Makine Öğrenmesi

Doğrusal Regresyon

İlhan AYDIN

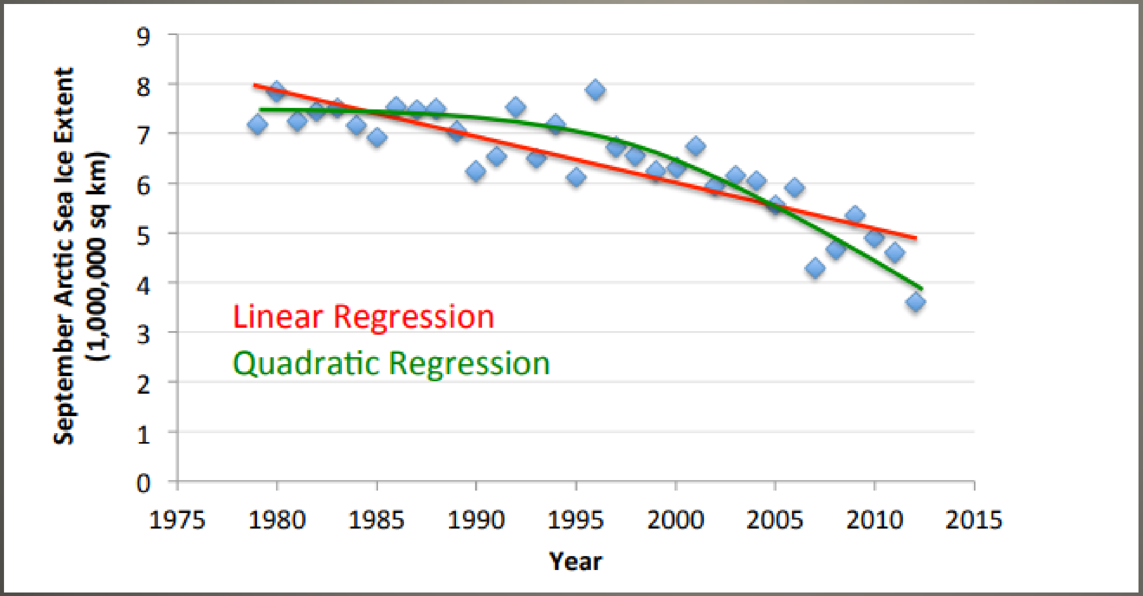
- Birçok boyutlu lineer regresyon
- Bileşik özellikler, lineer olmayan ilişki modelleme
- Model oluşturma
- Ek: *lineer regresyon normal denklemleri*

Ders planı

- Regresyon probleminde, olayın modeli sürekli bir modeldir, yani
- Modellenecek değişkenler sürekli şekilde değişmesi gerekiyor
- Makine öğrenme problemi, bu modellenecek değişkenlerin değişimi için karar verme için uygun modeli oluşturmak

Regresyon problemi

- Verilen
  - Veri X={x1,….xn} burada xi dboyutlu uzaydadır.
  - Verilee karşılık gelecek etiketler y={y1,…,yn} ve yi değerleri reel sayıdır.



Regresyon problemi

- 67 örnek eğitim / 30 teste bölünmüş toplam 97 örnek
- Sekiz tahmin edici (özellikler):
- – 6 sürekli (4 log dönüşümü), 1 binary, 1 sıralı
- Sürekli sonuç değişkeni:
  - lpsa: log(prostat spesifik anIgen seviyesi)

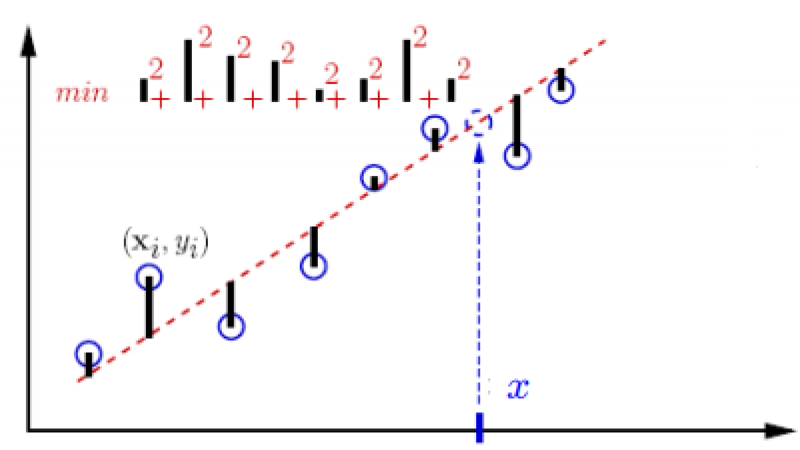
Term	Coefficient	Std. Error	Z Score
Intercept	2.46	0.09	27.60
lcavol	0.68	0.13	5.37
lweight	0.26	0.10	2.75
age	−0.14	0.10	−1.40
lbph	0.21	0.10	2.06
svi	0.31	0.12	2.47
lcp	−0.29	0.15	−1.87
gleason	−0.02	0.15	−0.15
pgg45	0.27	0.15	1.74

Prostat Kanser Veri Kümesi

- Hipotez:

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_d x_d = \sum_{j=0}^{d} \theta_j x_j$$

- Karesi alınan hataların toplamını en aza indirerek modeli sığdır



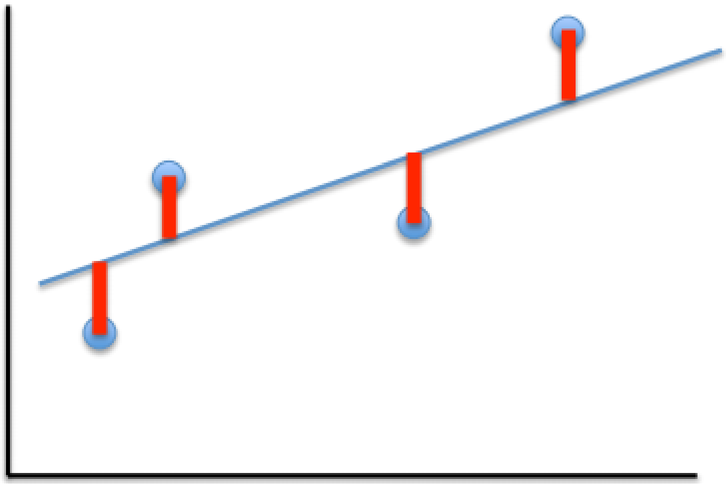
En küçük kareler
Uydurma çizgi, tahmin edici olarak kullanılır

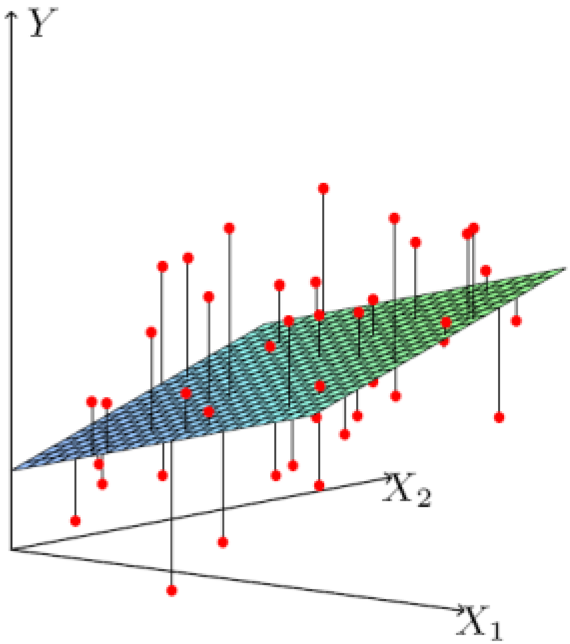
Doğrusal Regresyon

- Maliyet fonksiyonu:

$$J(\boldsymbol{\theta}) = \frac{1}{2n} \sum_{i=1}^{n} \left(h_{\boldsymbol{\theta}} \left(\boldsymbol{x}^{(i)} \right) - y^{(i)} \right)^2$$

