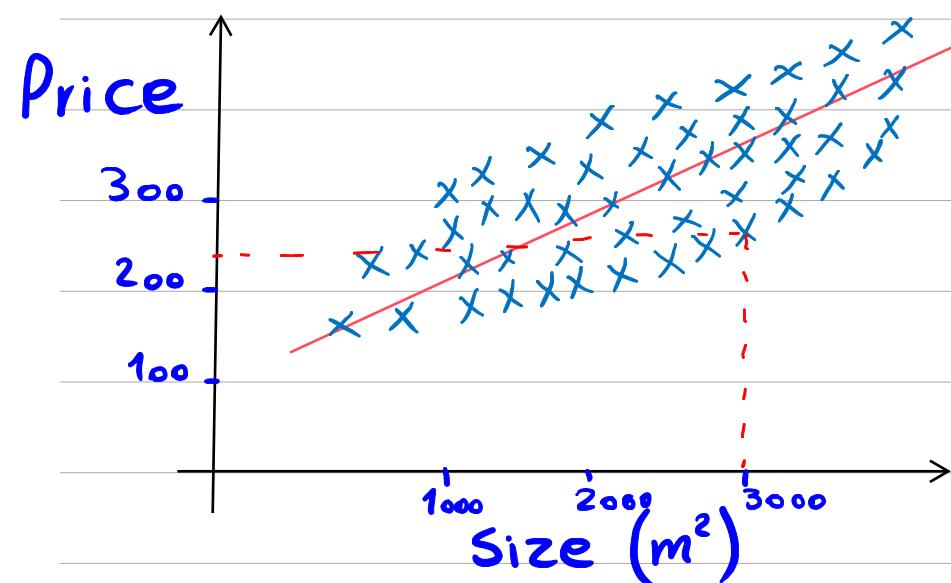


Session 10

Multiple Linear Regression with GD

Machine Learning | Zahra Amini

Linear Regression \rightarrow Continuous \rightarrow Housing Price Prediction



Notation:

fit X : Input

y : output

m : # examples $m=4$

X → feature

y → target

n : # features $n=1$

	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)

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

Gradient Descent (GD)

function $\rightarrow J(w, b)$

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

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

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

GD Algorithm:

$$w=0, b=0$$

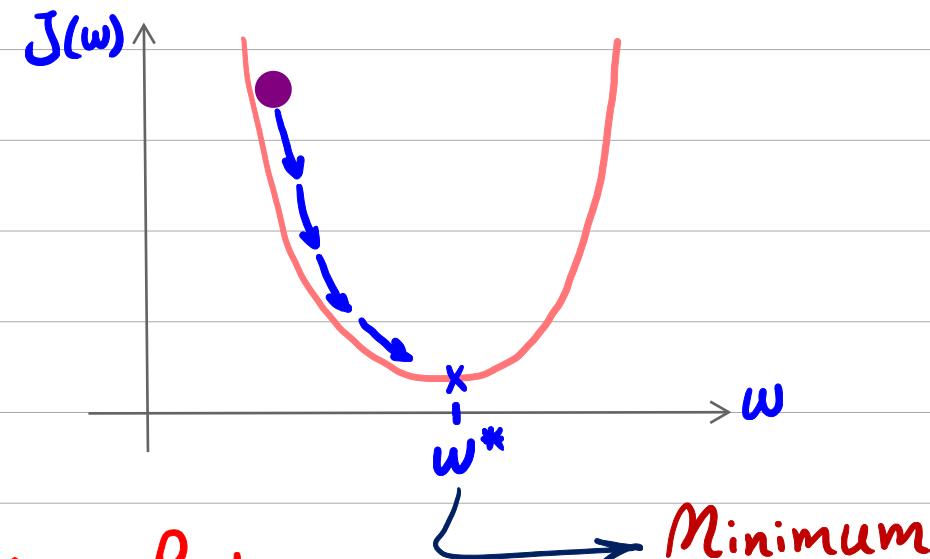
repeat until convergence

{

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

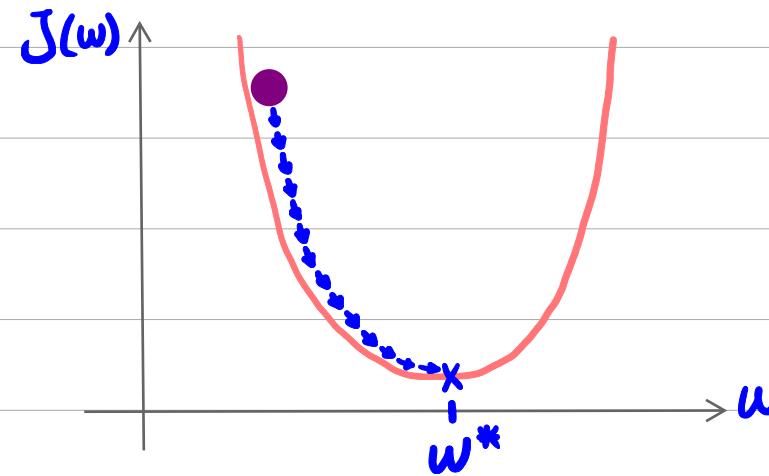
$$b = b - \alpha \frac{\partial}{\partial b} J(w, b)$$

α : Learning Rate

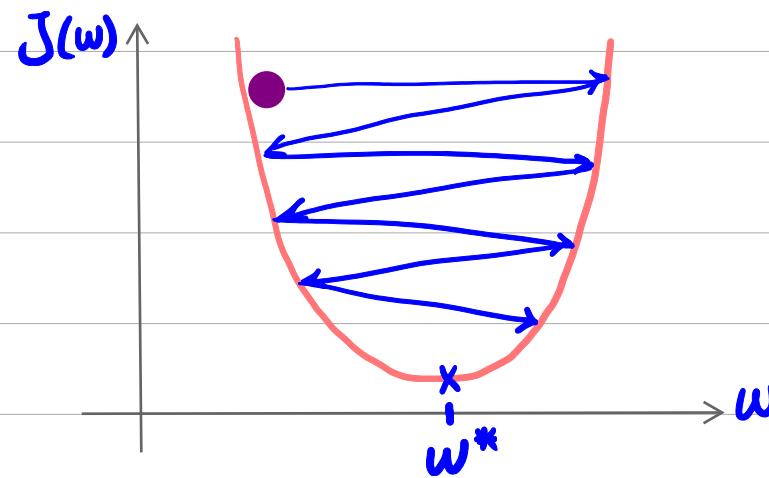


{}

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

$x_j = j^{\text{th}}$ feature $n = \# \text{features}$

4

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

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

$$x_2^{(3)} = 3$$

One feature

$$f_{w,b}(x) = Wx + b$$

n features

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

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

b = number

$$\vec{x} = \begin{bmatrix} x_1^1 & x_1^2 & \dots & x_1^n \\ x_2^1 & x_2^2 & \dots & x_2^n \\ \vdots & & & \\ x_m^1 & x_m^2 & \dots & x_m^n \end{bmatrix}_{m \times n}$$

$$X_{m \times n} \cdot W_{n \times 1} = \boxed{\quad}_{m \times 1}$$

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

dot product

Multiple Linear Regression:

$n=3$

$$\vec{x} = [x_1, x_2, x_3]$$

$$\vec{w} = [w_1, w_2, w_3]$$

b : number

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

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

$$f_+ = x[i] * w[i]$$

} vectorization

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

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

$$x = \begin{bmatrix} x_0 & x_1 & x_2 \\ 10 & 20 & 30 \end{bmatrix}$$

$$w = \begin{bmatrix} w_0 & w_1 & w_2 \\ 1 & 2.5 & -3.3 \end{bmatrix}$$

$$b = 4$$

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

10 50 -99

for i in range(3):

$$f_+ = b$$

GD:

One feature

repeat

{

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

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

}

n features

repeat

{

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

:

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

$$b = b - \alpha \frac{1}{m} \sum_{i=1}^m (f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)})$$

}

EX:

X_1 X_2 X_3 X_4 y

	Size(m ²)	#bedrooms	#floors	Age	Price
(1)	2104	5	1	45	460
(2)	1416	3	2	40	232
(3)	852	2	1	36	178

X

$$W = \begin{bmatrix} 0.3 \\ 18.5 \\ -53.3 \\ -26.4 \end{bmatrix}$$

y

$$b = 785$$

4×1

$$f_{w,b}(x) = w_1 x_1 + w_2 x_2 + \dots + w_n x_n + b \quad \underline{l} \quad f_{w,b} = W \cdot X + b$$

