Session 7&8

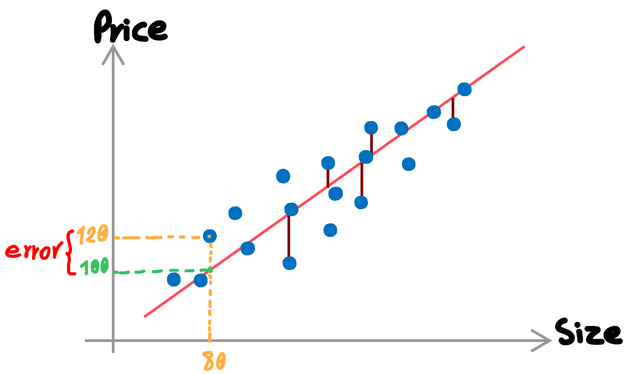
Linear Regression | Part02

NLP | Machine Learning | Zahra Amini

Telegram: @zahraamini_ai & Instagram:@zahraamini_ai &
LinkedIn: @zahraamini-ai

https://zil.ink/zahraamini

Loss Function ⟶ Error ⟶ به اختلاف بین مقدار واقعی و پیش‌بینی مدل خطا می گویند.



Regression Loss

1. $MSE = \frac{1}{2m}\sum_{i=1}^{m}(\hat{y}_i - y_i)^2$?
2. $MAE = \frac{1}{m}\sum_{i=1}^{m}|\hat{y}_i - y_i|$
3. $RMSE = \sqrt{MSE}$
4. $R2\text{-}Score = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$

SS_R (SUM of Squares for Regression)

SS_m (Sum of Squares for Model)

x	y	$\hat{y}$
1	2	1
2	4	2
3	6	3
4	8	4

$$x \xrightarrow{w=1} \bigcirc \longrightarrow \hat{y}$$

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(\hat{y}-y)^2 = \frac{1}{4}\left[\cancel{(1-2)^2}^{1} + \cancel{(2-4)^2}^{4} + \cancel{(3-6)^2}^{9} + \cancel{(4-8)^2}^{16}\right] = \frac{30}{4} = 7.5$$

$$R2 = 1 - \frac{SS_R}{SS_m} = 1 - \frac{30}{20} = -0.5$$? خوبه یا بد؟! -0.5

$$SS_R = (2-1)^2 + (4-2)^2 + (6-3)^2 + (8-4)^2 = 30$$

$$SS_m = (2-5)^2 + (4-5)^2 + (6-5)^2 + (8-5)^2 = 20$$

$$\bar{y} = \frac{2+4+6+8}{4} = 5$$

R2 چه عددی باشد عملکرد مدل خوبه؟

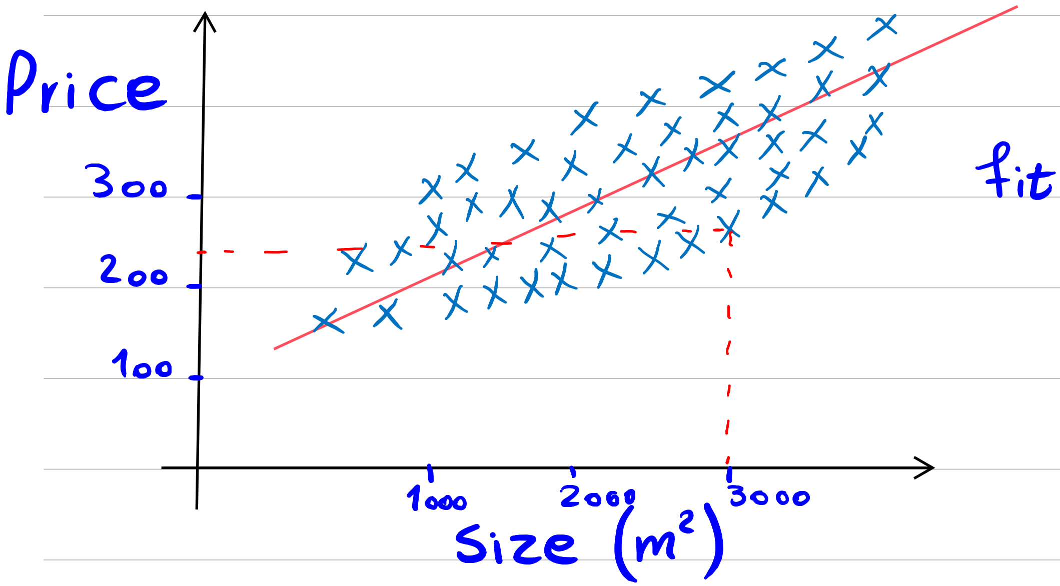
R2 =1: Perfect fit

0<R2 <1: Good fit

R2=0: equivalent to using the mean

R2<0: Worse than using the mean

Linear Regression → Continous → Housing Price Prediction



Notation:

X : Input

y: output

m: # examples   m=4

n: # features   n=1

X → feature   y → target

	Size (m^2)	Price
(1)	2104	460
(2)	1416	232
(3)	1534	315
(4)	852	178
⋮	⋮	⋮

Model: $\underbrace{f_{w,b}(X)}_{\hat{y}} = wX + b$

Cost function: MSE (Mean Square Error)

$$J(w,b) = \frac{1}{2m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)^2$$

Gradient Descent (GD)

Function $\rightarrow J(w,b)$

Goal $\rightarrow \min J(w,b)$

$$W = W - \alpha \frac{\partial}{\partial w} J(w,b) \rightsquigarrow \frac{\partial}{\partial w} J(w,b) = \frac{1}{m} \sum_{i=1}^{m} \left(f_{w,b}(x^{(i)}) - y^{(i)}\right) x^{(i)}$$

$$b = b - \alpha \frac{\partial}{\partial b} J(w,b) \rightsquigarrow \frac{\partial}{\partial b} J(w,b) = \frac{1}{m} \sum_{i=1}^{m} \left(f_{w,b}(x^{(i)}) - y^{(i)}\right)$$

GD Algorithm:

$W = 0$, $b = 0$

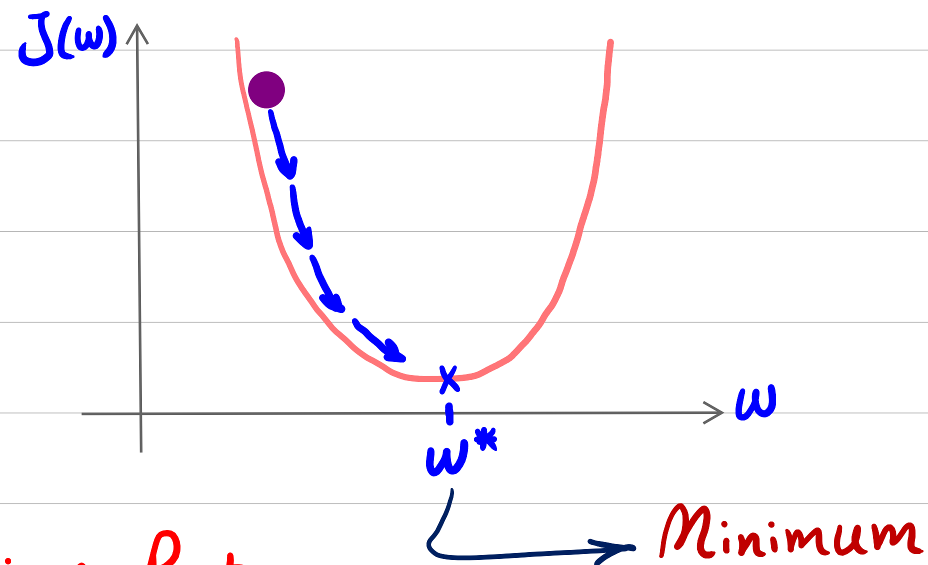
repeat until convergence

{

$w = w - \alpha \frac{\partial}{\partial w} J(w,b)$

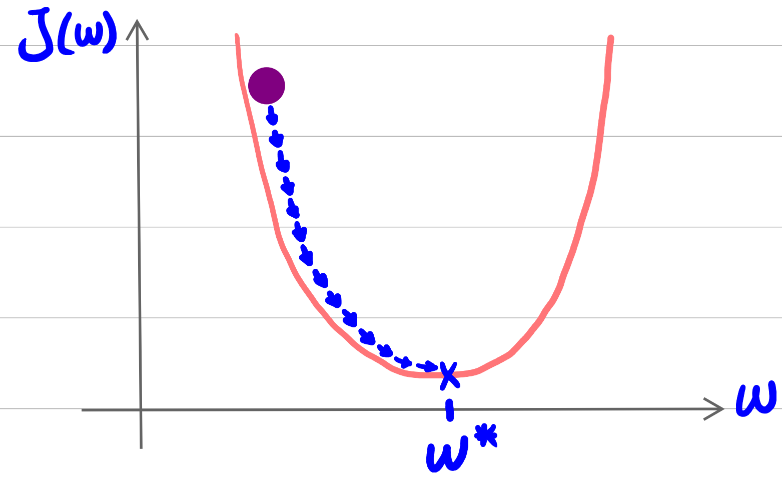
$b = b - \alpha \frac{\partial}{\partial b} J(w,b)$