$\min_{\boldsymbol{\theta}} J(\boldsymbol{\theta})$ değerini bul



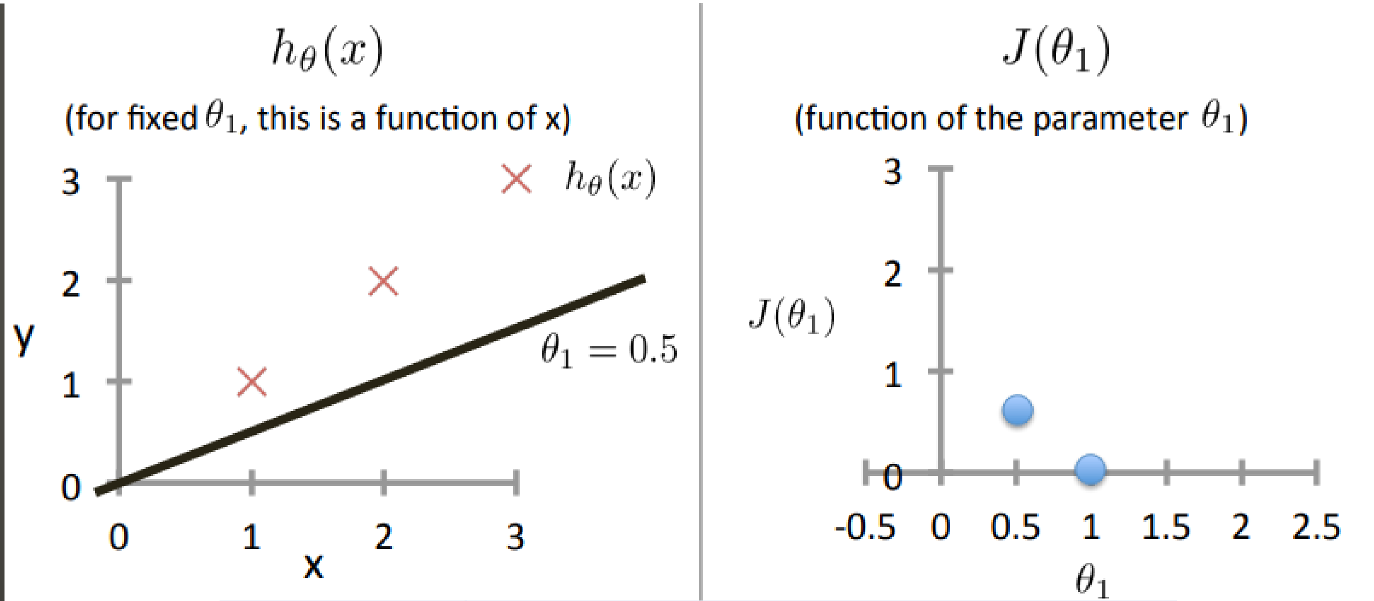


En Küçük Kareler Doğrusal Regresyon

$$J(\boldsymbol{\theta}) = \frac{1}{2n}\sum_{i=1}^{n}\left(h_{\boldsymbol{\theta}}\left(\boldsymbol{x}^{(i)}\right) - y^{(i)}\right)^2$$

J 'yi çözmek için elimizde x giriş özellikleri mevcut. O zaman aşağıdaki iki parametreyi ayarlamak gerekir.

$$\boldsymbol{\theta} = [\theta_0, \theta_1]$$



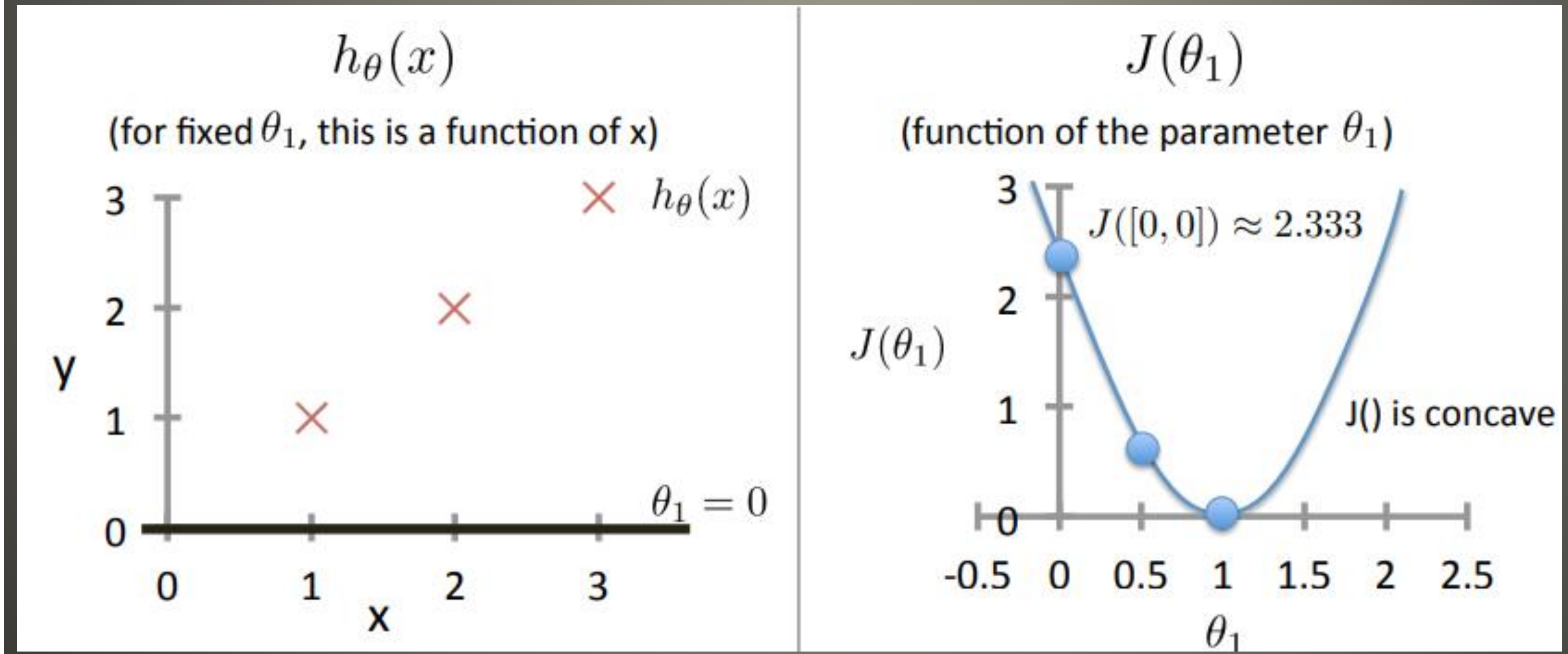
$$J([0, 0.5]) = \frac{1}{2 \times 3}\left[(0.5-1)^2 + (1-2)^2 + (1.5-3)^2\right] \approx 0.58$$