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

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} + b$$

$$a_0 = [(2104 \times 0.3) + (5 \times 18.5) + (1 \times (-53.3)) + (45 \times (-26.4))] =$$

$$a_1 =$$

$$a_2 =$$

Feature Size and Parameter Size:

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

$$\text{Price_Predict} = w_1 \underbrace{x_1}_{\text{Size}} + w_2 \underbrace{x_2}_{\# \text{bedroom}} + b$$

x_1 : Size (m^2) \rightarrow Range: 300, 2000
Large

x_2 : #bedroom \rightarrow Range: 0-5
Small

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

① $w_1 = 50$, $w_2 = 0.1$, $b = 50$

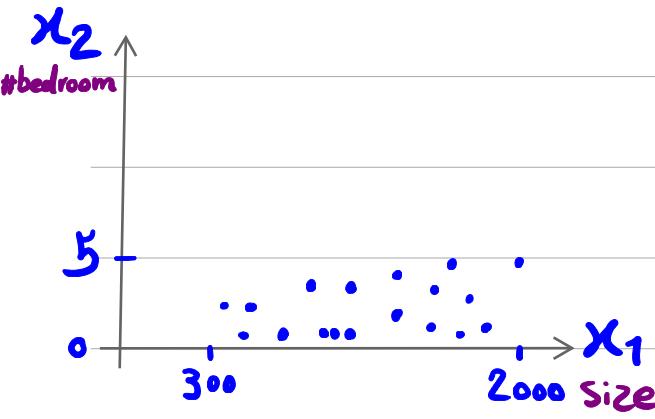
$$\begin{array}{cccccc} w_1 & x_1 & w_2 & x_2 & b \\ \text{Price_Predict} = 50 * 2000 + 0.1 * 5 + 50 & = 100,050.5 \text{ K} \end{array}$$

② $w_1 = 0.1$, $w_2 = 50$, $b = 50$

$$\text{Price_Predict} = 0.1 * 2000 + 50 * 5 + 50 = 500 \text{ K}$$

نکه لیبری: وقتی نہ range مقادیر یک دیتر کی بزرگ است مدلی خوب نداری یا نیز که پارامتر کو جگلی برای

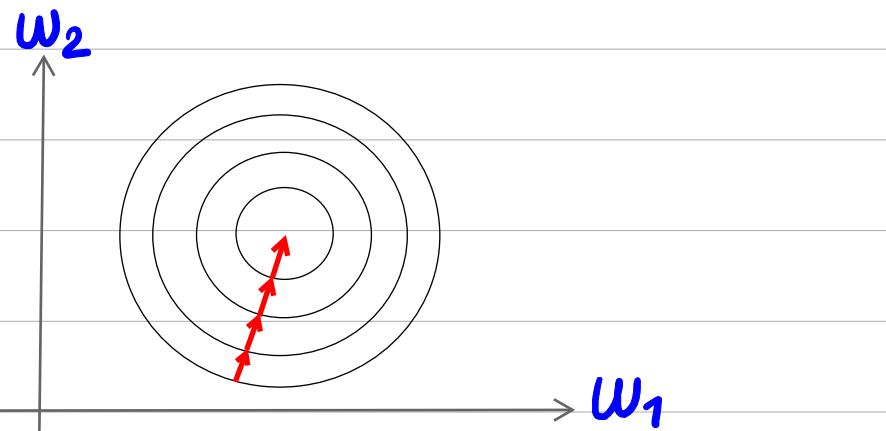
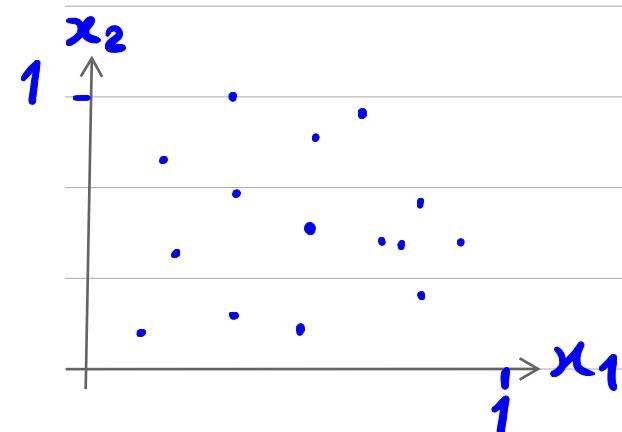
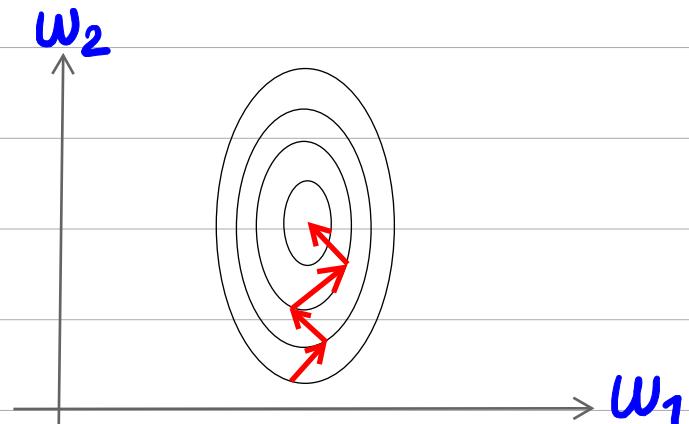
Features:



$$300 \leq x_1 \leq 2000$$

$$0 \leq x_2 \leq 5$$

Parameters: این دیتر کی انتخاب کنہ.



Feature Scaling:

$$x_{1\text{-Scaled}} = \frac{x_1}{\max}$$

$$0 \leq x_{1\text{-Scaled}} \leq 1$$

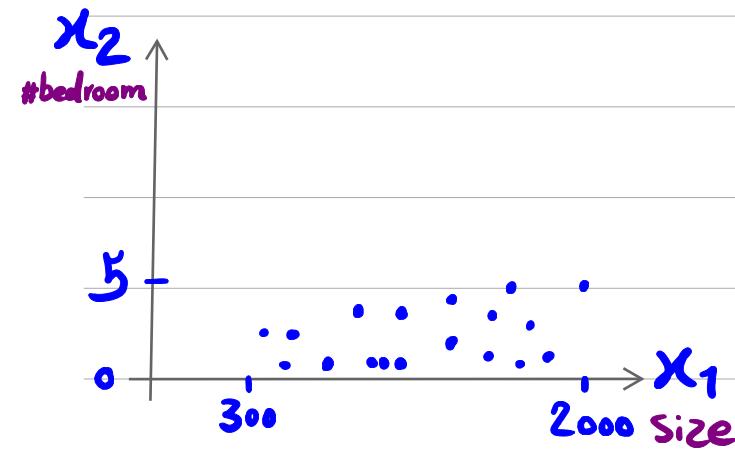
$$x_{2\text{-Scaled}} = \frac{x_2}{\max}$$

$$0 \leq x_{2\text{-Scaled}} \leq 1$$

Mean Normalization:

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

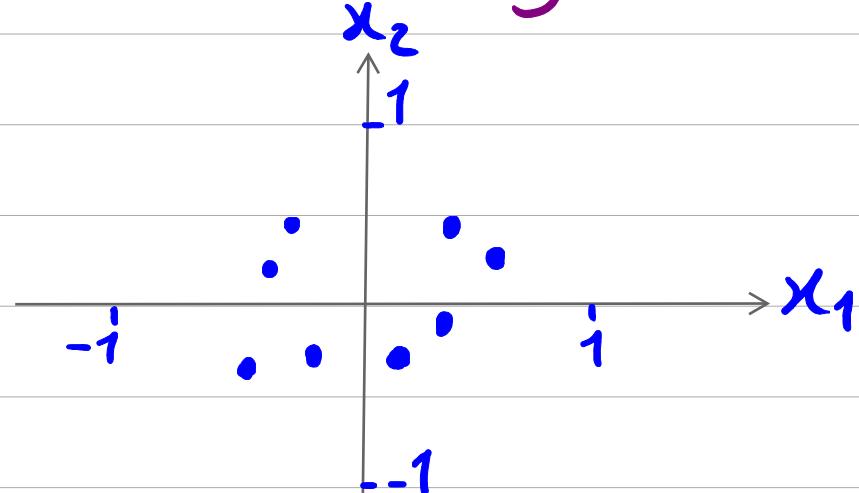
$$\mu_1 = 600, \mu_2 = 2.3$$



$$\frac{300 - \mu_1}{\max - \min} \leq x_1 \leq \frac{2000 - \mu_1}{\max - \min}$$
$$\frac{-300}{2000 - 300} \leq x_1 \leq \frac{1400}{2000 - 300}$$
$$\frac{300 - 600}{2000 - 300} \leq x_1 \leq \frac{2000 - 600}{2000 - 300}$$
$$\frac{-300}{1700} \leq x_1 \leq \frac{1400}{1700}$$
$$\rightarrow -0.18 \leq x_1 \leq 0.82$$

$$0 \leq x_2 \leq 5 \rightarrow \frac{0 - \mu_2}{\max - \min} \leq x_2 \leq \frac{5 - \mu_2}{\max - \min}$$

