bin YTL)	4	6	7	8
$mx + b$	$2m + b$	$5m + b$	$6m + b$	$9m + b$
$y - mx - b$	$4 - 2m - b$	$6 - 5m - b$	$7 - 6m - b$	$8 - 9m - b$

Artıkların kareleri toplamı aşağıdaki iki değişkenli fonksiyonu tanımlar:

$$f(m,b) = (4-2m-b)^2 + (6-5m-b)^2 + (7-6m-b)^2 + (8-9m-b)^2$$

Bu fonksiyonun hangi **m** ve **b** değerleri için minimum değeri aldığı belirlemeliyiz.

Kısmi türevleri hesaplayalım:

$$f_m(m,b) = 2(4-2m-b)(-2) + 2(6-5m-b)(-5) + 2(7-6m-b)(-6) + 2(8-9m-b)(-9) = 0,$$

$$f_b(m,b) = 2(4-2m-b)(-1) + 2(6-5m-b)(-1) + 2(7-6m-b)(-1) + 2(8-9m-b)(-1) = 0.$$

...

aritmetik işleminden sonra

$$\mathbf{146m+22b = 152}$$

$$\mathbf{22m+ 4b = 25} \quad \text{denklem sistemi elde edilir.}$$

Bu sistemi eliminasyon yöntemi ile çözersek;
İkinci denklem $-11/2$ ile çarpılıp birinci denkleme toplanırsa

$$25m = 14.5 \Rightarrow m = 0.58$$

ve m nin bu değeri ikinci denklemde yerine konulursa

$$12.76 + 4b = 25 \Rightarrow 4b = 11.24 \Rightarrow b = 3.06$$

elde edilir.

Görüldüğü gibi, sistemin tek çözümü vardır:

$$m = 0.58, b = 3.06. m \text{ ve } b \text{ nin}$$

bu değerleri için $f(m,b)$ ‘nin minimum olduğunu biliyoruz.

O halde regresyon doğrusu

$$y=0.58x+3.06$$

dur.

Başka bir deyimle, regresyon analizi sonucu ortaya çıkan gider fonksiyonu

$$C(x)=0.58x+3.06$$

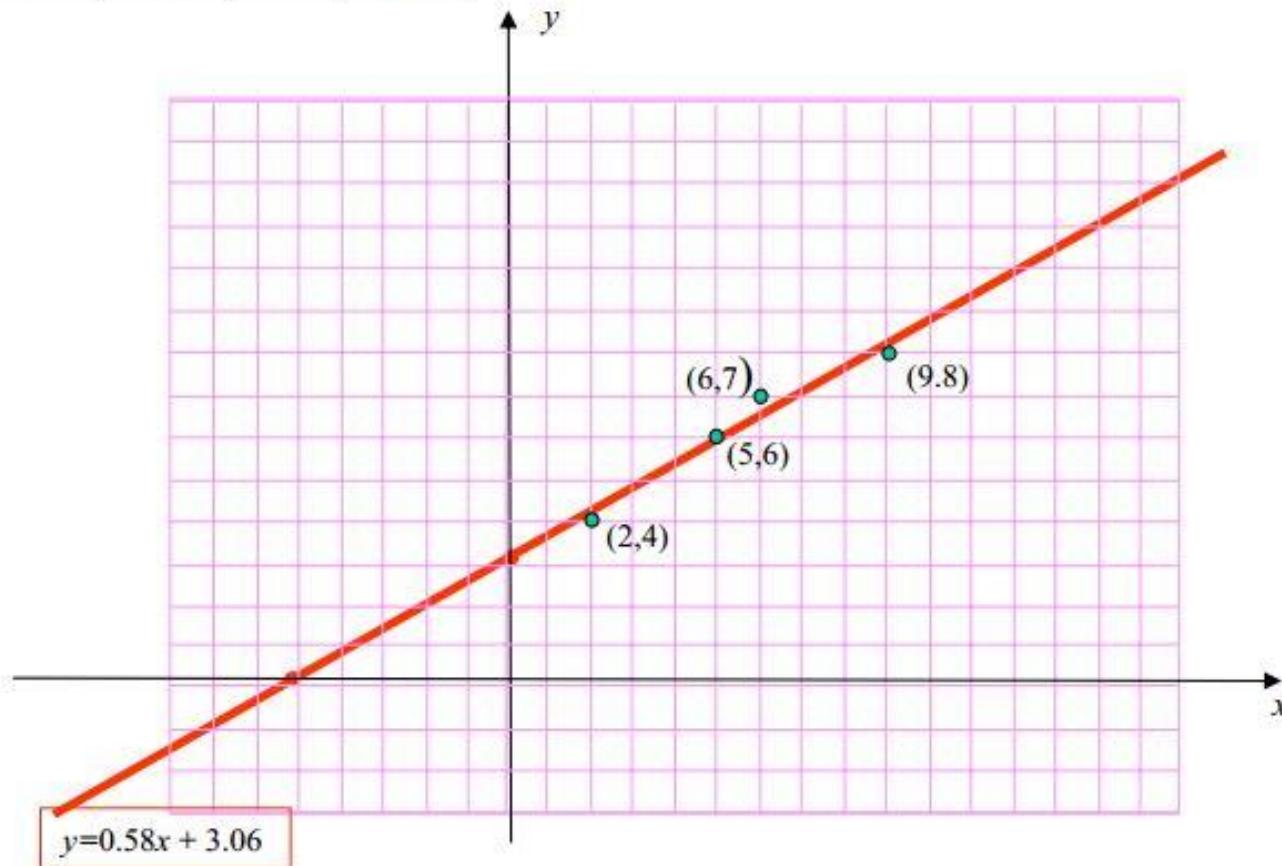
Denklemi ile verilenen fonksiyondur.