Maliyet Fonksiyonunun Arkasındaki Fikir

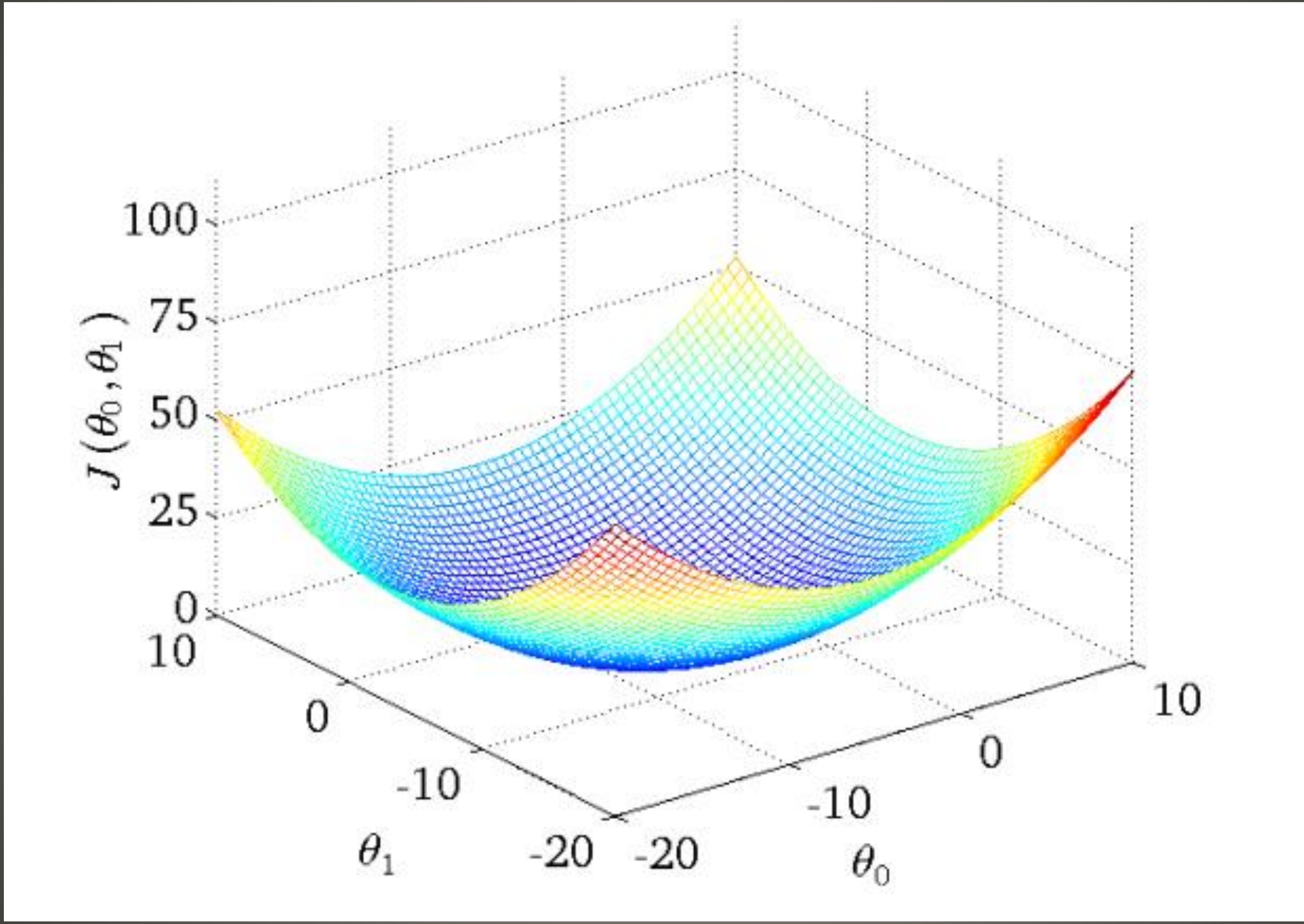
$$J(\boldsymbol{\theta}) = \frac{1}{2n}\sum_{i=1}^{n}\left(h_{\boldsymbol{\theta}}\left(\boldsymbol{x}^{(i)}\right) - y^{(i)}\right)^2$$

J ‘yi çözmek için elimizde x giriş özellikleri mevcut. O zaman aşağıdaki iki parametreyi ayarlamak gerekir.

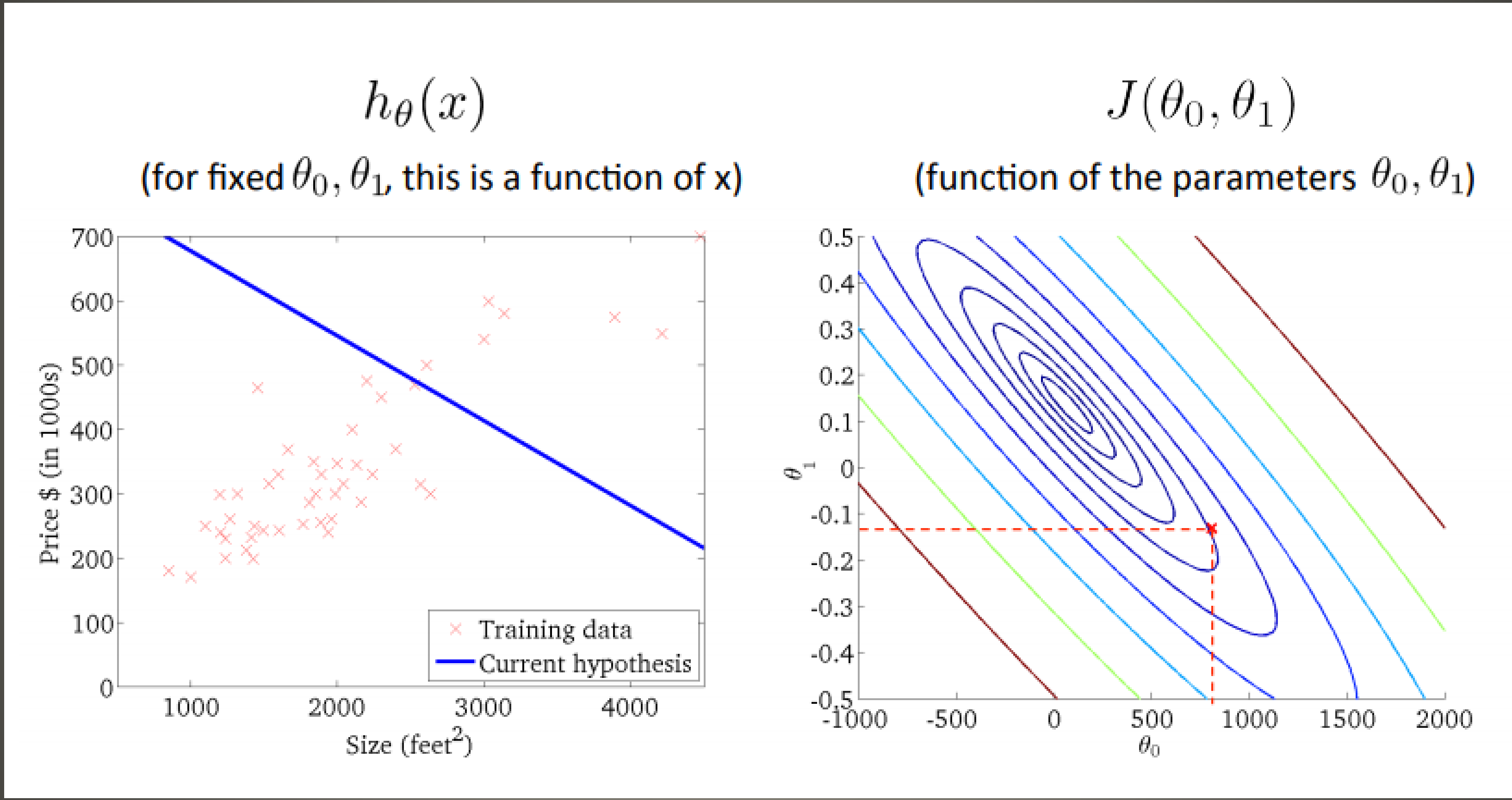
$$\boldsymbol{\theta} = [\theta_0, \theta_1]$$



Maliyet Fonksiyonunun Arkasındaki Fikir



Maliyet Fonksiyonunun Arkasındaki Fikir



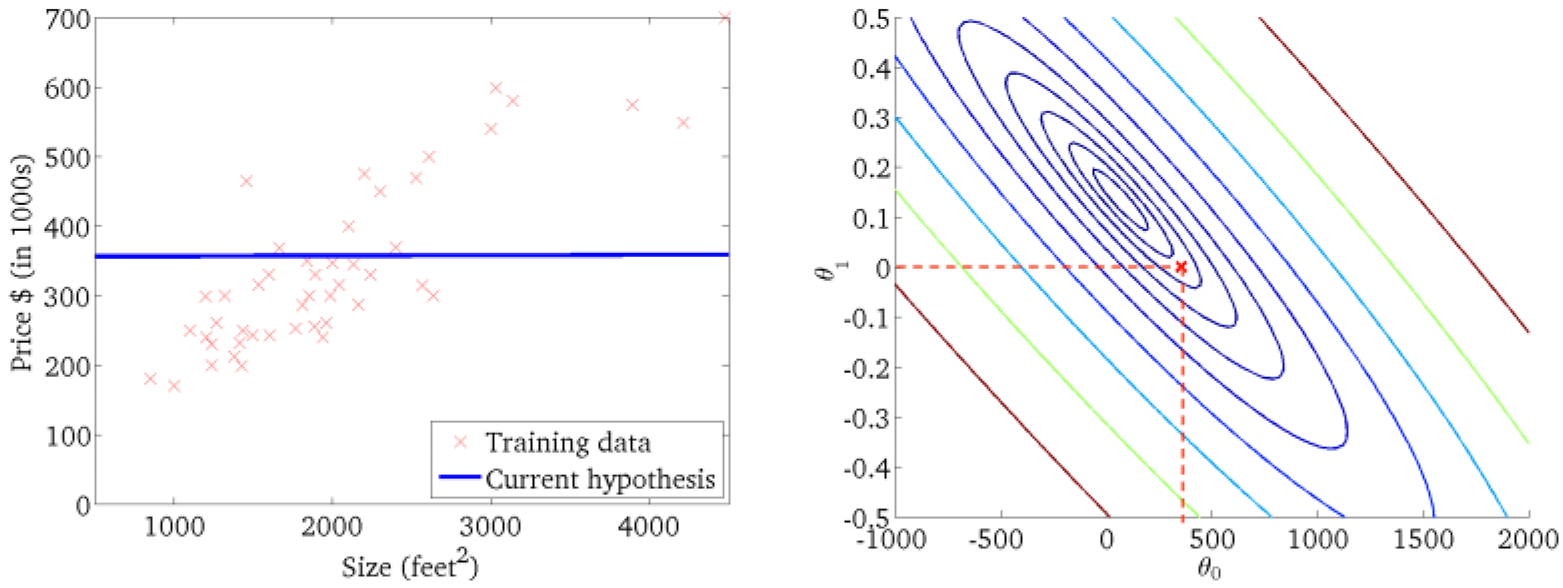
Maliyet Fonksiyonunun Arkasındaki Fikir

$h_\theta(x)$

(for fixed θ_0, θ_1, this is a function of x)

