Session 04

Linear Regression | Part02

Machine Learning | Zahra Amini

دیتاست چیست؟

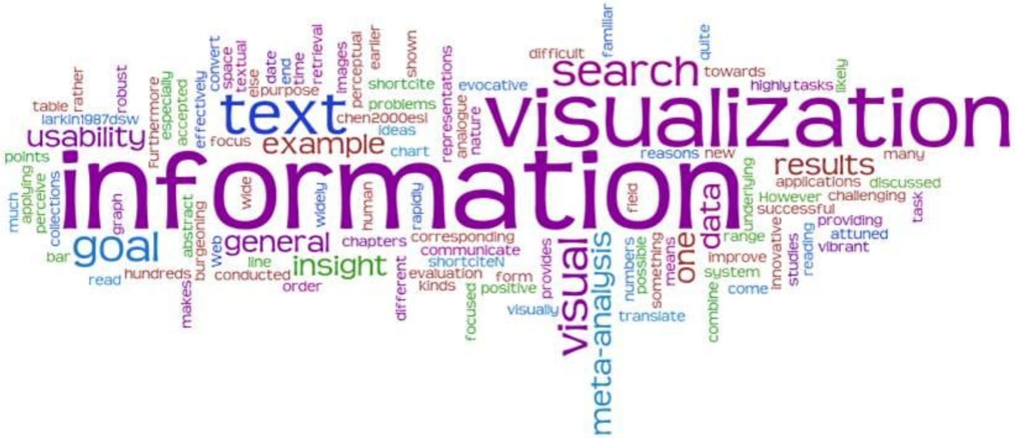
مجموعه داده‌های جمع آوری شده برای آموزش و ارزیابی مدل.

انواع داده‌ها

① ساختاریافته Structured Data

غیرساختاریافته Unstructured Data

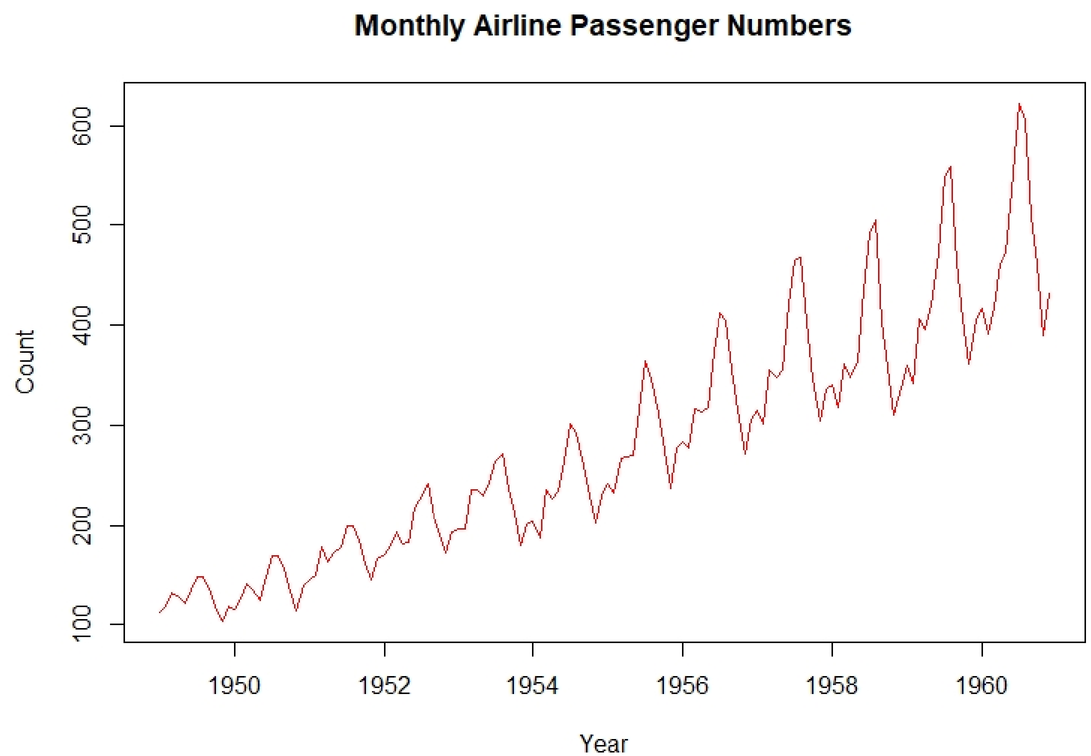
ویژگی	داده‌های ساختاریافته	داده‌های غیرساختاریافته
ساختار	سازمان‌یافته و قابل پیش‌بینی	سازمان‌دهی نشده یا دارای ساختار نامنظم
ذخیره‌سازی	در پایگاه‌ های داده رابطه‌ای (SQL)	در سیستم‌های پایگاه داده NoSQL یا به صورت فایل
مدیریت	آسان برای مدیریت به دلیل یکپارچگی و فهرست‌بندی	مدیریت دشوارتر به دلیل عدم وجود ساختار مشخص
پردازش	پردازش سریع و ساده به خاطر ساختار مشخص	نیاز به ابزارهای پیشرفته برای تحلیل و پردازش
تحلیل داده‌ها	تحلیل ساده و مستقیم	تحلیل پیچیده و گاهی اوقات نیازمند پردازش خاص
مثال‌ها	داده‌های مشتری، تراکنش‌های بانکی	ایمیل‌ها، فایل‌های ویدئویی، پست‌های شبکه‌های اجتماعی