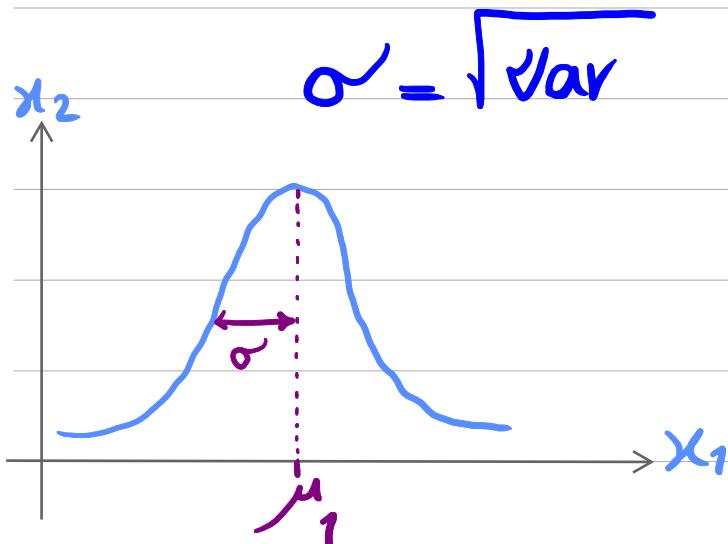
$$\frac{0 - 2.3}{5 - 0} \leq x_2 \leq \frac{5 - 2.3}{5 - 0}$$
$$\rightarrow -0.46 \leq x_2 \leq 0.54$$



Z-Score Normalization:

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

Standard deviation σ اکراف معیار یا

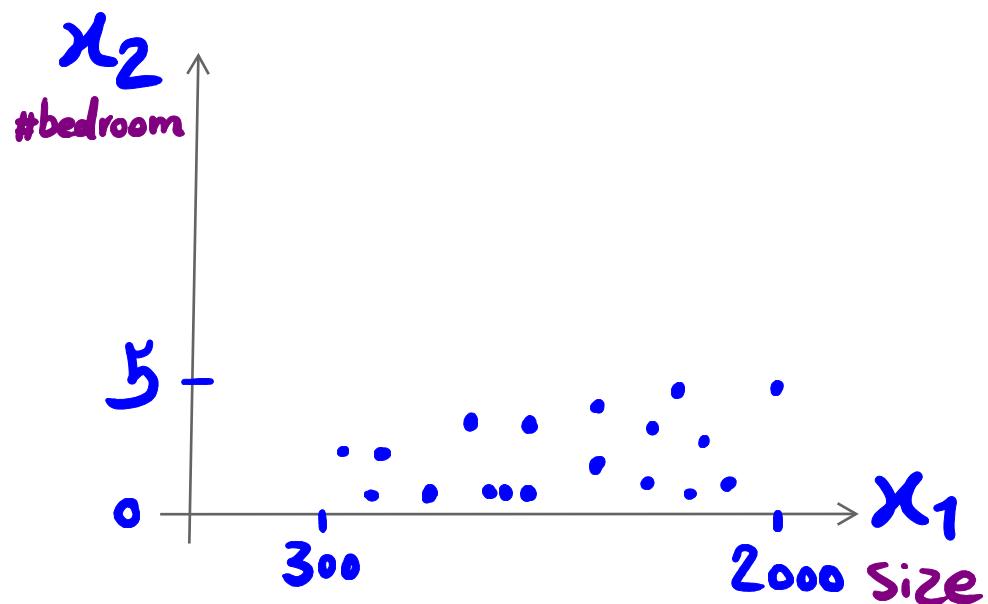


$$x = \frac{x - \mu}{\sigma}$$

StandardScaler()