Üretici, örneğin, 4 ürün üretince giderinin ne olacağını tahmin edebilir :

$$C(4)=(0.58)(4)+3.06 = 2.32+3.06 = 5.38$$

Örnek veri tablosu ve regresyon doğrusu

x	2	5	6	9
y	4	6	7	8



Önceki örnekte izlenen yol genel duruma uygulanarak **m** ve **b** nin doğrudan hesaplanması sağlayacak formüller elde edilebilir.
Artıkların karelerinin toplamı olarak tanılanan

$$F(m,b) = \sum_{i=1}^n (y_i - mx_i - b)^2 = (y_1 - mx_1 - b)^2 + \dots + (y_n - mx_n - b)^2$$

fonksiyonunun kritik noktaları

$$F_m(m,b) = \sum_{i=1}^n 2(y_i - mx_i - b)(-x_i) = -2 \left(\sum_{i=1}^n x_i^2 \right) m - 2 \left(\sum_{i=1}^n x_i \right) b + 2 \left(\sum_{i=1}^n x_i y_i \right) = 0$$

$$F_b(m,b) = \sum_{i=1}^n 2(y_i - mx_i - b)(-1) = -2 \left(\sum_{i=1}^n x_i \right) m - 2 \left(\sum_{i=1}^n 1 \right) b + 2 \left(\sum_{i=1}^n y_i \right) = 0$$

ya da

$$\begin{cases} \left(\sum_{i=1}^n x_i^2 \right) m + \left(\sum_{i=1}^n x_i \right) b = \sum_{i=1}^n x_i y_i \\ \left(\sum_{i=1}^n x_i \right) m + nb = \sum_{i=1}^n y_i \end{cases}$$

denklem sistemi çözüлerek bulunur.

Bu denklem sisteminin daima tek bir çözüm vardır.

$$m = \frac{n \left(\sum_{k=1}^n x_k y_k \right) - \left(\sum_{k=1}^n x_k \right) \left(\sum_{k=1}^n y_k \right)}{n \left(\sum_{k=1}^n x_k^2 \right) - \left(\sum_{k=1}^n x_k \right)^2},$$

$$b = \frac{\sum_{k=1}^n y_k - m \left(\sum_{k=1}^n x_k \right)}{n}.$$

Örnek : $(0, 6.4)$, $(1, 2.6)$, $(2, 0.5)$, $(3, 0.6)$ ve $(4, 0.3)$ veri noktalarına en iyi uyan doğrunun denklemini bulunuz.

Çözüm: Aranılan doğrunun denklemi $y = mx + b$ olmak üzere veri noktalarına karşılık gelen aşağıdaki tabloyu oluşturalım:

x_i	0	1	2	3	4
y_i	6.4	2.6	0.5	0.6	0.3

m ve **b** yi formüllerden elde edebilmek için bu formüldeki her bir terimin değerini buluruz.

