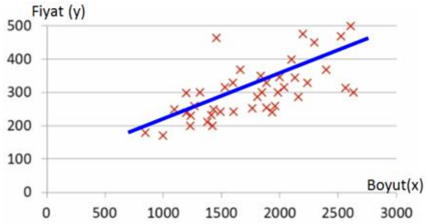
İçindekiler

Regresyon Nedir

Veri kümesini en iyi ifade edebilen eğrinin veriden elde edilmesi işlemine regresyon denir. Örneğin bir emlak firması tarafından satılan evlerin boyut ve fiyat bilgileri Şekil 1(a) gibi olsun. Şekil 1 (b)'de grafiksel olarak gösterilen bu veriyi tek bir doğru ile ifade etmek gerekirse mavi renkli doğru elde edilir. Regresyonun amacı, bu doğruyu oluşturan parametrelerin hesaplanabilmesidir.

Boyut (x)	Fiyat (y)
2104	460
1416	232
1534	315
852	178
...	...



(a) (b)

Şekil 1. Satılan ev bilgileri (a) matris (b) grafik (mavi çizgi: doğrusal hipotez)

Mavi renkli doğru $y = \theta_0 + \theta_1 x$ denklemiyle ifade edilebildiğine göre, regresyon işlemi $\theta = \{\theta_0, \theta_1\}$ parametrelerini veri kullanılarak hesaplanmasıdır.

Normal Eşitlik Yöntemi

Normal eşitlik yöntemi, $X\theta = y$ denklem sistemini linear cebirdeki denklem çözme tekniğini kullanarak θ parametrelerini hesaplar. Bunun için giriş ve çıkış verilerinin matris formuna dönüştürülmesi gerekir (bak Şekil 2).

x_0	x_1	y
1	2104	460
1	1416	232
1	1534	315
1	852	178

$$X = \begin{bmatrix} 1 & 2104 \\ 1 & 1416 \\ 1 & 1534 \\ 1 & 852 \end{bmatrix} \quad y = \begin{bmatrix} 460 \\ 232 \\ 315 \\ 178 \end{bmatrix}$$

(a) Tek değişkenli

x_0	x_1	x_2	x_3	x_4	y
1	2104	5	1	45	460
1	1416	3	2	40	232
1	1534	3	2	30	315
1	852	2	1	36	178

$$X = \begin{bmatrix} 1 & 2104 & 5 & 1 & 45 \\ 1 & 1416 & 3 & 2 & 40 \\ 1 & 1534 & 3 & 2 & 30 \\ 1 & 852 & 2 & 1 & 36 \end{bmatrix} \quad y = \begin{bmatrix} 460 \\ 232 \\ 315 \\ 178 \end{bmatrix}$$

(b) Çok değişkenli

Şekil 2. Tek ve çok değişkenli verinin matrissel forma dönüştürülmesi

Matris formuna dönüşüm tamamlandıktan sonra, θ parametreleri aşağıdaki şekilde hesaplanır.

