

MAKİNE ÖĞRENMESİ

Çok Değişkenli Lineer Regresyon

1. Çok Değişkenler (Multiple Features)

Çok Değişkenler

	x_1 R&D Spend	x_2 Administration	x_3 Marketing Spend	x_4 State	x_5 Profit
0	165349.20	136897.80	471784.10	New York	192261.83
1	162597.70	151377.59	443898.53	California	191792.06
2	153441.51	101145.55	407934.54	Florida	191050.39
3	144372.41	118671.85	383199.62	New York	182901.99
4	142107.34	91391.77	366168.42	Florida	166187.94

$$n = 5$$

$$x_3 = \begin{bmatrix} \text{New York} \\ \text{California} \\ \text{Florida} \\ \text{New York} \\ \text{Florida} \end{bmatrix}$$

$$x_1^0 = 165349.20$$

Notasyon

n : kolon ~ özellik sayısı

$x^{(i)}$: girdi matrisinin i . örneği

$x_j^{(i)}$: girdi matrisinin j kolonunun i . örneği



Hipotez

$$h_w(x) = w_0 \times x w_1$$

$$h_w(x) = w_0 + w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4$$

$$h_w(x) = 50 + 0.5x_1 + 0.05x_2 + 7x_3 - 6x_4$$

$$h_w(x) = \sum_{i=1}^n w_i x_i + w_0$$

$$h_w(x) = w_0 \vec{x}_0 + w_1 \vec{x}_1 + w_2 \vec{x}_2 + w_3 \vec{x}_3 + \dots + w_n \vec{x}_n$$

$$h_w(x) = w_0 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{m \times 1} + w_1 \begin{bmatrix} | \\ x_1 \\ | \end{bmatrix}_{m \times 1} + w_2 \begin{bmatrix} | \\ x_2 \\ | \end{bmatrix}_{m \times 1} + \dots + w_n \begin{bmatrix} | \\ x_n \\ | \end{bmatrix}_{m \times 1}$$

Matris

$$X = \begin{bmatrix} \vec{x}_0 & \vec{x}_1 & \vec{x}_2 & \dots & \vec{x}_n \end{bmatrix}_{m \times (n+1)}$$

$$\vec{w} = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}_{(n+1) \times 1}$$

Vektör

Skaler

Yaş

1 vektörü

$$x_0^{(i)} = 1$$

$$h_w(x) = X\vec{w}$$



MAKİNE ÖĞRENMESİ

Çok Değişkenli Lineer Regresyon

2. Gradyenik Alçalma (Gradient Descent)

Hipotez

$$h_w(x) = X\vec{w} = \sum_{i=1}^n w_i x_i + w_0$$

Parametreler : $w_0, w_1, w_2, \dots, w_n$

Maliyet Fonksiyonu : $J(w) = \frac{1}{2m} \sum_{i=1}^m (y^{(i)} - h_w(x^{(i)}))^2$

Gradyenik Alçalma Yaklaşımı (Gradient Based Approach)

İterasyon {

$$w_j := w_j - \alpha \frac{\partial J(w_j)}{\partial w_j}$$

}



Gradyenik Alçalma Yaklaşımı(Gradient Based Approach)

$$\frac{\partial J(w_j)}{\partial w_j} = \frac{\partial}{\partial w_j} \frac{1}{2m} \sum_{i=1}^m (y^{(i)} - h_w(x^{(i)}))^2$$

$$\frac{\partial J(w_j)}{\partial w_j} = \frac{\partial}{\partial w_j} \frac{1}{2m} \sum_{i=1}^m (y^{(i)} - (w_0x_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n))^2$$

İterasyon {

$$w_j := w_j + \alpha \frac{1}{m} \sum_{i=1}^m (y^{(i)} - h_w(x^{(i)})) x_j^{(i)}$$

}

Bazı kaynaklarda ağırlık bileşenleri aşağıdaki şekilde formulize edilmektedir.

$$w_j := w_j - \alpha \frac{1}{m} \sum_{i=1}^m (h_w(x^{(i)}) - y^{(i)}) x_j^{(i)}$$



Gradyenik alçalma algoritması

$$\begin{aligned} &\text{İterasyon } \{ \\ &w_j := w_j + \alpha \frac{1}{m} \sum_{i=1}^m (y^{(i)} - h_w(x^{(i)})) x_j^{(i)} \\ &\} \end{aligned}$$