Sklearn ↗

Z-Score Normalization: $x = \frac{x - \mu}{\sigma}$



$$\mu_1 = 600$$

$$\mu_2 = 2.3$$

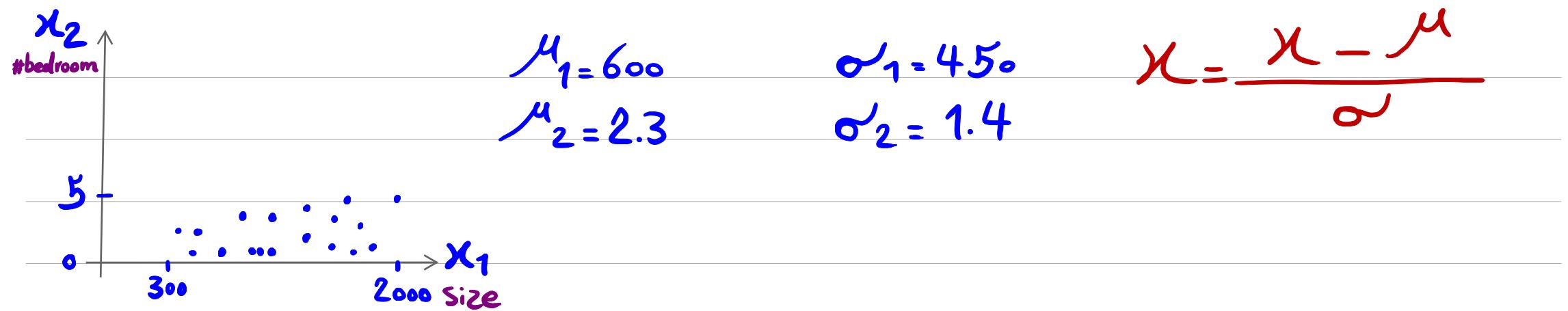
$$\sigma_1 = 450$$

$$\sigma_2 = 1.4$$

$$a \leq x_1 \leq b$$

$$c \leq x_2 \leq d$$

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



$$300 \leq x_1 \leq 2000$$

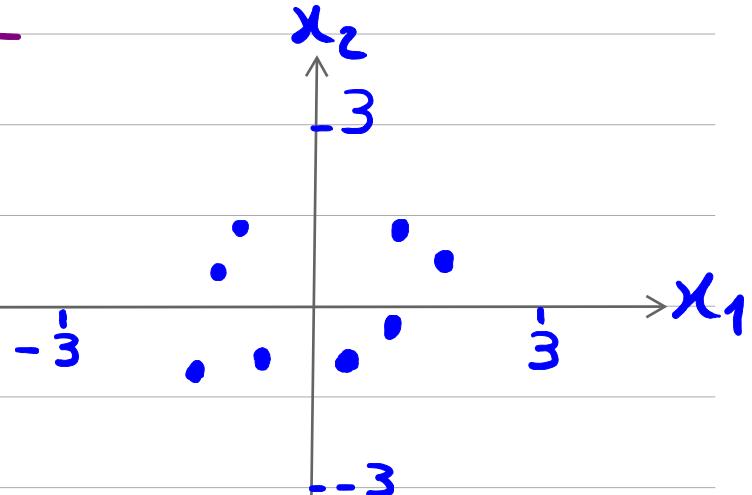
$$\chi_1 = \frac{x_1 - \mu_1}{\sigma_1} \rightsquigarrow \frac{300 - 600}{450} \leq \chi_1 \leq \frac{2000 - 600}{450}$$