$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)



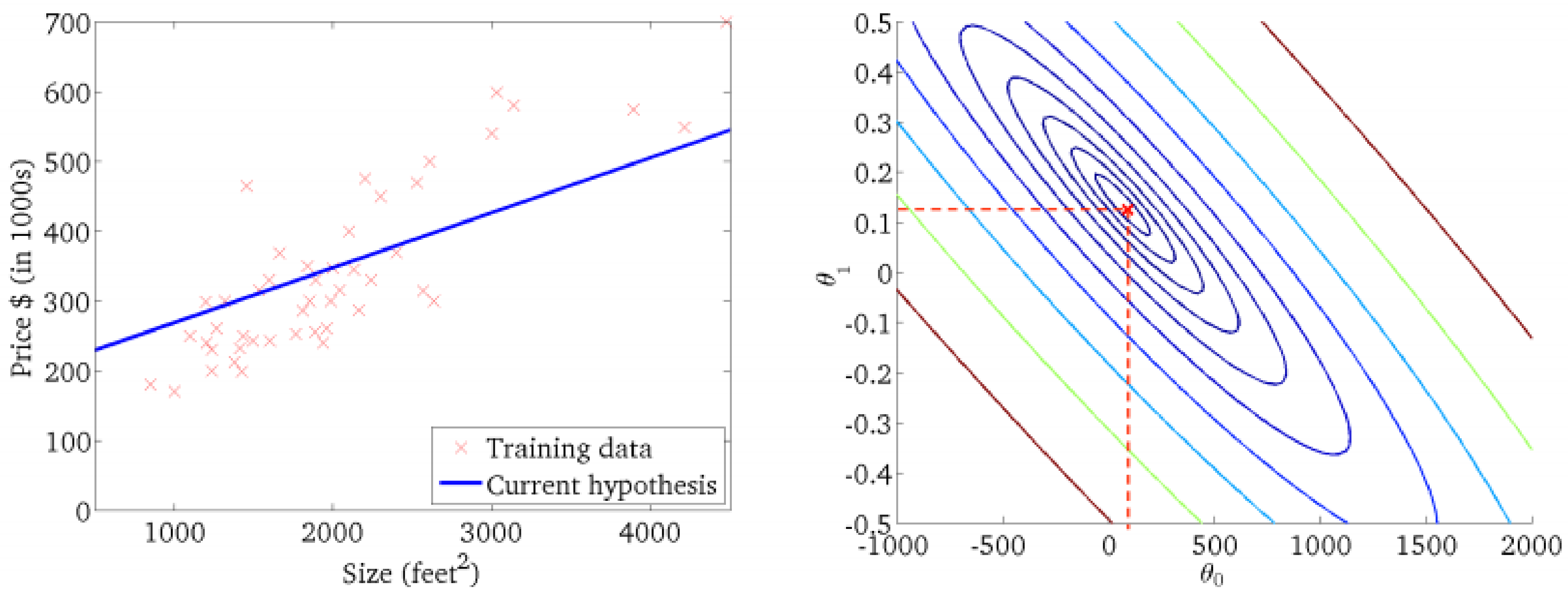
Maliyet Fonksiyonunun Arkasındaki Fikir

$h_\theta(x)$

(for fixed θ_0, θ_1, this is a function of x)

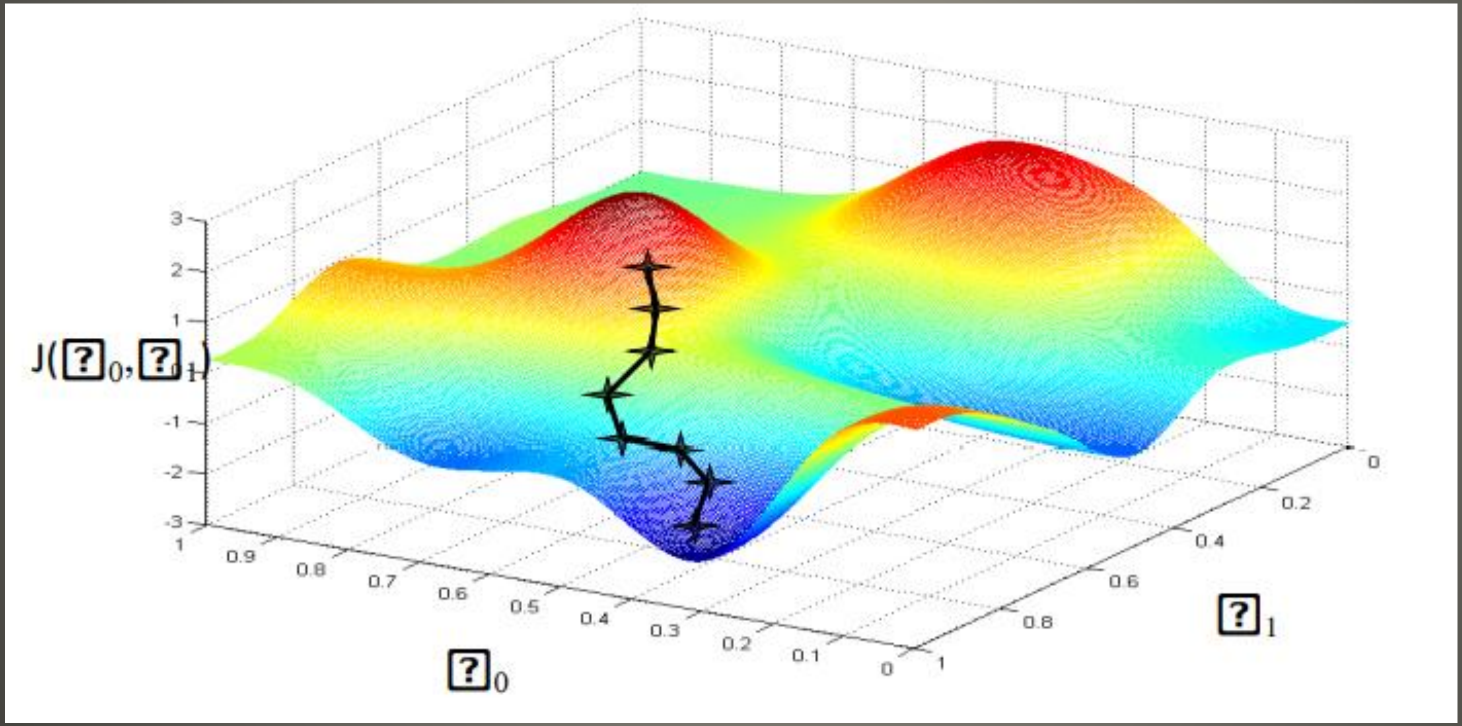
$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)