Tabular Data

ID	TOTAL ACTIONS	ACTION 1	ACTION 2	TOTAL TIME
10	120	80	40	0:50:05
11	255	130	125	1:40:03
12	180	100	80	1:20:19
13	305	205	100	1:58:58
14	71	50	21	0:35:41
15	418	310	108	2:08:18
16	222	150	72	1:32:58

Time series data



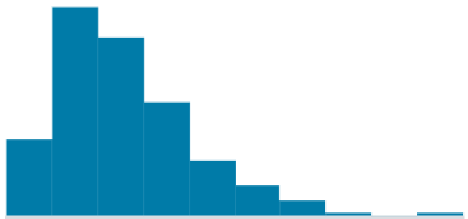
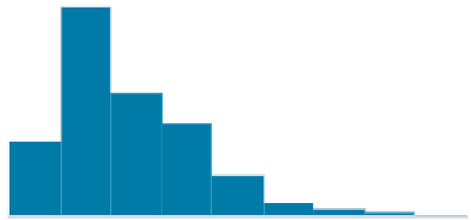
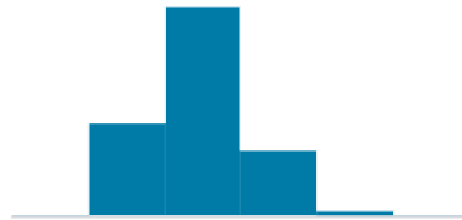
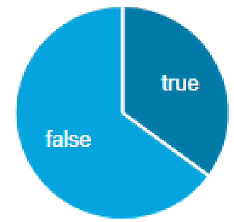
انواع داده ها ← ② ← کتگوریکال Categorical Data