}

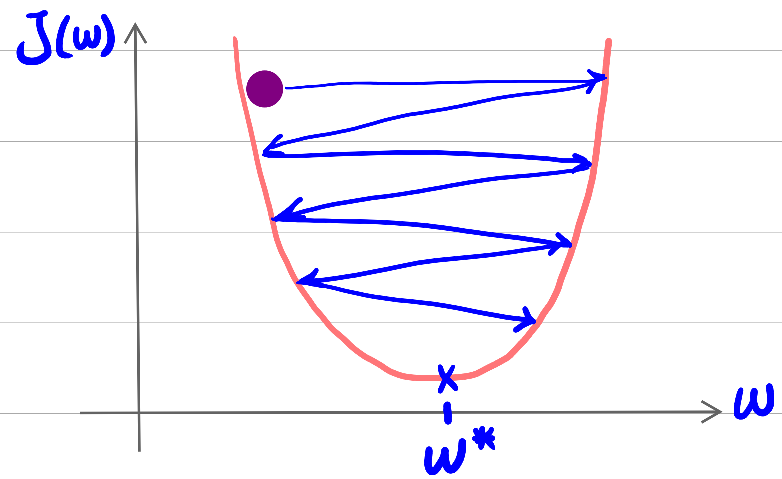


α : Learning Rate

If α is too Small:



If α is too Large:



Multiple Linear Regression:

Size(m^2)	Price
2104	460
1416	232
1534	315
852	178
⋮	⋮

X_1	X_2	X_3	X_4	y
Size(m^2)	#bedrooms	#floors	Age	Price
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
852	2	1	36	178

$x_j = j^{th}$ feature

$n = \#$ features

4

$\vec{x}^{(i)}$ = features of i^{th} training example

$x^{(3)} = [1534 \quad 3 \quad 2 \quad 30]$

$x_2^{(3)} = 3$

One feature	n features
$f_{w,b}(x) = wx + b$	$f_{w,b}(x) = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n + b$

$$\vec{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_n \end{bmatrix}_{n\times 1} \qquad b = \text{number}$$

$$\vec{x} = \begin{bmatrix} x_1^1 & x_1^2 & \cdots & x_1^n \\ x_2^1 & x_2^2 & \cdots & x_2^n \\ & \vdots & & \\ x_m^1 & x_m^2 & \cdots & x_m^n \end{bmatrix}_{m\times n} \qquad X_{m\times n} \cdot W_{n\times 1} = \boxed{}_{m\times 1}$$

$$f_{\vec{w},b}(\vec{x}) = \vec{x} \cdot \vec{w} + b$$

dot product

Multiple Linear Regression:

$n=3$

$\vec{x}=[x_1, x_2, x_3]$ $\quad$ $\vec{w}=[w_1, w_2, w_3]$ $\quad$ b : number

$$f_{\vec{w},b}(\vec{x}) = w_1 x_1 + w_2 x_2 + w_3 x_3 + b$$

```
f = w[0] * x[0] + w[1] * x[1] + w[2] * x[2] + b
```

```
for i in range(3):
    f += x[i] * w[i]
f += b
```

vectorization

$$f_{\vec{w},b}(\vec{x}) = \vec{w}.\vec{x} + b$$

```
f = np.dot(w, x) + b
```

$$x = \begin{array}{c} \begin{matrix} x_0 & x_1 & x_2 \end{matrix} \\ [10, 20, 30] \end{array} \qquad w = \begin{array}{c} \begin{matrix} w_0 & w_1 & w_2 \end{matrix} \\ [1, 2.5, -3.3] \end{array} \qquad b = 4$$

$$(\underbrace{\cancel{1 \times 10}}_{10}) + (\underbrace{\cancel{2.5 \times 20}}_{50}) + (\underbrace{\cancel{-3.3 \times 30}}_{-99}) + \overset{b}{4} = -35$$

GD:

One feature	n features

repeat
{
$w = w - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(f_{w,b}(x^{(i)}) - y^{(i)}\right) x^{(i)}$
$b = b - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(f_{w,b}(x^{(i)}) - y^{(i)}\right)$
}
	repeat
	{
	$w_1 = w_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right) x_1^{(i)}$
	$\vdots$
	$w_n = w_n - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right) x_n^{(i)}$
	$b = b - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right)$
	}