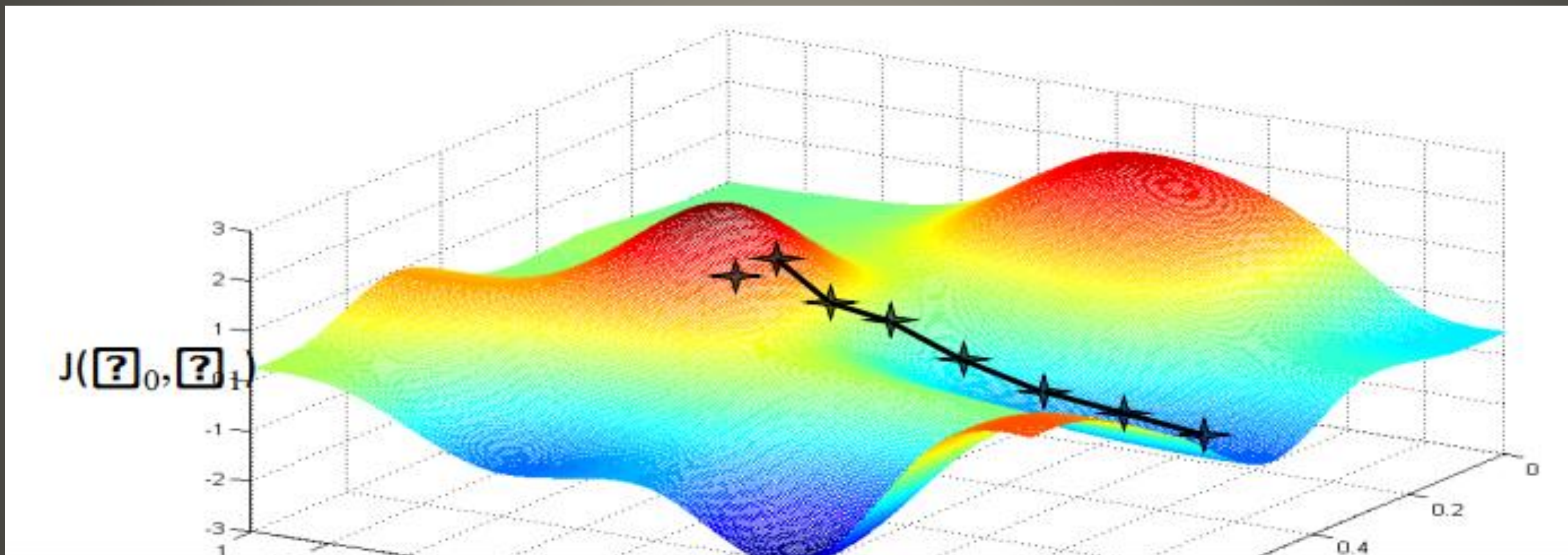
Maliyet Fonksiyonunun Arkasındaki Fikir

- θ değeri için başlangıç değeri seç
- Minimuma ulaşıncaya kadar devam et
  - J(θ)'yı azaltacak θ için yeni değerleri seç



Temel Arama Prosedürü

- θ değeri için başlangıç değeri seç
- Minimuma ulaşıncaya kadar devam et
  - J(θ)’yı azaltacak θ için yeni değerleri seç

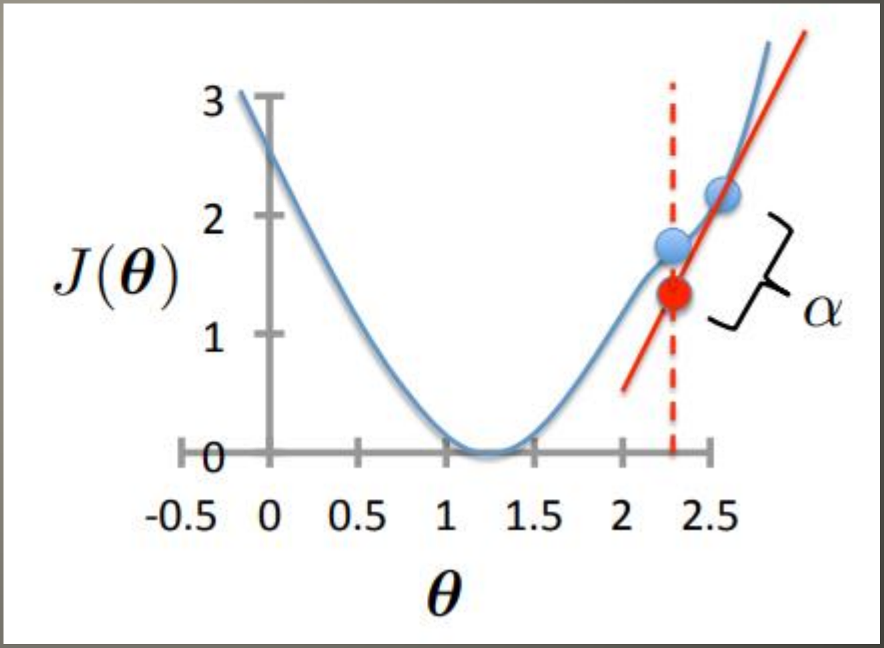


En küçük kareler amaç fonksiyonu dışbükey (içbükey) olduğundan, yerel minimumlar hakkında endişelenmemize gerek yoktur.

Temel Arama Prosedürü

- θ değerlerini başlat
- Yakınsama sağlanana kadar
  - $\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$ j=0,1,…,d için anlık güncelleme

Öğrenme oranı α=0.05
gibi küçük bir değer seçilir



Dereceli Azaltma

- θ değerlerini başlat
- Yakınsama sağlanana kadar
  - $\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$ j=0,1,…,d için anlık güncelleme

$$Lineer\ regresyon\ için: \frac{\partial}{\partial \theta_j} J(\theta) = \frac{\partial}{\partial \theta_j} \frac{1}{2n} \sum_{i=1}^{n} (h_\theta\left(x^{(i)} - y^{(i)}\right)^2$$

Dereceli Azaltma

- θ değerlerini başlat
- Yakınsama sağlanana kadar
  - $\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$ j=0,1,…,d için anlık güncelleme

$$Lineer\ regresyon\ için: \frac{\partial}{\partial \theta_j} J(\theta) = \frac{\partial}{\partial \theta_j} \frac{1}{2n} \sum_{i=1}^{n} (h_\theta\left(x^{(i)} - y^{(i)}\right)^2$$

$$= \frac{\partial}{\partial \theta_j} \frac{1}{2n} \sum_{i=1}^{n} \left(\sum_{k=0}^{d} \theta_k x_k^{(i)} - y^{(i)}\right)^2$$

Dereceli Azaltma

- θ değerlerini başlat
- Yakınsama sağlanana kadar
  - $\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$ j=0,1,…,d için anlık güncelleme

Lineer regresyon için

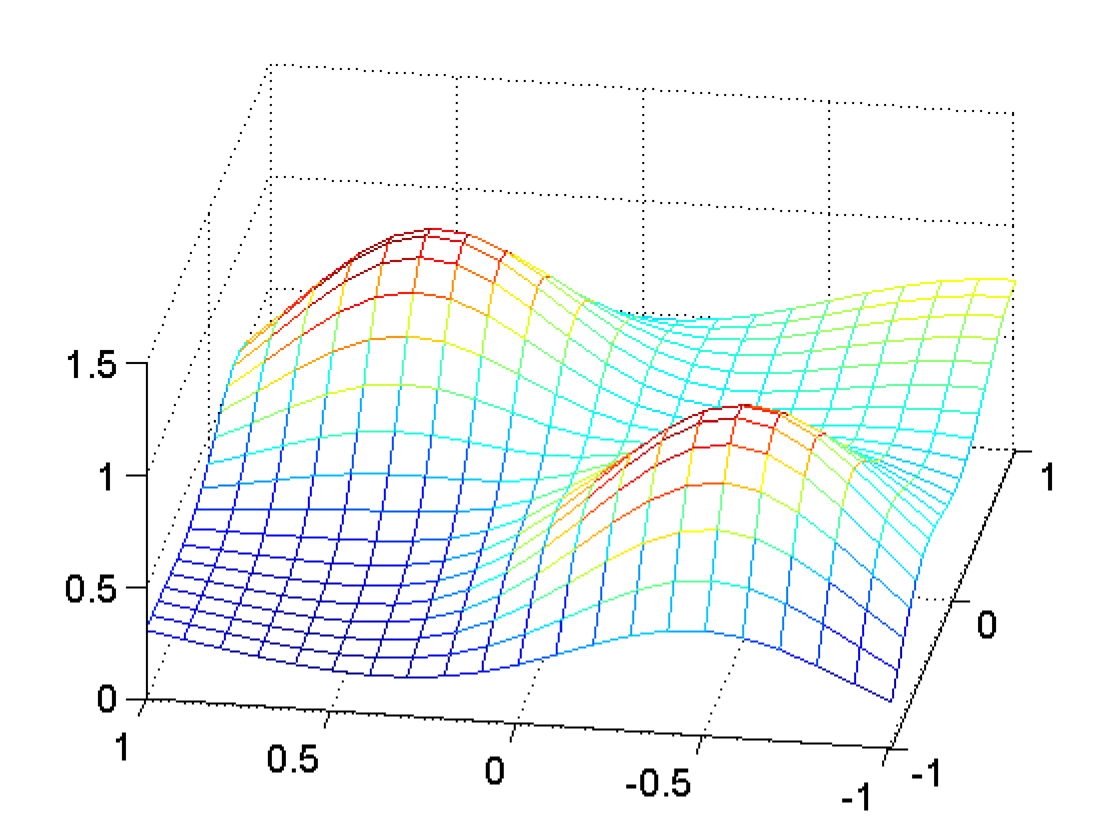
$$
\begin{aligned}
\frac{\partial}{\partial \theta_j} J(\boldsymbol{\theta}) &= \frac{\partial}{\partial \theta_j} \frac{1}{2n} \sum_{i=1}^{n} \left(h_{\boldsymbol{\theta}}\left(\boldsymbol{x}^{(i)}\right) - y^{(i)} \right)^2 \\
&= \frac{\partial}{\partial \theta_j} \frac{1}{2n} \sum_{i=1}^{n} \left(\sum_{k=0}^{d} \theta_k x_k^{(i)} - y^{(i)} \right)^2 \\
&= \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{k=0}^{d} \theta_k x_k^{(i)} - y^{(i)} \right) \times \frac{\partial}{\partial \theta_j} \left(\sum_{k=0}^{d} \theta_k x_k^{(i)} - y^{(i)} \right) \\
&= \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{k=0}^{d} \theta_k x_k^{(i)} - y^{(i)} \right) x_j^{(i)}
\end{aligned}
$$