② ← عددی Numerical Data

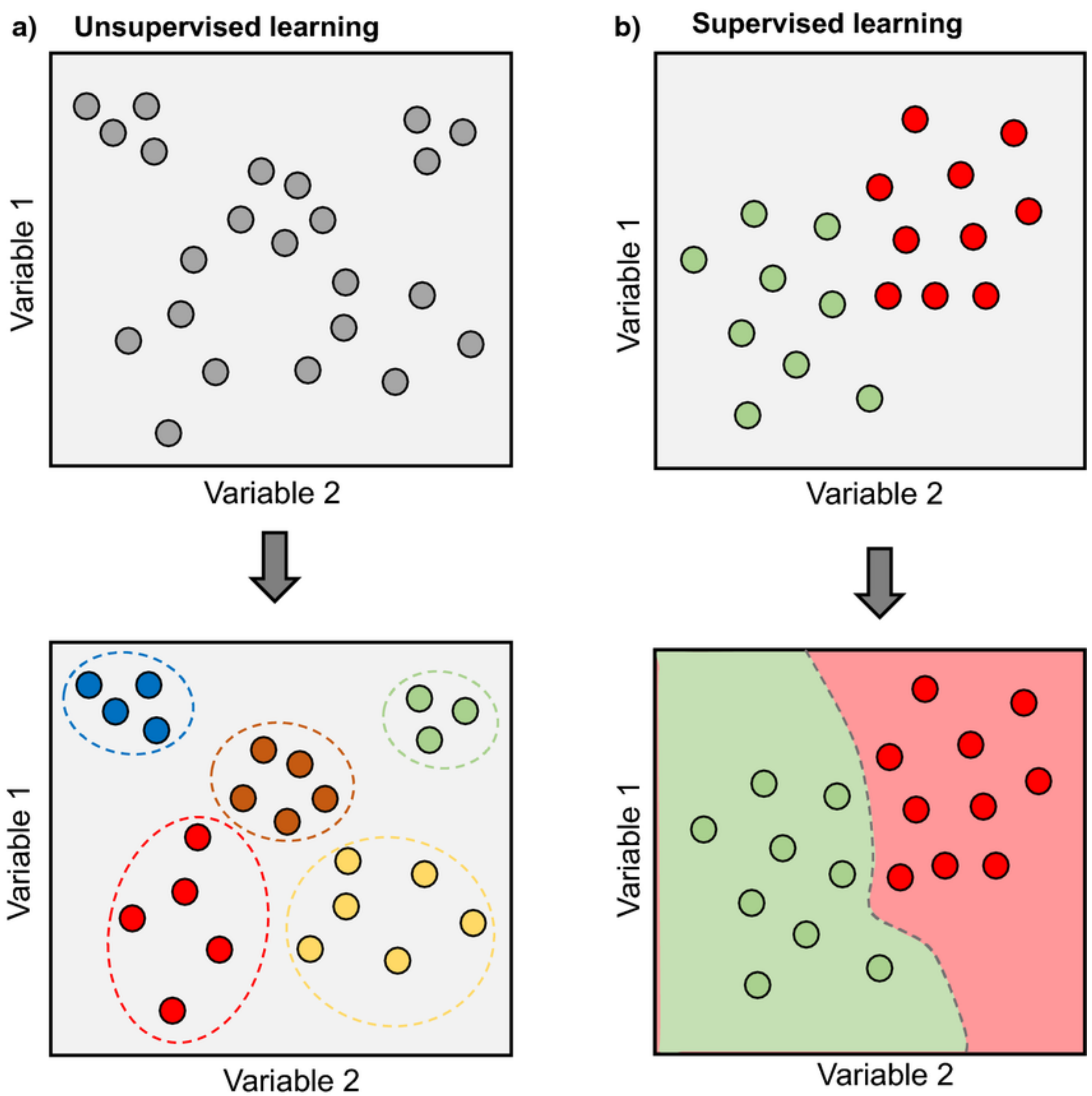
عددی ← پیوسته Continues

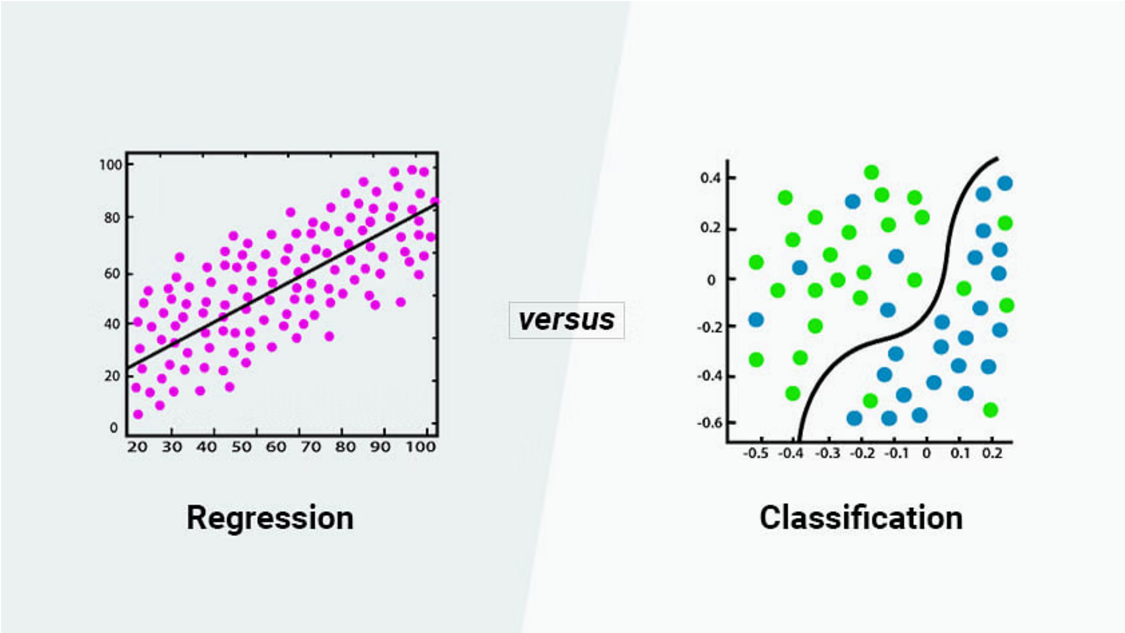
عددی ← گسسته Discrete

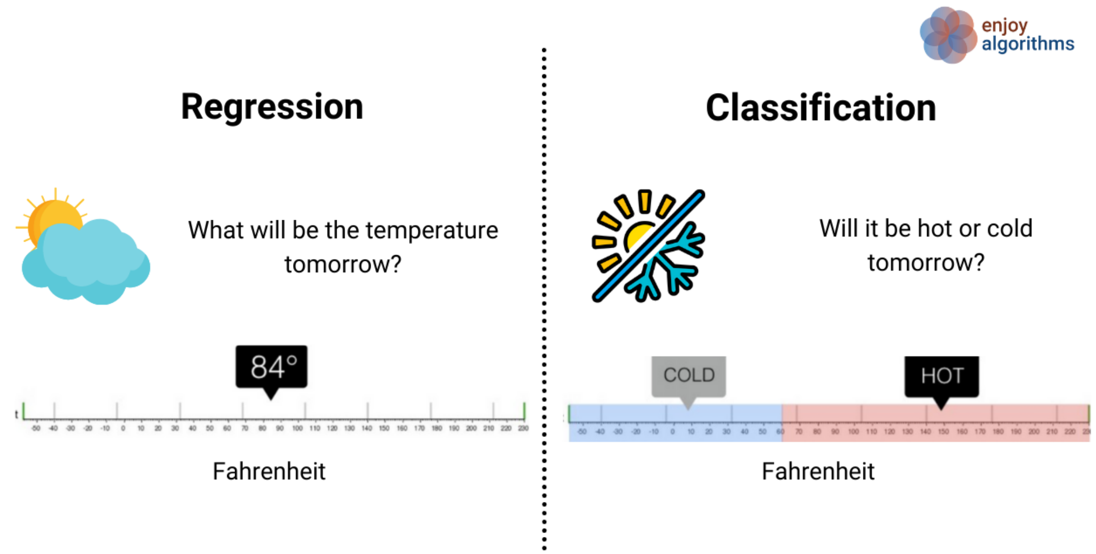
ویژگی	داده‌های عددی	داده‌های کتگوریکال
تعریف	داده‌ها که به صورت اعداد بیان می‌شوند	داده‌ها که به صورت دسته‌بندی‌های مشخص بیان می‌شوند
نوع	پیوسته یا گسسته	نامینال یا ترتیبی
مثال‌ها	سن، قد، وزن، دما	جنسیت، رنگ، رتبه‌بندی، نوع بیمه
عملیات ریاضی	معنادار (جمع، تفریق، میانگین)	غیرمعنادار
استفاده در تجزیه و تحلیل	تجزیه و تحلیل‌های کمی، رگرسیون	تجزیه و تحلیل فراوانی، تست‌های آماری کتگوریکال
تجسم	نمودارهای میله‌ای، خطی، پراکندگی	نمودارهای میله‌ای، دایره‌ای، پله‌ای