EX:

	x_1 Size (m^2)	x_2 #bedrooms	x_3 #floors	x_4 Age	y Price
(1)	2104	5	1	45	460
(2)	1416	3	2	40	232
(3)	852	2	1	36	178

$X_{3\times4}$ $\quad y_{3\times1}$

$$W = \begin{bmatrix} 0.3 \\ 18.5 \\ -53.3 \\ -26.4 \end{bmatrix}_{4\times1} \qquad b = 785$$

$$f_{w,b}(x) = w_1x_1 + w_2x_2 + \dots + w_nx_n + b$$

l

$$f_{w,b} = w \cdot x + b$$

$$\begin{bmatrix} 2104 & 5 & 1 & 45 \\ 1416 & 3 & 2 & 40 \\ 852 & 2 & 1 & 35 \end{bmatrix}_{3\times4} \begin{bmatrix} 0.3 \\ 18.5 \\ -53.3 \\ -26.4 \end{bmatrix}_{4\times1} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}_{3\times1} \qquad \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} + b$$

$$a_0 = [(2104\times0.3) + (5\times18.5) + (1\times(-53.3)) + (45\times(-26.4))] =$$

$a_1 =$ $\qquad a_2 =$

Feature Size and Parameter Size:

$\hat{y} = w_1 x_1 + w_2 x_2 + \dots + w_n x_n + b$

$$\text{Price\_Predict} = \underset{\text{Size}}{w_1 x_1} + \underset{\#\text{bedroom}}{w_2 x_2} + b$$

x_1: Size (m^2) ⟶ range: 300, 2000 (Large) x_2: #bedroom ⟶ range: 0-5 (Small)

Sample: $x_1 = 2000$ $x_2 = 5$ Price = 500k

① $w_1 = 50$, $w_2 = 0.1$, $b = 50$

$$\text{Price\_Predict} = \underset{w_1}{50} * \underset{x_1}{2000} + \underset{w_2}{0.1} * \underset{x_2}{5} + \underset{b}{50} = 100{,}050.5\text{k}$$

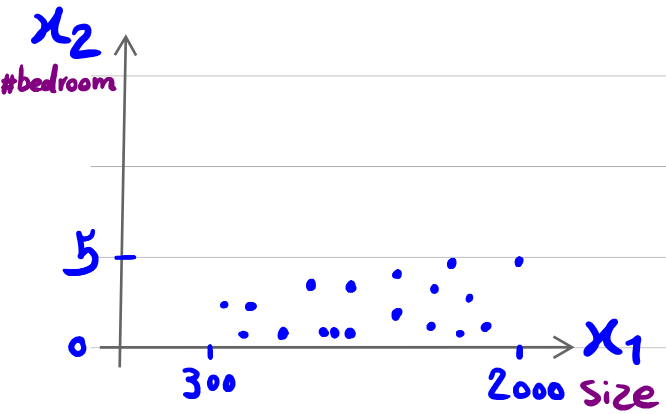
② $w_1 = 0.1$, $w_2 = 50$, $b = 50$

$$\text{Price\_Predict} = 0.1 * 2000 + 50 * 5 + 50 = 500\text{k}$$

نتیجه‌گیری: وقتی که range مقادیر یک ویژگی بزرگ است مدلی خوب یاد می‌گیرد که پارامتر کوچکی برای این ویژگی انتخاب کند.

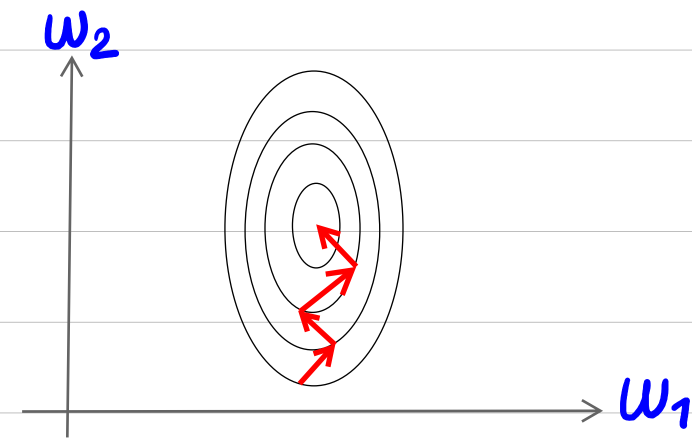
Features:

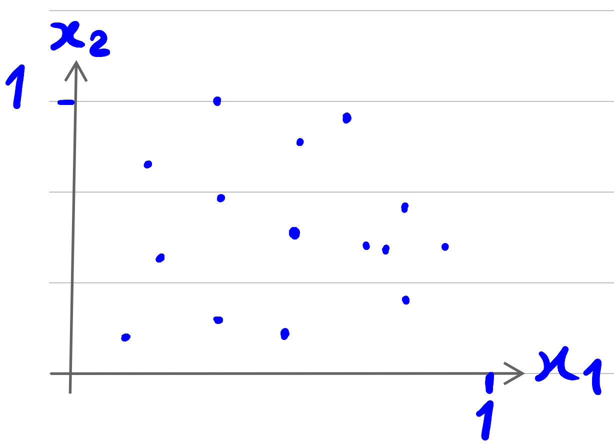
Parameters:

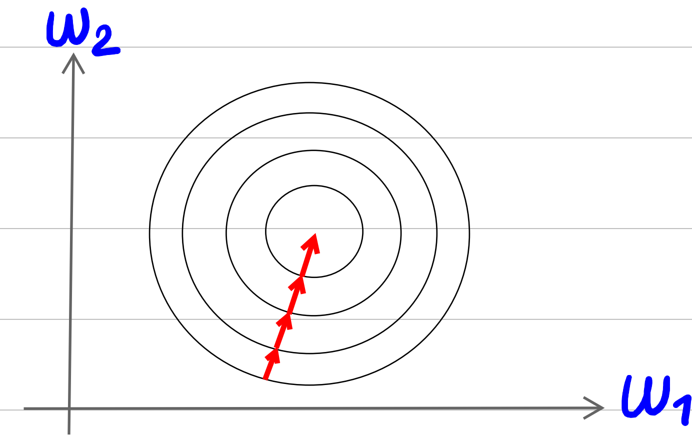


$300 \leqslant x_1 \leqslant 2000$

$0 \leqslant x_2 \leqslant 5$







Feature Scaling:

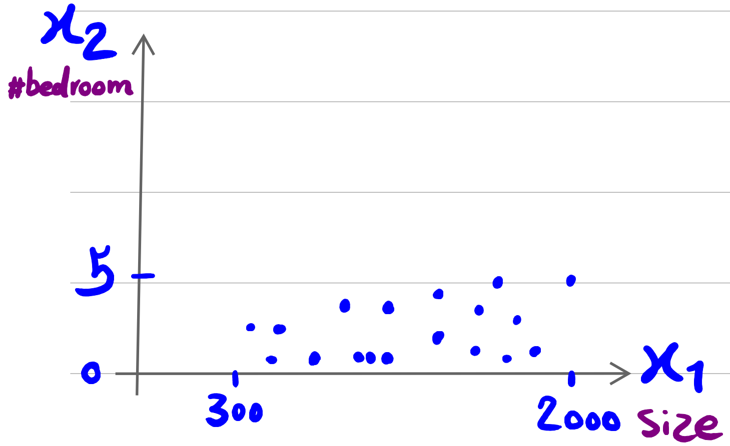
$$x_{1\text{-Scaled}} = \frac{x_1}{\underset{max}{2000}} \qquad 0 \leqslant x_{1\text{-Scaled}} \leqslant 1$$

$$x_{2\text{-Scaled}} = \frac{x_2}{\underset{\to max}{5}} \qquad 0 \leqslant x_{2\text{-Scaled}} \leqslant 1$$

Mean Normalization:

$$\boxed{x = \frac{x-\mu}{max-min}}$$

$\mu_1 = 600$, $\mu_2 = 2.3$



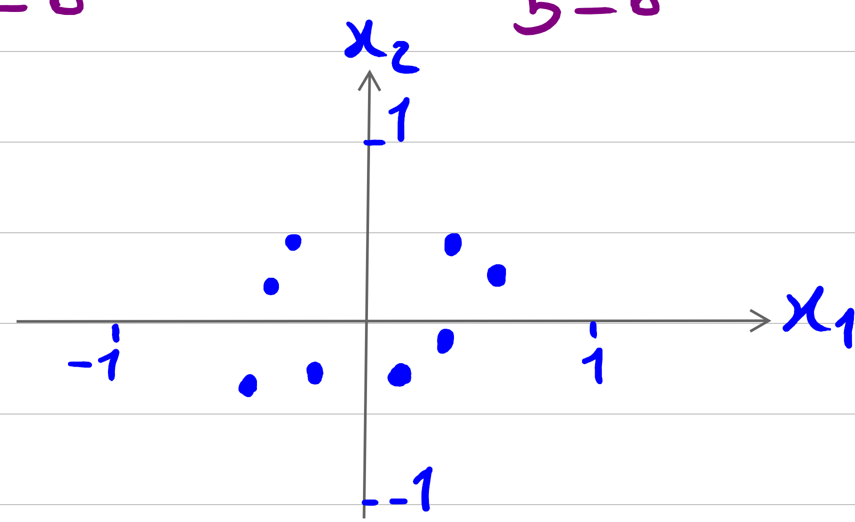
$$300 \leqslant x_1 \leqslant 2000$$

$$\frac{300-\mu_1}{max-min} \leqslant x_1 \leqslant \frac{2000-\mu_1}{max-min}$$

$$\frac{\overset{-300}{\cancel{300-600}}}{\underset{1700}{\cancel{2000-300}}} \leqslant x_1 \leqslant \frac{\overset{1400}{\cancel{2000-600}}}{\underset{1700}{\cancel{2000-300}}} \rightsquigarrow -0.18 \leqslant x_1 \leqslant 0.82$$

$$0 \leqslant x_2 \leqslant 5 \rightsquigarrow \frac{0-\mu_2}{max-min} \leqslant x_2 \leqslant \frac{5-\mu_2}{max-min}$$

$$\frac{0-2.3}{5-0} \leqslant x_2 \leqslant \frac{5-2.3}{5-0} \rightsquigarrow -0.46 \leqslant x_2 \leqslant 0.54$$

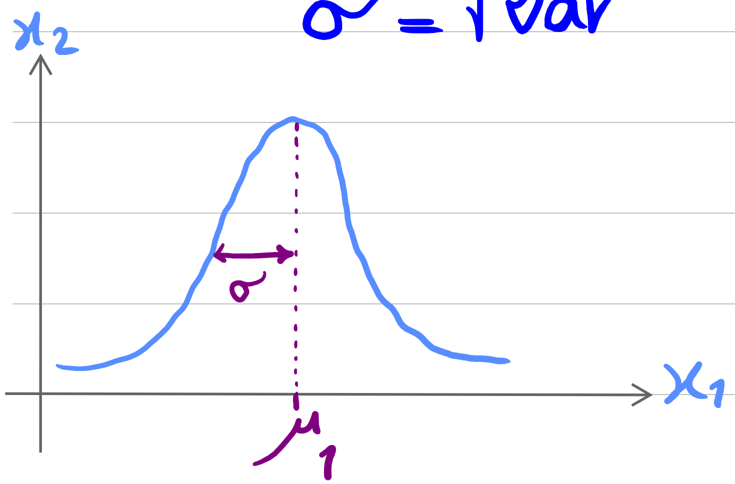


Z-Score Normalization :

$$Var(x) = \frac{1}{n} \sum_{i=1}^{n} (x-\mu)^2$$

standard deviation یا σ یا انحراف معیار

$$\sigma = \sqrt{Var}$$



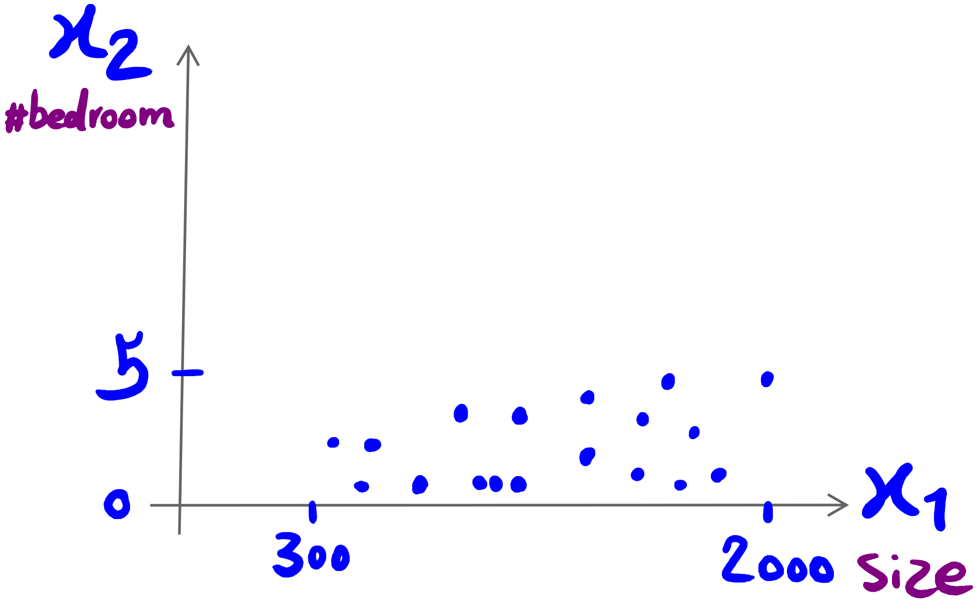
$$\boxed{x = \frac{x-\mu}{\sigma}}$$

StandardScaler()