Dereceli Azaltma

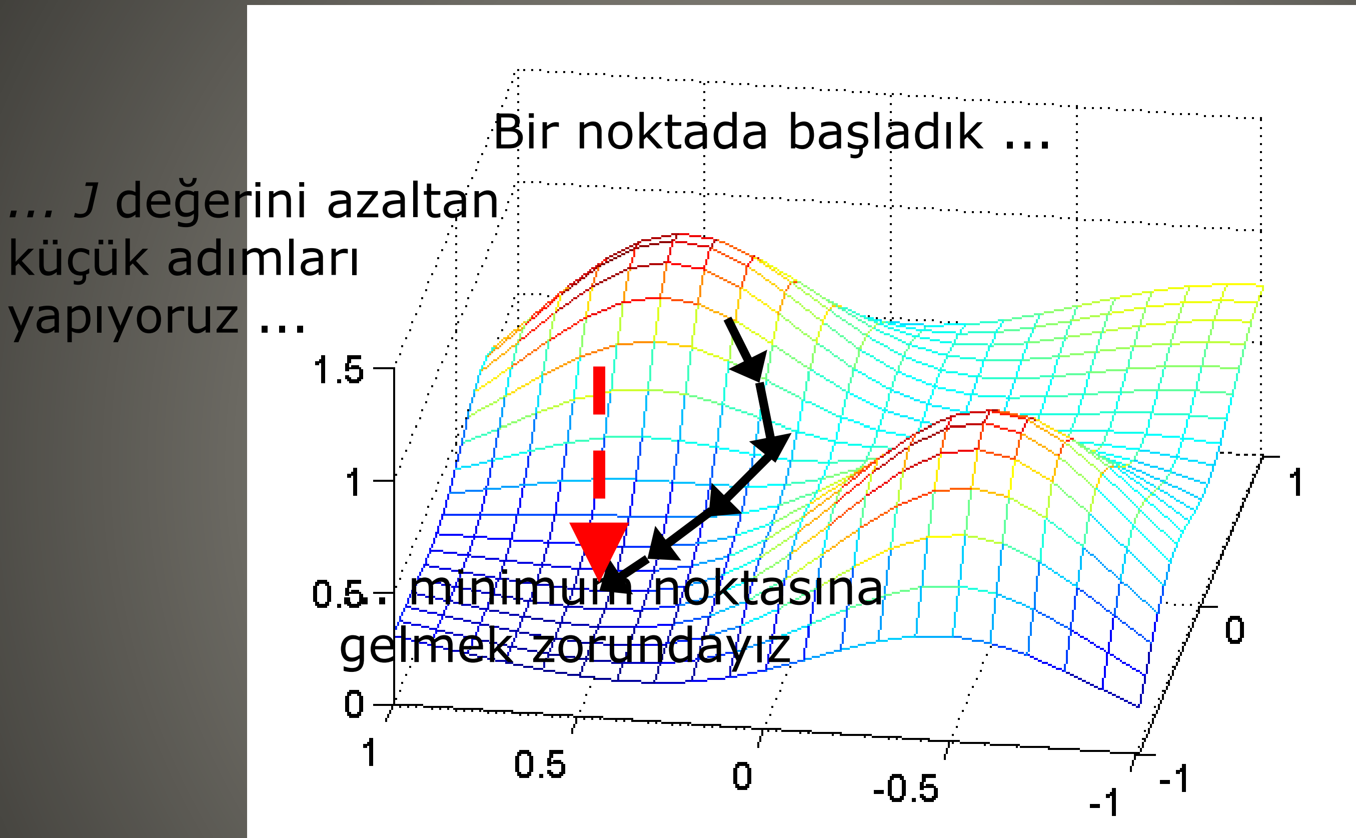
Dereceli azaltma metodu

- Dereceli azaltma (gradient descent) metodu, çok güçlü ve çok genel optimizasyon metodudur
  - Bir (θ_0, θ_1) noktasında başlıyoruz
  - Devamlı, J'nin değerlerini azaltmak için (θ_0, θ_1) uzayda küçük adımlarını yapıyoruz
  - J'nin değerleri sürekli düşmek zorunda
  - Çünkü $J \geq 0$, bu süreç sonunda bir noktaya gelmek zorundadır (sonsuz devam edemez yani)