Vektörsel Yaklaşım

$$w_j := w_j + \alpha \frac{1}{m} \sum_{i=1}^m x_j^{(i)} (y^{(i)} - h_w(x^{(i)}))$$

$$\mathbf{w} := \mathbf{w} + \alpha \frac{1}{m} \mathbf{X}^T (\mathbf{y} - \mathbf{X}\mathbf{w})$$



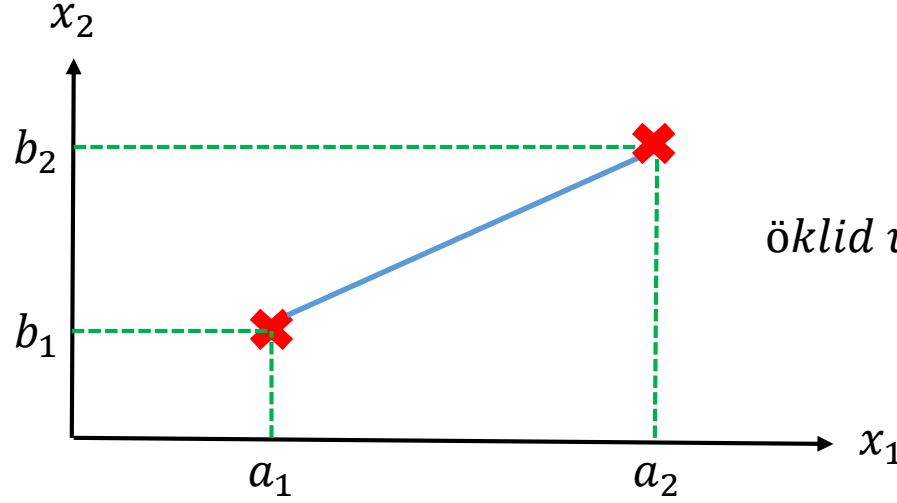
MAKİNE ÖĞRENMESİ

*Çok Değişkenli Lineer
Regresyon*

3. Normalizasyon (Normalization)

Girdinin bütün kolonları aynı ölçekte olursa daha iyi olur.

Age	Salary
44	72000
27	48000
30	54000
38	61000
40	63777.77778
35	58000
38.77777778	52000
48	79000
50	83000
37	67000