Sklearn در

Z-Score Normalization:

$$x = \frac{x - \mu}{\sigma}$$



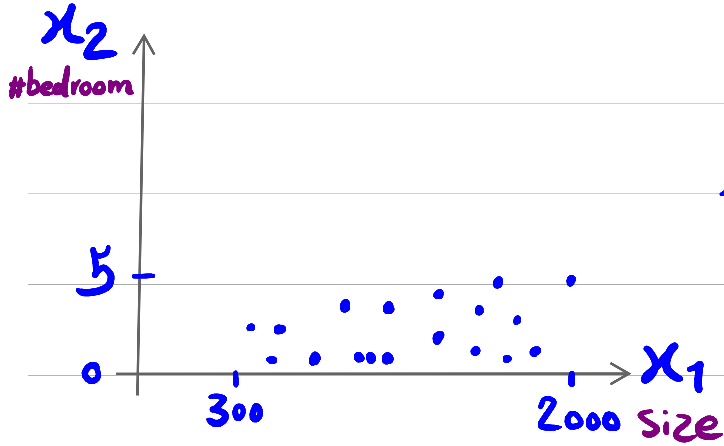
$\mu_1 = 600$ $\sigma_1 = 450$

$\mu_2 = 2.3$ $\sigma_2 = 1.4$

$$a \leqslant x_1 \leqslant b$$

$$c \leqslant x_2 \leqslant d$$

$$a, b, c, d = ?$$



$\mu_1 = 600 \qquad \sigma_1 = 450$

$\mu_2 = 2.3 \qquad \sigma_2 = 1.4$

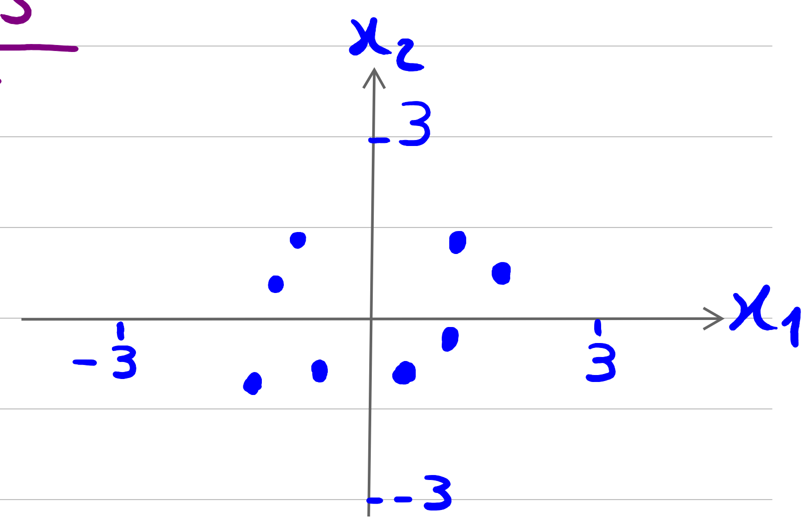
$$x = \frac{x - \mu}{\sigma}$$

$$300 \leqslant x_1 \leqslant 2000$$

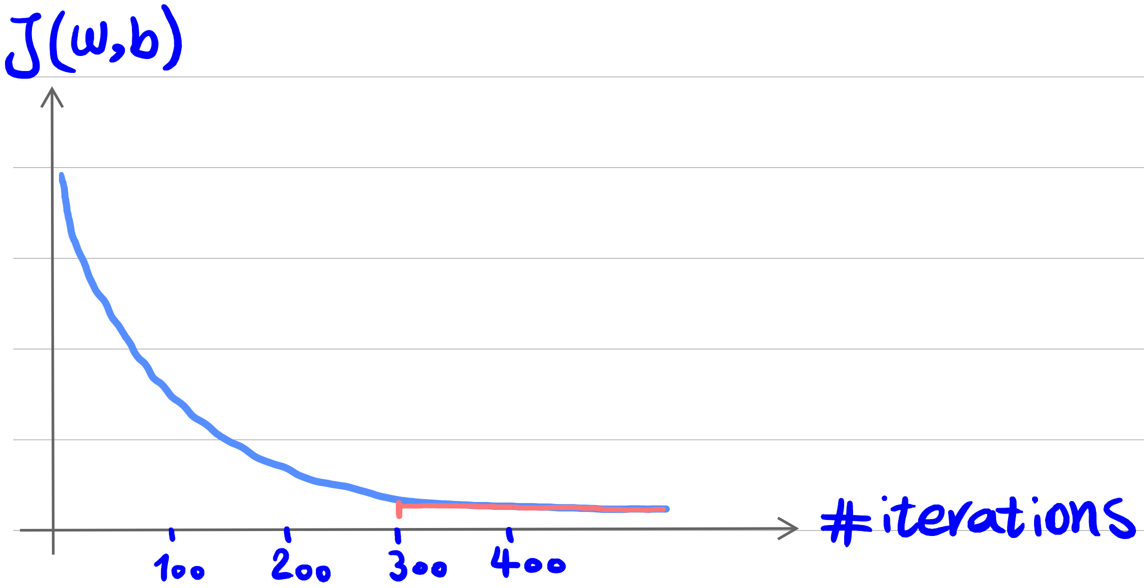
$$x_1 = \frac{x_1 - \mu_1}{\sigma_1} \rightsquigarrow \overset{-0.67}{\frac{300-600}{450}} \leqslant x_1 \leqslant \overset{3.1}{\frac{2000-600}{450}}$$

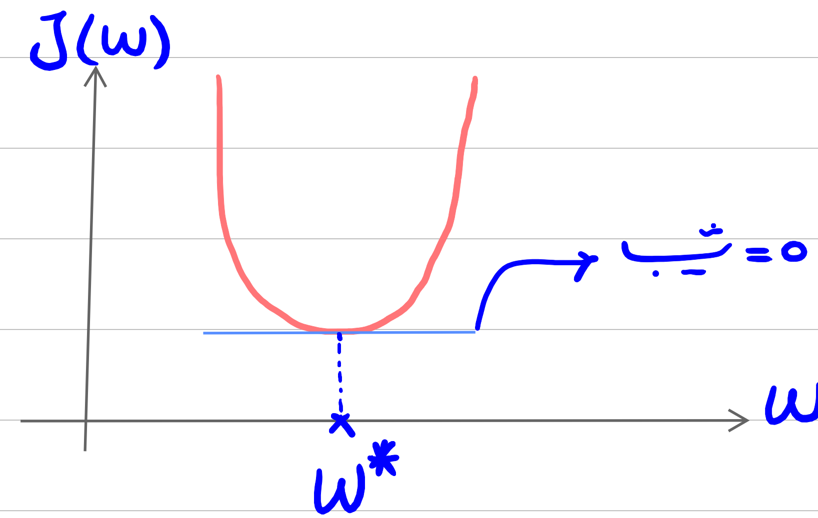
$$0 \leqslant x_2 \leqslant 5$$

$$x_2 = \frac{x_2 - \mu_2}{\sigma_2} \rightsquigarrow \overset{-1.6}{\frac{0-2.3}{1.4}} \leqslant x_2 \leqslant \overset{1.9}{\frac{5-2.3}{1.4}}$$



؟: چه زمانی GD همگرا می‌شود؟





Goal: Converge to Global minimum

(1) #iterations

1. تعیین مقدار Iteration

2. تعیین مقدار J(w,b) ← تست همگرایی

```
for i in range(1000):
  {
    w = w - α ∂/∂w J(w,b)

    b = b - α ∂/∂b J(w,b)
  }
```

$$w = w - \alpha \frac{\partial}{\partial w} J(w,b)$$

$$b = b - \alpha \frac{\partial}{\partial b} J(w,b)$$

② ε

$J(w,b)$

ε

While decreas > thr:

{

$$w = w - \alpha \frac{\partial}{\partial w} J(w,b)$$

$$b = b - \alpha \frac{\partial}{\partial b} J(w,b)$$

}

Regression

→ $MAE = \frac{1}{m} \sum_{i=1}^{m} | \hat{y}_i - y_i |$

Mean Absolute Error

→ $MSE = \frac{1}{m} \sum_{i=1}^{m} (\hat{y}_i - y_i)^2$

Mean Squared Error

→ $RMSE = \sqrt{MSE}$

Root Mean Square Error