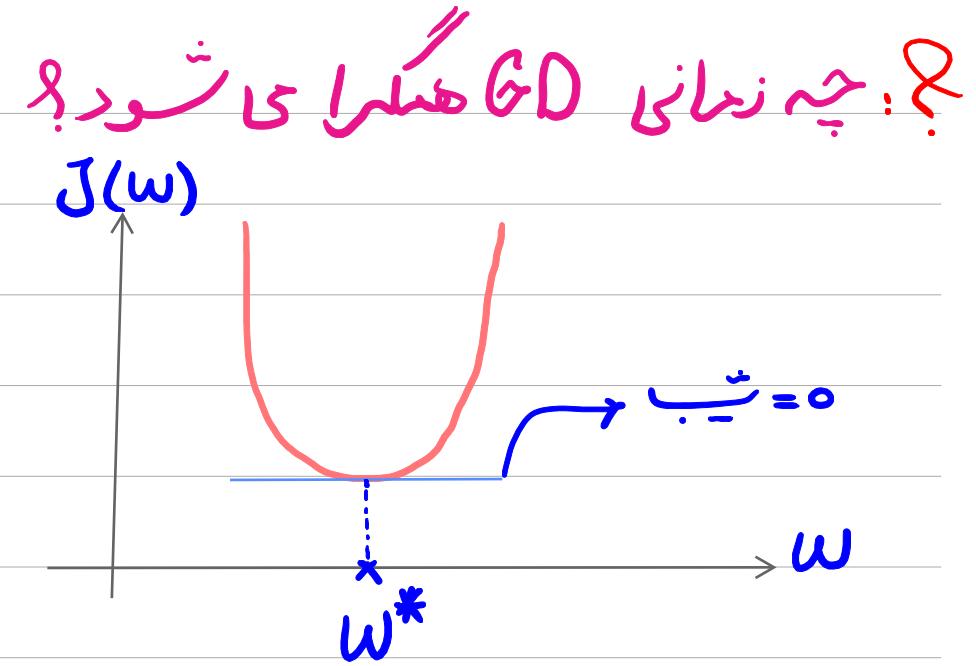
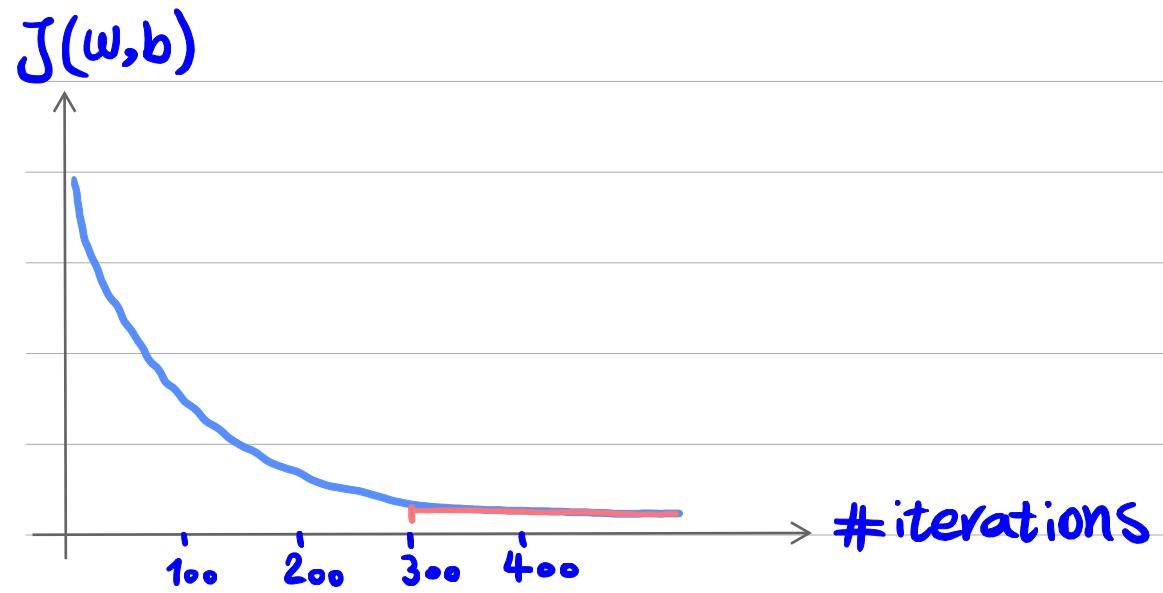
-0.67 3.7

$$0 \leq x_2 \leq 5$$

$$\chi_2 = \frac{x_2 - \mu_2}{\sigma_2} \rightsquigarrow \frac{0 - 2.3}{1.4} \leq \chi_2 \leq \frac{5 - 2.3}{1.4}$$

-1.6 1.9





Goal: Converge to Global minimum

① # iterations

for i in range (1000):

{

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

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

}

1. تعیین حدودار Iteration

2. تعیین حدودار $J(w, b) \leftarrow$ تست هنگرایی

② ε

$$J(w, b)$$

ε

While $\text{decreas} > \text{thr}$:

{

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

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

}

Regression

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

Mean Absolute Error

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

Mean Squared Error

$$\text{RMSE} = \sqrt{\text{MSE}}$$

Root Mean Square Error

x_1	x_2	y

80% Train {

20% Test {

} 70% Train

} 30% Test