$$\text{öklid uzaklığı} = \sqrt{(a_2 - a_1)^2 + (b_2 - b_1)^2}$$

$Age(x_1)$

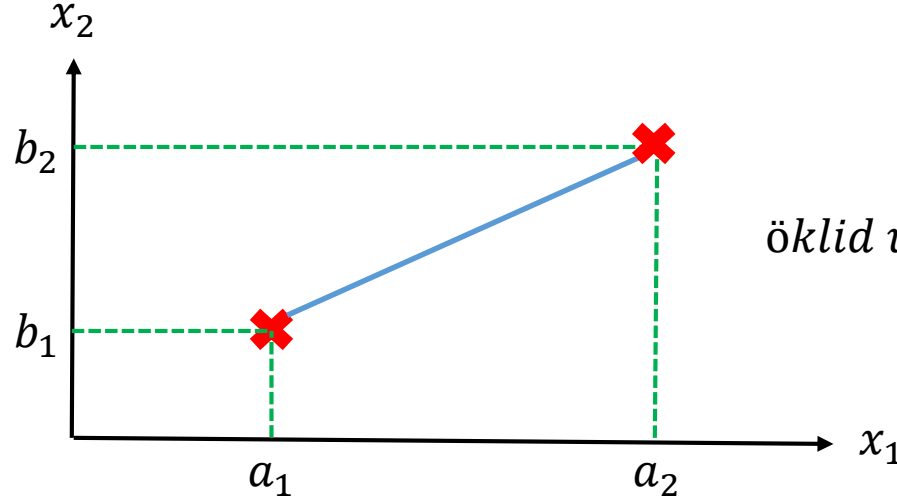


$Salary(x_2)$

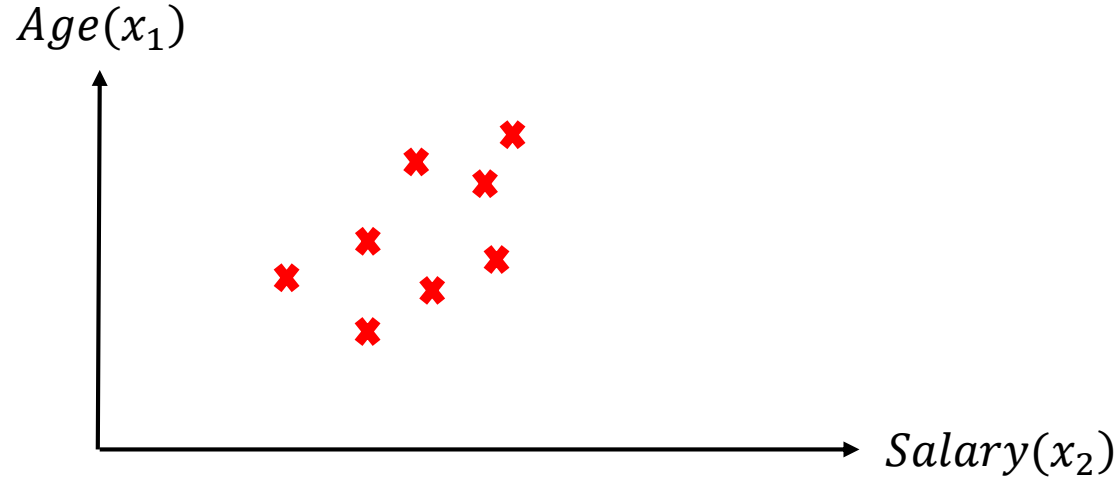


Girdinin bütün kolonları aynı ölçekte olursa daha iyi olur.

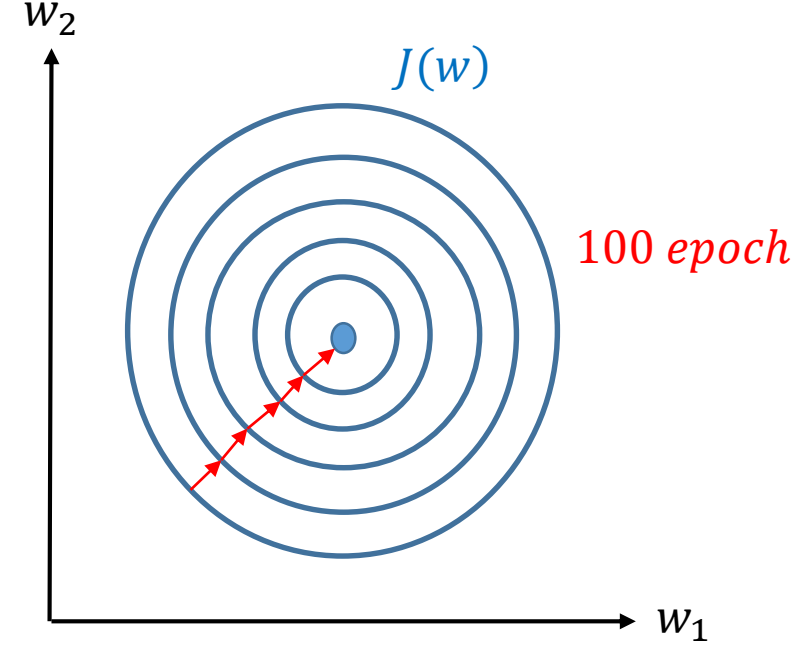
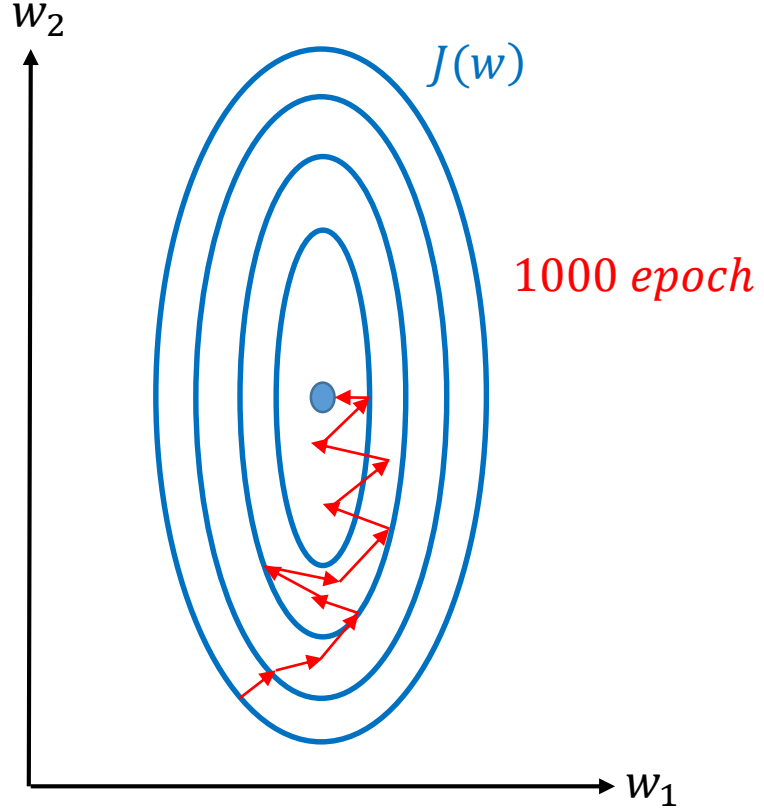
Age	Salary
44	72000
27	48000
30	54000
38	61000
40	63777.77778
35	58000
38.77777778	52000
48	79000
50	83000
37	67000