Öncelikle, **n = 5** olduğuna dikkat ederek

x_i	0	1	2	3	4
y_i	6.4	2.6	0.5	0.6	0.3

$$\sum_{k=1}^5 x_k = 0 + 1 + 2 + 3 + 4 = 10$$

$$\sum_{k=1}^5 y_k = 6.4 + 2.6 + 0.5 + 0.6 + 0.3 = 10.4$$

$$\sum_{k=1}^5 x_k y_k = 0 \cdot (6.4) + 1 \cdot (2.6) + 2 \cdot (0.5) + 3 \cdot (0.6) + 4 \cdot (0.3) = 2.6 + 1 + 1.8 + 1.2 = 6.6$$

$$\sum_{k=1}^5 x_k^2 = 0 + 1 + 4 + 9 + 16 = 30$$

$$(\sum_{k=1}^5 x_k)(\sum_{k=1}^5 y_k) = 10 \cdot (10.4) = 104$$

$$m = \frac{n(\sum_{k=1}^n x_k y_k) - (\sum_{k=1}^n x_k)(\sum_{k=1}^n y_k)}{n(\sum_{k=1}^n x_k^2) - (\sum_{k=1}^n x_k)^2} = \frac{5 \cdot (6.6) - 104}{5 \cdot 30 - 100} = \frac{33 - 104}{50} = \frac{-71}{50} = -1.42,$$

$$b = \frac{\sum_{k=1}^n y_k - m(\sum_{k=1}^n x_k)}{n} = \frac{10.4 - (-1.42) \cdot 10}{5} = \frac{10.4 + 14.2}{5} = \frac{24.6}{5} = 4.92.$$

O halde istenilen doğrunun denklemi

$$y = -1.42x + 4.92$$

dir.

Problemler

1. Aşağıdaki veri tablolarından her biri için en küçük kareler doğrusu(regresyon doğrusu)nu bulunuz. Veri tablosuna karşılık gelen noktaları ve regresyon doğrusunu grafikle gösteriniz.

a)

x	1	2	3	4
y	1	3	4	3

b)

x	1	2	3	4
y	8	5	4	0

c)

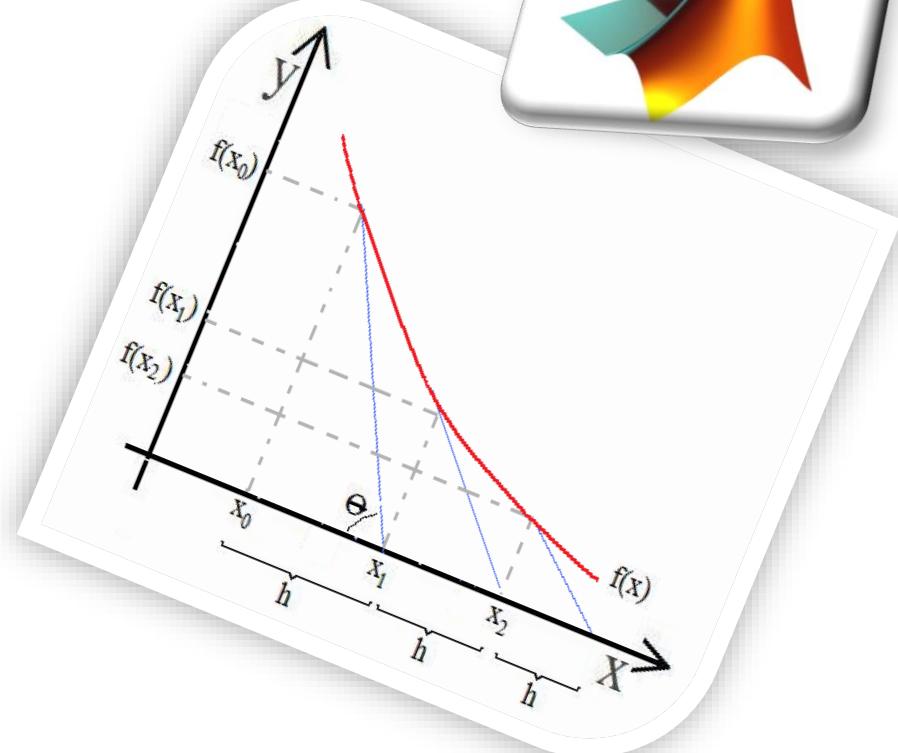
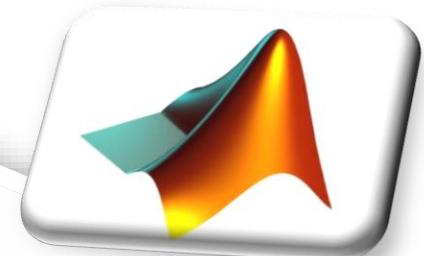
x	1	2	3	4
y	2	3	3	2

ç)

x	1	2	3	4	5
y	2	3	3	2	3

Kaynaklar

Sayısal Analiz S.Akpınar
Başkent Univ. Mat.Ders Notları



Sonraki Hafta :

Eğri uydurma, aradeğer ve dış değer bulma yöntemleri...