$$X_{mxn} * \theta_{mx1} = y_{mx1}$$

$$X^T X * \theta = X^T * y$$

$$(X^T X)^{-1} X^T X * \theta = (X^T X)^{-1} X^T * y$$

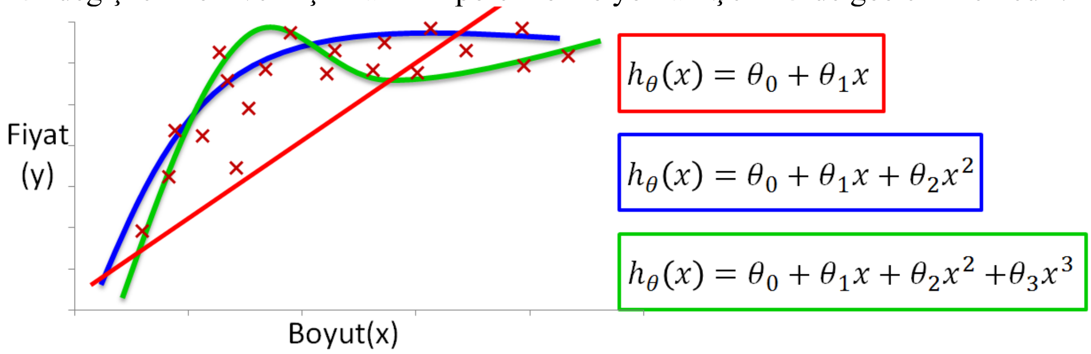
$$\theta = (X^T X)^{-1} X^T y$$

Gradient Azalması (Gradient Descent) Yöntemi

Bir hipotez modelinin veriye olan toplam uzaklığını minimize etmeye dayanır.

Hipotez Fonksiyonu

Tek değişkenli bir veri için farklı hipotez fonksiyonları Şekil 3'de gösterilmektedir.



Şekil 3. Aynı veri üzerinde farklı hipotez fonksiyonları

x örneğinin θ katsayılı hipotez değeri $h_\theta(x)$ ile gösterilir ve girişin eğri üzerindeki değeridir.

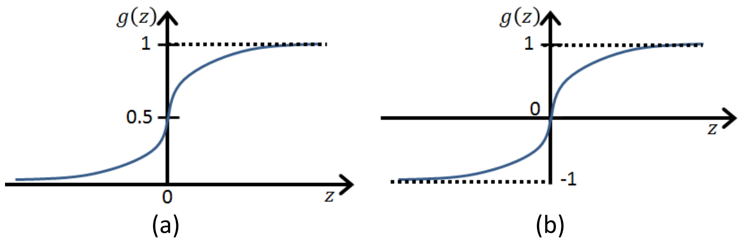
Hipotez fonksiyonunda kullanılacak değişken derecesi veriye göre ayarlanır. Benzer şekilde verinin çok değişkenli olması durumunda hipotez fonksiyonu aşağıdaki gibi güncellenir:

Tek değişkenli	Çok değişkenli
$h_\theta(x) = \theta_0 + \theta_1 x$	$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \cdots + \theta_n x_n$

Tüm durumlar için θ_0 katsayısı $x_0 = 1$ girişine ait olduğu düşünülürse, hipotez değeri tek işlemle elde edilebilir.

$$x = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ . \\ . \\ x_n \end{bmatrix}, \qquad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ . \\ . \\ \theta_n \end{bmatrix} \Rightarrow h_\theta(x) = \theta^T x$$

Doğrusal olmayan verilerin çıkış değerleri [0,1] veya [-1,1] aralığına dönüştürüldükten sonra regresyon işlemine başlanır. Çıkış değerleri üzerindeki yapılan bu dönüşüm, hipotez üzerinde de yapılmak zorundadır. Dönüşümler Şekil 4'de gösterilen sigmoid/lojistik ve hiperbolik tanjant fonksiyonlarıyla gerçekleşir.



Şekil 4. Dönüşüm fonksiyonları (a) Sigmoid/lojistik (b) Hiperbolik tanjant

Sigmoid/Lojistik: $g(z) = \frac{1}{1+e^{-z}}$

Hiperbolik Tanjant: $g(z) = \frac{e^{z}-e^{-z}}{e^{z}+e^{-z}}$

Çıkış değerlerinde yapılan dönüşüm işlemi hipotez fonksiyon değerleri için de yapılır. Buna göre yeni hipotez fonksiyonu aşağıdaki gibidir:

$$h_\theta(x) = \frac{1}{1 + e^{-(\theta_0 + \theta_1 x_1 + \cdots + \theta_n x_n)}}$$