$$\text{öklid uzaklığı} = \sqrt{(a_2 - a_1)^2 + (b_2 - b_1)^2}$$



Girdinin bütün kolonları aynı ölçekte olursa daha iyi olur.



$-1 \leq x_1 \leq 1$
 $-1 \leq x_2 \leq 1$ \longrightarrow örnek olabilir.

Ortalama Normalizasyonu (Mean Normalization) – StandardScaler

$$x_{mean} = \frac{1}{m} \sum_{i=1}^m x_i$$

$$x_{std} = \sqrt{\frac{1}{m} \sum_{i=1}^m (x_i - x_{mean})^2}$$

$$x_{scaler} = \frac{x - x_{mean}}{x_{std}}$$

Min Max Normalizasyonu – MinMaxScaler

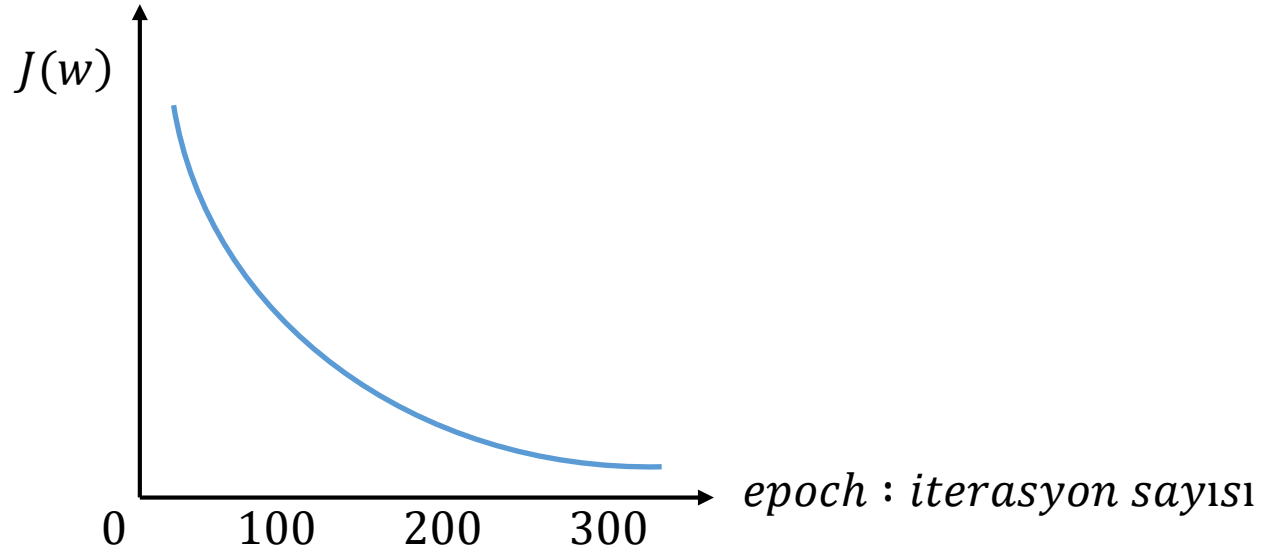
$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}$$



MAKİNE ÖĞRENMESİ

Çok Değişkenli Lineer Regresyon

4. Öğrenme Oranı (Learning Rate)