Azaltılacak fonksiyon



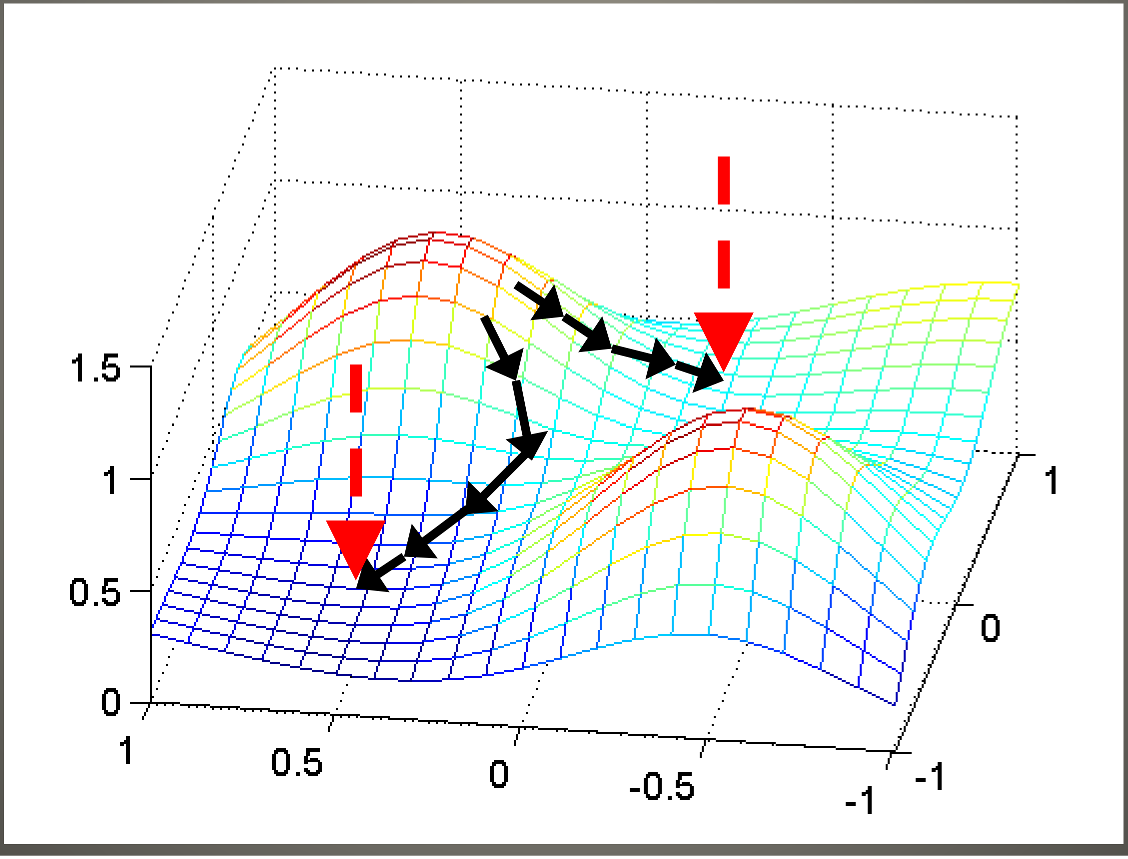
Dereceli azaltma metodu



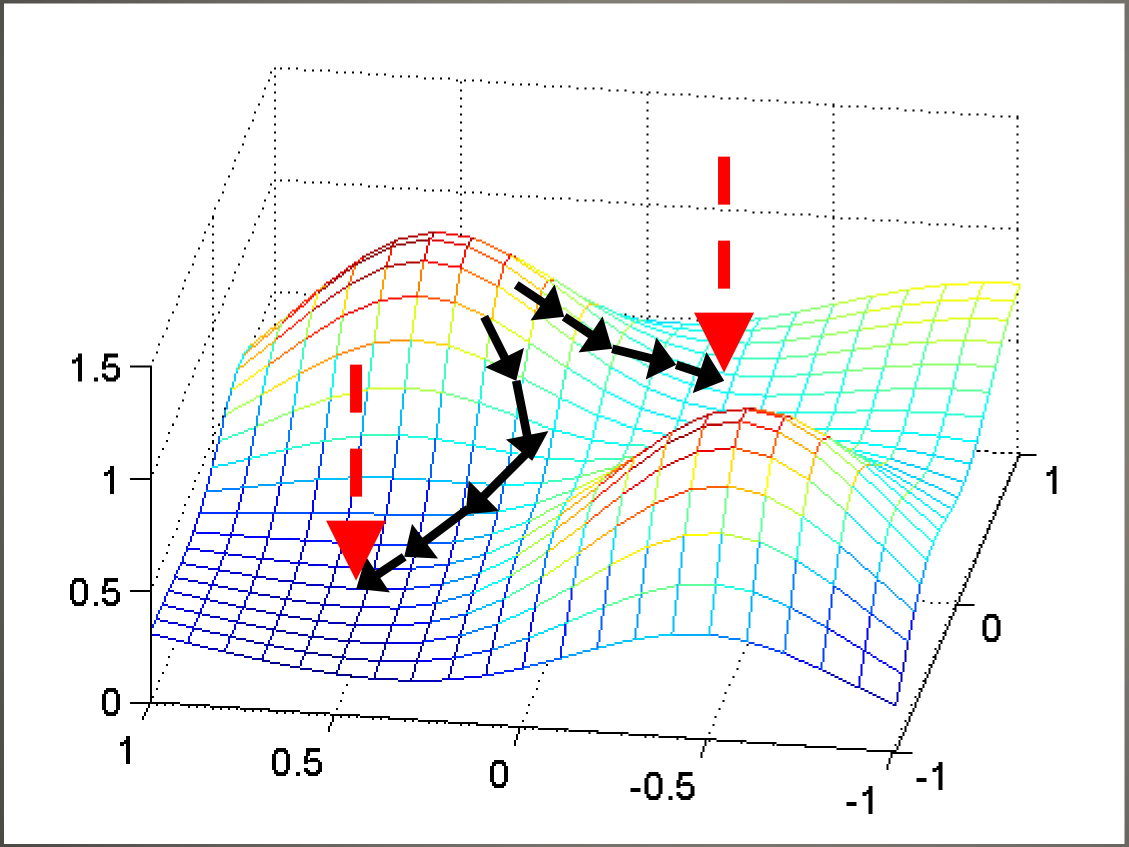
Dereceli azaltma metodu

Dereceli azaltma metodu

Sadece lokal olarak bir minimumdur: başka noktadan başlayınca başka noktaya gelmek mümkündür



Genellikle, bu metot birkaç rastgele başlangıç nokta ile çalıştırılmalı, ve en iyi minimum seçilmeli



Dereceli azaltma metodu

- Ortadaki adımları, J değerini en çok azaltması isteriz
- Bunun için, adımları "gradient" (yani eğim) yönünde yapılmaktadır
- Dereceli azaltma algoritması:
  - Yakınsamaya kadar tekrarlama {

j=1,2 için;

}

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

Dereceli azaltma metodu

- Yakınsamaya kadar tekrarlayın {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

j=1,2 için;

}

Dereceli azaltma metodu

Dereceli azaltma metodu

- Bu formülde

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

Değer güncelleştirilmesi

Maliyet fonksiyonununtürevleri

Özel bir parametre (öğrenme hız parametresi)

- Yakınsamaya kadar tekrarlayın {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

j=1,2 için;

}

Önemli Not:

Türevler, şu andaki (θ_0,θ_1) noktası için hepsi döngüden önce hesaplanmalı. Sonra, θ_0 ve θ_1 değerleri güncelleştirilmesi gerekiyor.

θ_0,θ_1'nin güncelleştirilmesi türevler hesaplanmasıyla aynı zamanda yapılmaz ! yani, (θ_0,θ_1) parça-parça şekilde güncelleştirilmez!

Dereceli azaltma metodu

- Yakınsamaya kadar tekrarlayın {

  j=1,2 için;

}

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

Yakınsamaya kadar tekrarlayın {

$$\theta_0 := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$

$$\theta_1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$

}

Dereceli azaltma metodu

- Yakınsamaya kadar tekrarlayın {

  j=1,2 için;

}

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

Yakınsamaya kadar tekrarlayın {

$$\theta_0 := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$

$$\theta_1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$

}

Yanlış

İlk adımda değişmiş oldu !!

Dereceli azaltma metodu

- Yakınsamaya kadar tekrarlayın {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

j=1,2 için;

}

Doğru

Yakınsamaya kadar tekrarlayın {

$$\theta_0 := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$

$$\theta_1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$

}

Yakınsamaya kadar tekrarlayın {

$$temp_0 := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$

$$temp_1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$

$$\theta_0 := temp_0$$

$$\theta_1 := temp_1$$

}

AYNI !

Dereceli azaltma metodu

- Lineer regresyonu

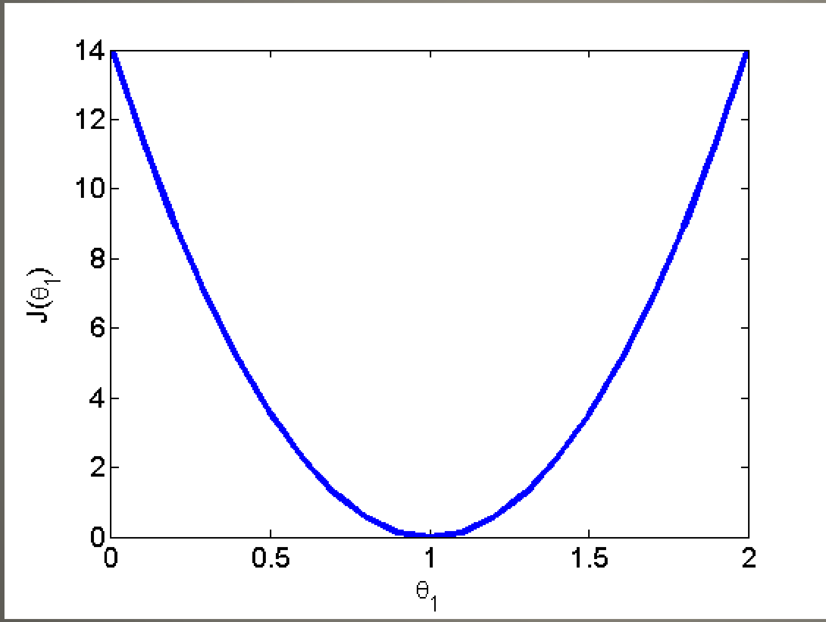
$$(J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^i) - y^i)^2)$$

$$\theta_0 := \theta_0 - \alpha \frac{1}{m}\sum_{i=1}^{m}(\overline{\theta}_0 + \overline{\theta}_1 x^i - y^i)$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m}\sum_{i=1}^{m} x^i \cdot (\overline{\theta}_o + \overline{\theta}_1 x^i - y^i)$$