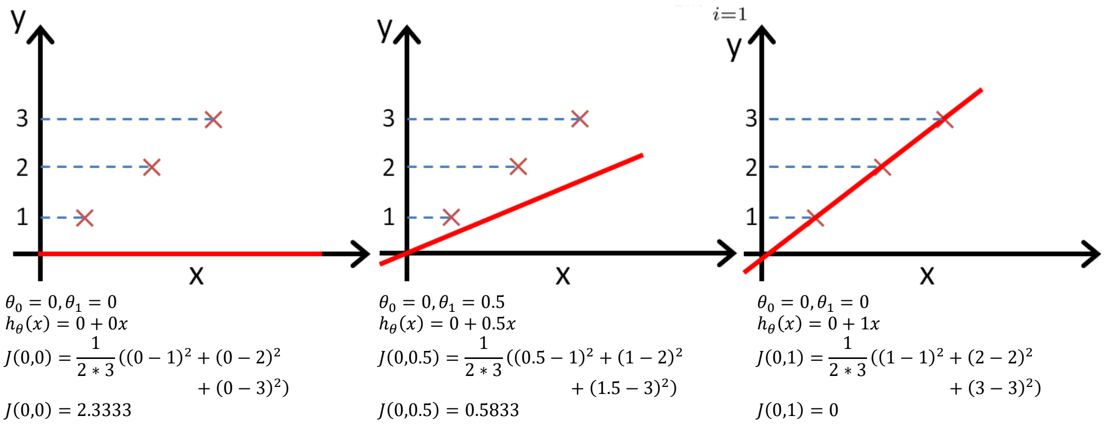
Maliyet Fonksiyonu

Veri kümesindeki örneklerin θ katsayılı hipoteze olan uzaklıklar toplamına maliyet denir ve $J(\theta)$ olarak gösterilir. Regresyon probleminin amacı, minimum $J(\theta)$ için en uygun θ katsayılarını bulmaktır.

$$\min_\theta J(\theta)$$

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

Denklemlerle ifade edilen maliyet fonksiyonunun daha iyi anlaşılabilmesi için Şekil 5'de aynı veri üzerinde üç farklı hipotezden elde edilen maliyet değerleri gösterilmektedir. Maliyeti minimuma indiren θ katsayıları optimaldir.



Şekil 5. Maliyet fonksiyonu örneği

Katsayıların Güncellenmesi

Veriyi en iyi ifade eden hipotez modeli başlangıç aşamasında bilinmediğinden dolayı θ katsayıları sıfır alınır. Katsayılar maliyet fonksiyonunun ilgili katsayıya göre türevi yönünde güncellenmektedir:

Yakınsayana kadar devam et{

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

}

α öğrenme katsayısıdır. Maliyet fonksiyonunun ilgili katsayıya göre türevi:

$$\frac{\partial J(\theta)}{\partial \theta_i} = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)x_j^{(i)}$$

Buna göre katsayılar aşağıdaki gibi güncellenir:

$$\theta_j = \theta_j - \alpha\frac{1}{m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)x_j^{(i)}$$

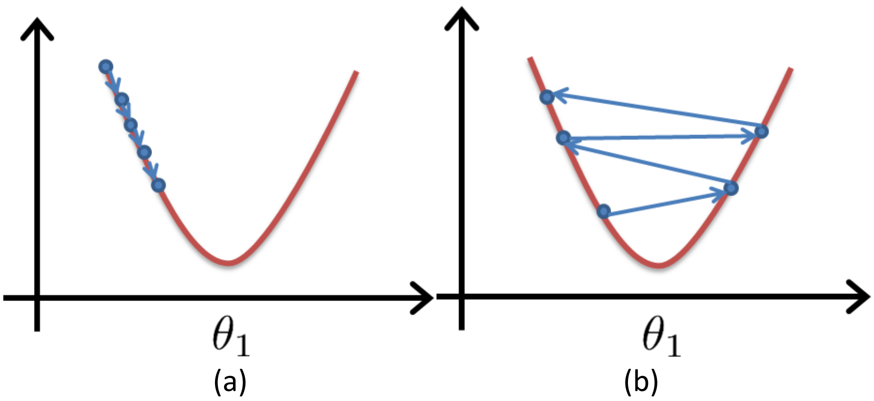
Güncelleme işlemi eş zamanlı yapılmak zorundadır:

Doğru Güncelleme	**Yanlış Güncelleme**
$temp0 := \theta_0 - \alpha\frac{\partial}{\partial\theta_0}J(\theta_0, \theta_1)$	$temp0 := \theta_0 - \alpha\frac{\partial}{\partial\theta_0}J(\theta_0, \theta_1)$
$temp1 := \theta_1 - \alpha\frac{\partial}{\partial\theta_1}J(\theta_0, \theta_1)$	$\theta_0 := temp0$

$\theta_0 \coloneqq temp0$ $\theta_1 \coloneqq temp1$	$temp1 \coloneqq \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$ $\theta_1 \coloneqq temp1$

Öğrenme Katsayısının Etkisi

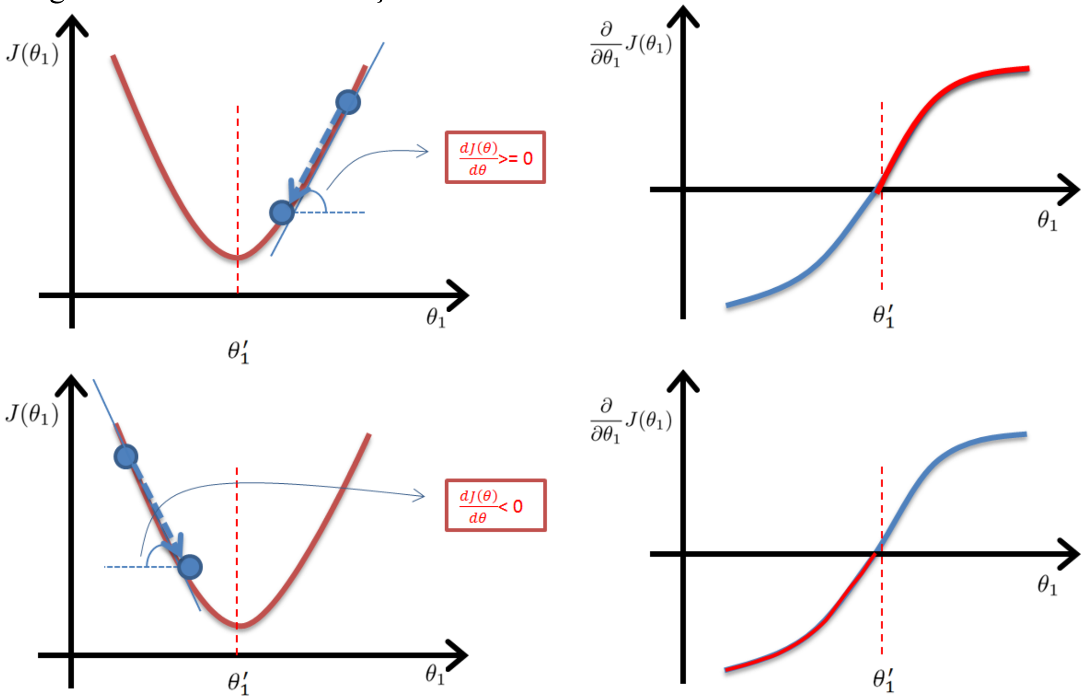
α katsayısıyla doğru orantılı bir şekilde güncelleme gerçekleşir. Şekil 6'da gösterildiği gibi, eğer α çok küçükse, güncelleme adımları küçük olacağından algoritma yavaş olacaktır. Eğer α çok büyükse, algoritmanın hedeflenen noktayı aşabilmesi söz konusudur. Dolayısıyla minimum noktaya yakınsamak yerine uzaklaşabilir. Lokal minimum noktasına yaklaştığımız zaman, algoritma otomatik olarak daha küçük adımlarla ilerleyecektir. Bu nedenle zamanla α değerini güncellemeye gerek kalmaz.