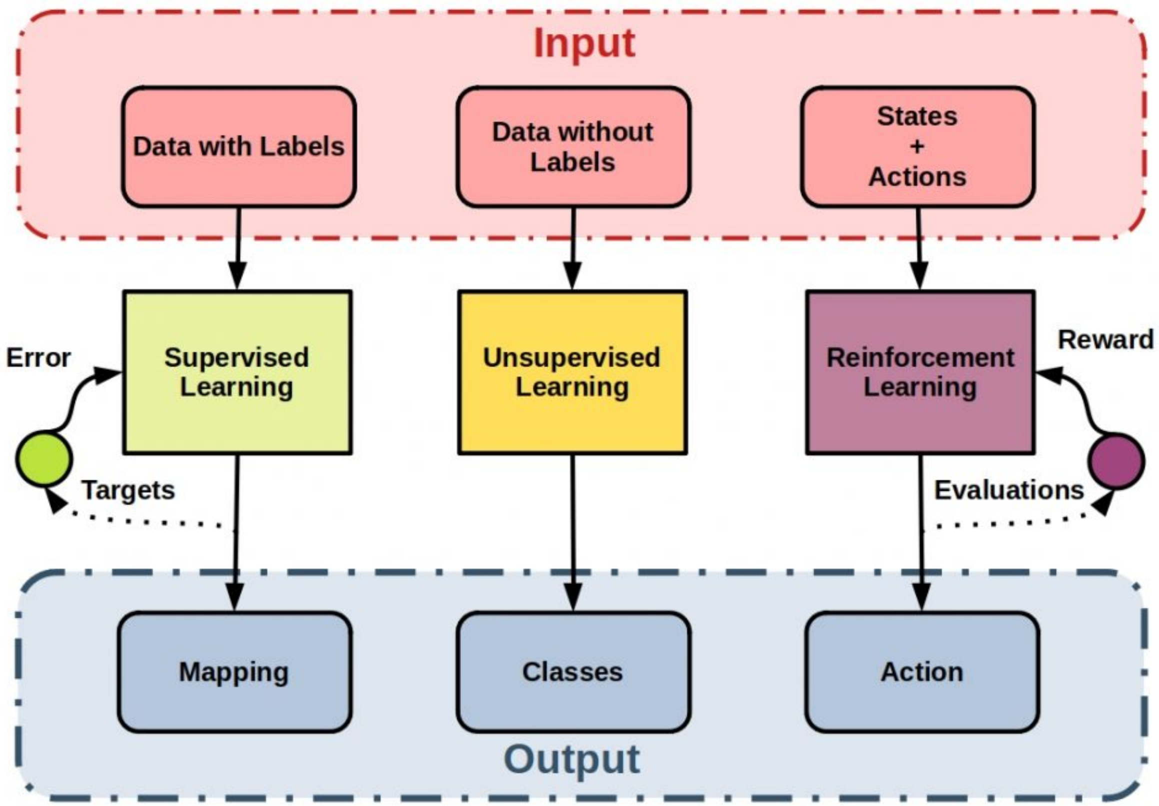
# price	# area	# bedrooms	✓ basement
Price of the Houses	Area of a House	Number of House Bedrooms	Weather has a basement
 1.75m 13.3m	 1650 16.2k	 1 6	 true 191 35% false 354 65%
13300000	7420	4	no
12250000	8960	4	no
12250000	9960	3	yes
12215000	7500	4	yes
11410000	7420	4	yes
10850000	7500	3	yes
10150000	8580	4	no
10150000	16200	5	no
9870000	8100	4	yes
9800000	5750	3	no