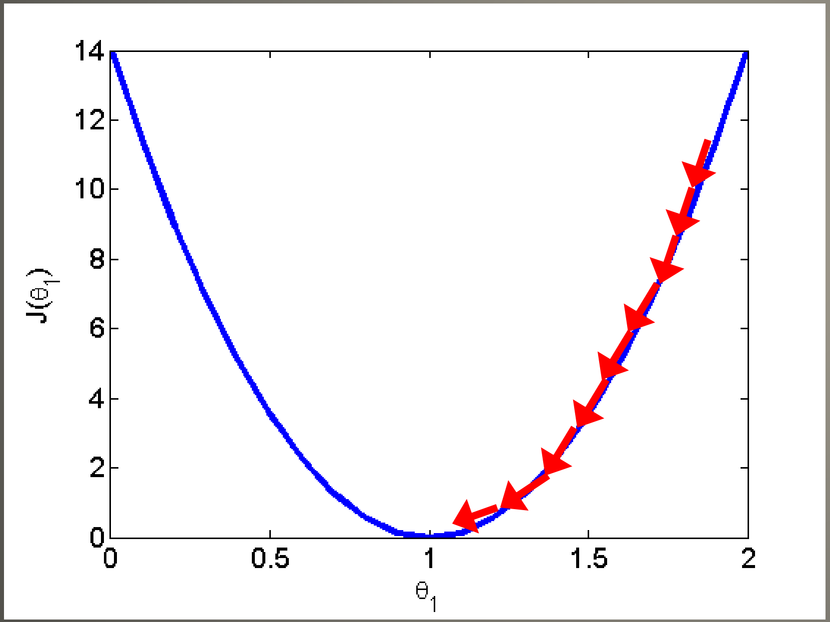
Dereceli azaltma metodu

- Alpha seçme (biraz dikkat edilmeli)
  - Küçük alpha – yavaş yakınsama
  - Büyük alpha –ilerli geri yakınsama



Dereceli azaltma metodu

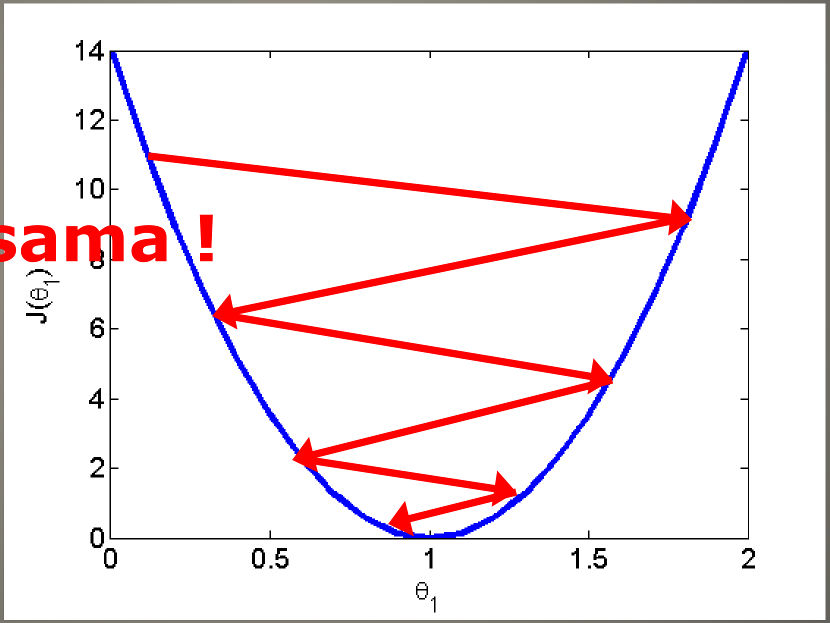
- Alpha seçme
  - Küçük alpha – yavaş yakınsama
  - Büyük alpha –ilerli geri yakınsama



yavaş!

Dereceli azaltma metodu

- Alpha seçme
  - Küçük alpha – yavaş yakınsama
  - Büyük alpha –ilerli geri yakınsama



İleri geri yakınsama !

Dereceli azaltma metodu

- Alpha seçme
  - Küçük alpha – yavaş yakınsama
  - Büyük alpha –ilerli geri yakınsama

- İyi bir alpha seçmek için, birkaç alpha değerini denemek lazım
- Bu değerler için, dereceli azaltma metodunu çalıştırıp performansını incelemek lazım

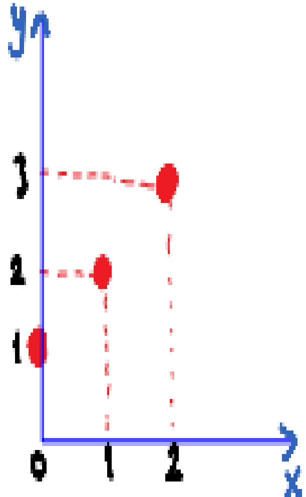
Tipik denenecek alpha değerleri:

alpha=0.001, 0.003, 0.01, 0.03, 0.1, 0.3, 1, 3, 10,...

yavaşyavaş ↑ iyi ilerli geri ilerli geri hkt

Dereceli azaltma metodu

Örnek)

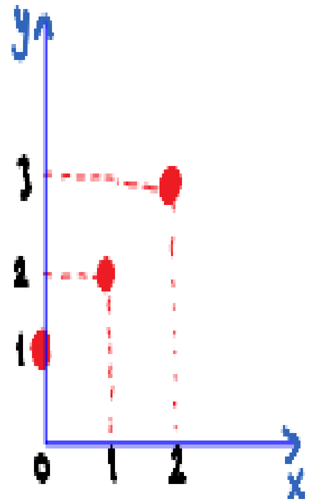


x	y
0	1
1	2
2	3

Doğrusal regresyon yöntemiyle veriden model çıkarın?

El ile Doğrusal Regresyon Çözümü

Örnek)



x	y
0	1
1	2
2	3

Doğrusal regresyon yöntemiyle veriden model çıkarın?

Çözüm) Model polinomu $h_\theta(x) = \theta_0 + \theta_1 x_1$

Başlangıç aşamasında $\theta_0 = 0$ $\theta_1 = 0$

θ kullanılarak x girişlerine karşılık hipotez tahminlerini hesaplayalım:

x	$h_\theta(x)$	y	Hata
0	$0 + 0.0 = 0$	1	-1
1	$0 + 0.1 = 0$	2	-2
2	$0 + 0.2 = 0$	3	-3

El ile Doğrusal Regresyon Çözümü

Çözüm) Model polinomu $h_\theta(x) = \theta_0 + \theta_1 x_1$

Başlangıç aşamasında $\theta_0 = 0$ $\theta_1 = 0$

θ kullanılarak x girişlerine karşılık hipotez tahminlerini hesaplayalım:

x	$h_\theta(x)$	y	Hata
0	$0 + 0.0 = 0$	1	-1
1	$0 + 0.1 = 0$	2	-2
2	$0 + 0.2 = 0$	3	-3

Ortalama maliyeti hesaplayalım: $J(\theta) = \frac{1}{2m} \cdot \sum_{i=1}^{m} \left(h_\theta(x_i) - y_i\right)^2$

$$J(\theta) = \frac{1}{2.3} \sum_{i=1}^{3} \left(h_\theta(x_i) - y_i\right)^2 = \frac{1}{6}\left((-1)^2 + (-2)^2 + (-3)^2\right) = \frac{14}{6}$$

El ile Doğrusal Regresyon Çözümü

Ortalama maliyeti hesaplayalım: $J(\theta)=\frac{1}{2m}\cdot\sum_{i=1}^{m}\left(h_\theta(x_i)-y_i\right)^2$

$$J(\theta)=\frac{1}{2.3}\sum_{i=1}^{3}\left(h_\theta(x_i)-y_i\right)^2=\frac{1}{6}\left((-1)^2+(-2)^2+(-3)^2\right)=\frac{14}{6}$$

Şimdi ortalama hatanın herbir parametreye türevini alarak parametreleri uygun yönde güncelleyelim

$$\theta_j=\theta_j-\alpha.\frac{\partial J(\theta)}{\partial\theta_j}$$

$$\frac{\partial J(\theta)}{\partial\theta_0}=\frac{1}{m}\sum_{i=1}^{m}\overbrace{(\theta_0+\theta_1x_i-y_i)}^{hata^{(i)}}.1=\frac{1}{3}\left[(-1)+(-2)+(-3)\right]=-2$$

$$\frac{\partial J(\theta)}{\partial\theta_1}=\frac{1}{m}\sum_{i=1}^{m}(\theta_0+\theta_1x_i-y_i).x_i=\frac{1}{3}\left[(-1).0+(-2).1+(-3).2\right]=-\frac{8}{3}=-2.6$$

$$\alpha=0.1,\quad \theta_0=0-(0.1).(-2)=0.2$$

$$\theta_1=0-(0.1).(-2.6)=0.26$$

El ile Doğrusal Regresyon Çözümü