Şekil 6. Öğrenme katsayısının etkisi (a) α çok küçük (b) α çok büyük

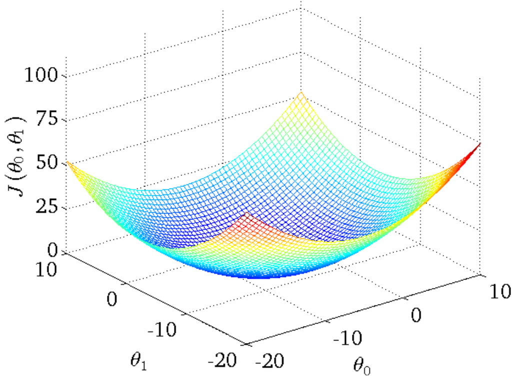
Türevin Etkisi

Türev işlemi θ parametresinin güncellenme yönünü (negatif/pozitif) belirler. Türev sonucu sıfırdan büyükse önündeki negatif işaretten dolayı θ değeri azalır. Aksi halde θ değeri artar. Her iki durum Şekil 3-8'de ifade edilmektedir.



Şekil 7. Türev fonksiyonunun etkisi (ilk satır) türev>0 (ikinci satır) türev<0

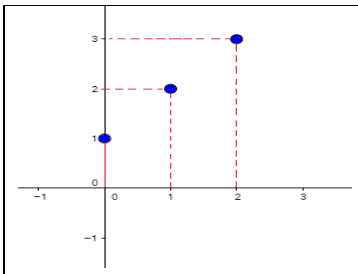
Doğrusal regresyonun hipotez fonksiyonunda iki parametrenin (θ_0 ve θ_1) maliyet fonksiyonu ($J(\theta)$) ile güncelleme grafiği Şekil 7'de gösterilmektedir.



Şekil 7. Maliyet ve hipotez parametreleri arasındaki ilişki

Tek Girişli Doğrusal Regresyon Örneği

Örnek: Doğrusal regresyon yöntemiyle verilen modeli çıkarınız.



x	y
0	1
1	2
2	3

Çözüm:

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

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

θ kullanılarak x girişine 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:

$$J(\theta) = \frac{1}{2n} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)^2$$

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

Şimdi ortalama hatanın her bir parametreye türevini alarak parametreleri uygun yönde güncelleyelim ($\alpha = 0.1$):