$J(w)$ her epoch'ta azalmalı

- α çok küçük olursa gradyenik alçalma minimuma çok yavaş bir şekilde yakınsar.
- α çok büyük olursa gradyenik alçalma minimum noktasını aşırabilir.

ideal $\alpha = 0.001, 0.003, 0.01, 0.03, 0.1, 0.3$

MAKİNE ÖĞRENMESİ

Çok Değişkenli Lineer Regresyon

5. Kukla Değişkenler (Dummy Variables)

Kukla Değişkenler (Dummy Variables)

New York = 0

California = 1

Florida = 2

	R&D Spend	Administration	Marketing Spend	State	Profit		
0	165349.20	136897.80	471784.10	New York	192261.83	0	0
1	162597.70	151377.59	443898.53	California	191792.06	0	1
2	153441.51	101145.55	407934.54	Florida	191050.39	1	0
3	144372.41	118671.85	383199.62	New York	182901.99	0	0
4	142107.34	91391.77	366168.42	Florida	166187.94	1	0

New York	California	Florida
0	0	0
0	1	0
1	0	0
1	0	1

MAKİNE ÖĞRENMESİ

Çok Değişkenli Lineer Regresyon

6. Polinomik Regresyon (Polynomial Regression)

Polinomik Regresyon

Ev fiyat tahmini

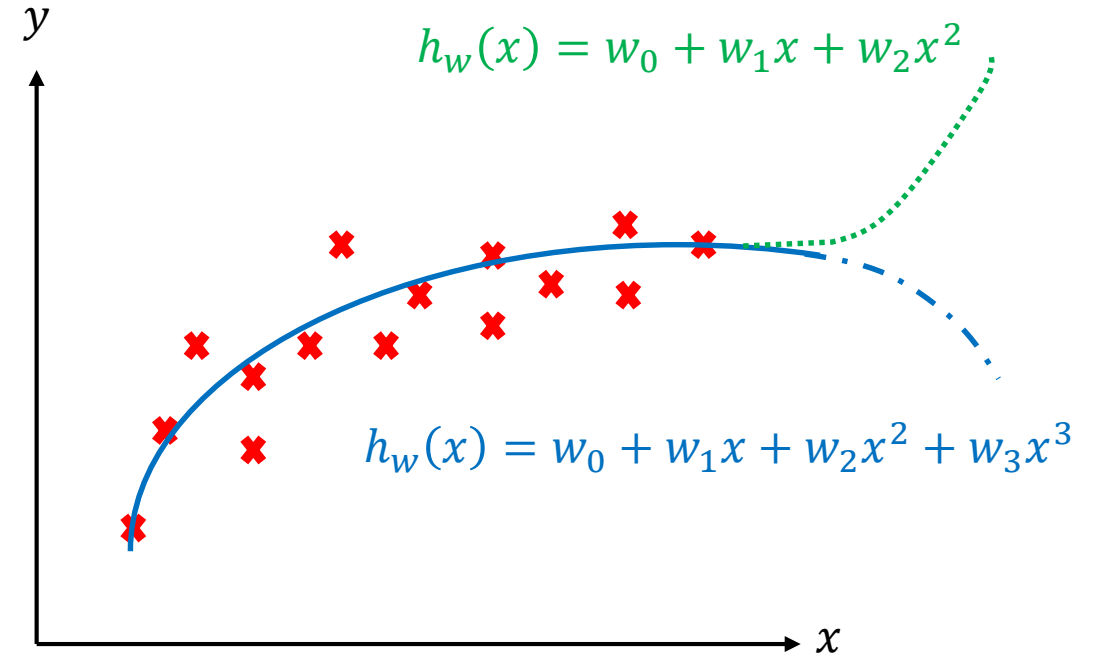
$$h_w(x) = w_0 \vec{x}_0 + w_1 \vec{x}_1 + w_2 \vec{x}_2$$

$$h_w(x) = w_0 + w_1 \cdot \text{ön cephe} + w_2 \cdot \text{derinlik}$$

$$\text{alan} : x = \text{ön cephe} \cdot \text{derinlik}$$

$$h_w(x) = w_0 + w_2 x$$

$$h_w(x) = w_0 + w_2 \cdot \text{alan}$$



MAKİNE ÖĞRENMESİ

*Çok Değişkenli Lineer
Regresyon*

7. Normal Denklem