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

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

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

$$= -\frac{8}{3} = -2.6$$

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

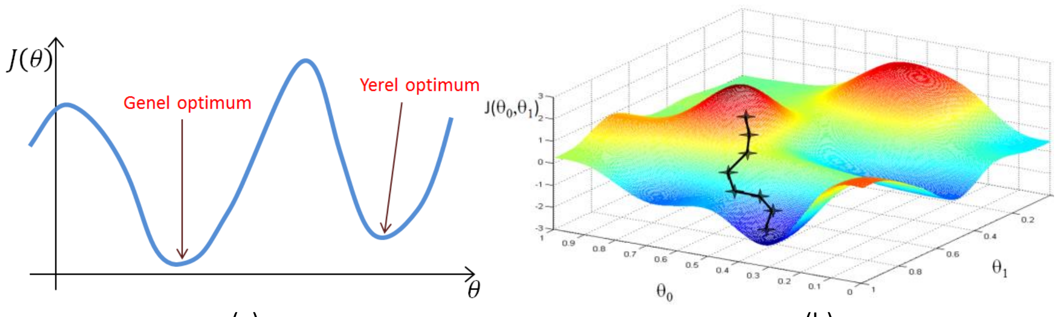
$$\theta_1 = 0 - (0.1).(-2.6) = 0.26$$

Kıyaslama: Normal Eşitlik-Gradient Azalma

Normal eşitlik yönteminin gradient azalma yönteminden üstünlüğü iteratif olmamasıdır. Tek adımda bilinmeyen katsayıları hesaplayabilmektedir. Ancak normal eşitlik yönteminde $X^T X$ matrisinin tersini hesaplama zorluğu vardır. Matris boyutu arttıkça tersini hesaplama işlemi aşırı güçleşir. Ayrıca matris tersinin hesaplanamadığı durumların (satır veya sütun bağımlılığı olduğunda) olduğu bilinmektedir. Bu durumda, X matrisine özellik ekleme/çıkarma yapılır veya regülerizasyon işlemine başvurulur. Özellik sayısının örnek sayısından büyük olması durumunda $(n > m)$ ise uygun θ değerleri hesaplanamamaktadır.

Gradient azalma yönteminin dezavantajları:

- α parametre değerine ihtiyaç duyar.
- İterasyonel çalışır.
- Başlangıç θ değerlerine karşı oldukça hassastır. Buna göre yerel veya global optimum noktasına yakınsayabilir (bak Şekil 8).



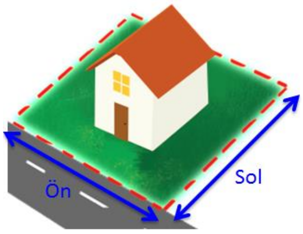
Şekil 8. Dereceli alçaltma yönteminin dezavantajı (a) tek boyut (b) iki boyut

Ön işlemler

Regresyon işlemine başlamadan önce veri üzerinde bir takım ön işlemler yapılmaktadır. Bunlar özelliklerin birleştirilmesi, normalizasyon ve anormal örnek temizlenmesidir.

Polinomsal Özellikler

Çok özellikli verilerde, birden fazla özellik birleştirilerek tek özellik olarak kullanılabilir. Örneğin Şekil 9'daki evin ön ve sol cephe uzunluklarını ayrı kullanmak yerine birleştirilerek tek bir özellik halinde kullanılabilir.



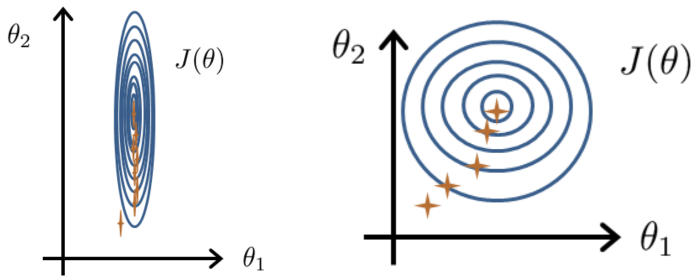
Önce: $h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$, x_1: Ön, x_2: Sol

Sonra: $h_\theta(x) = \theta_0 + \theta_1 x$, x: Ön $* Sol$

Şekil 9.Polinomsal özellikler

Normalizasyon

Eğitim verisindeki her bir örnek farklı büyüklüklerde özellikler içerebilir. Örneğin evin boyutu x_1 ve oda sayısı x_2 özelliklerinin alabileceği değer aralığı çok farklı olabilir. Bu farklılık özelliklerin katsayıları olan θ parametrelerine de yansır. Böylece Şekil 10'da gösterildiği gibi $\theta - J(\theta)$ analizi zorlaştırır.



(a) (b)

Şekil 10. Özellik ölçekleme (a) zor analiz (b) kolay analiz

Ortalama normalizasyonu yöntemi en sık kullanılan yöntemlerden biridir. Buna göre x özelliği aşağıdaki gibi normalleştirilir:

$$x = \frac{x - \mu_x}{\sigma_x}$$

$x - \mu_x$ işlemiyle özellik sütunundaki değerler sıfır merkezine çekilir. σ_x ya bölünerek sütun değerleri yaklaşık olarak [-1,1] aralığında olması sağlanır. Bu işlem verideki tüm sütunlar için ayrı ayrı gerçekleştirilir. Normalleşen verinin $\theta - J(\theta)$ analizi rahatlıkla yapılabilir.