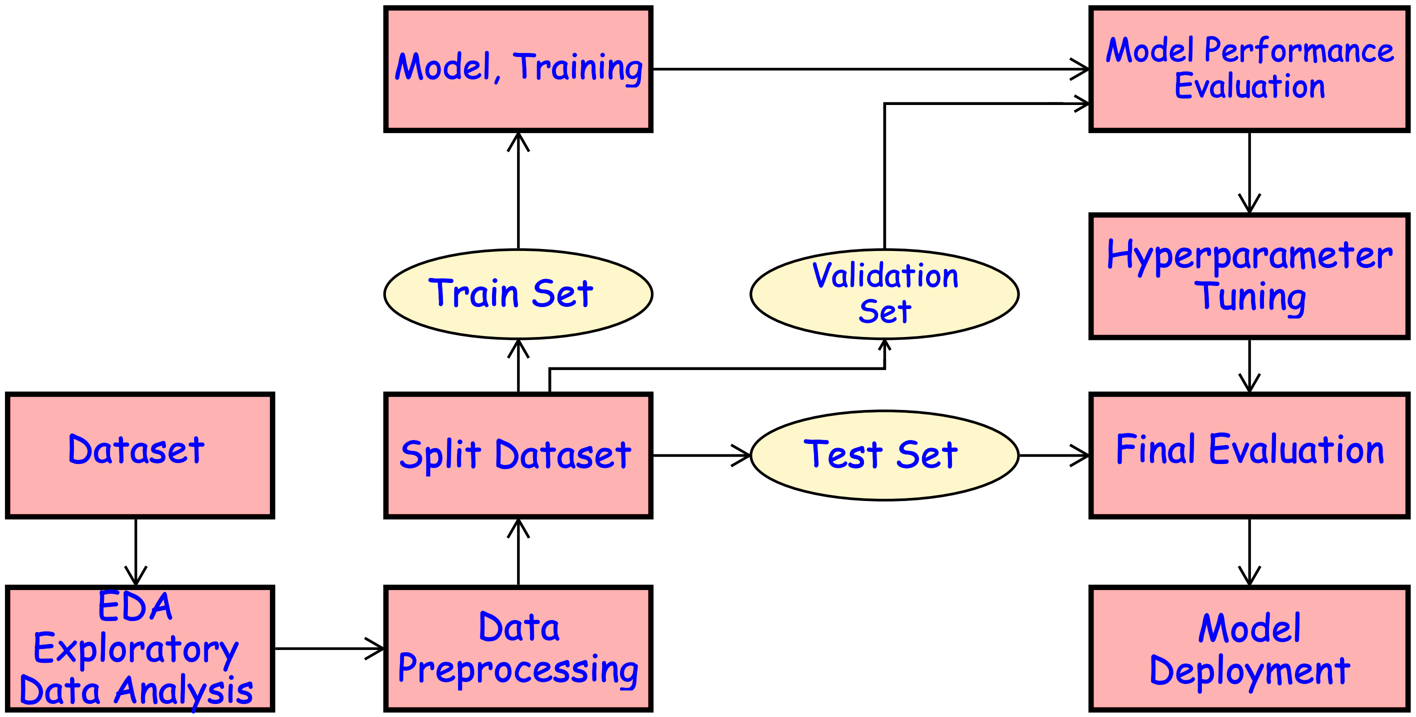
انواع یادگیری
- با نظارت Supervised → انواع مسائل
  - رگرسیون Regression
  - طبقه‌بندی/کلاسه‌بندی Classification
- بدون نظارت Unsupervised → مسئله خوشه‌بندی
- تقویتی Reinforcement

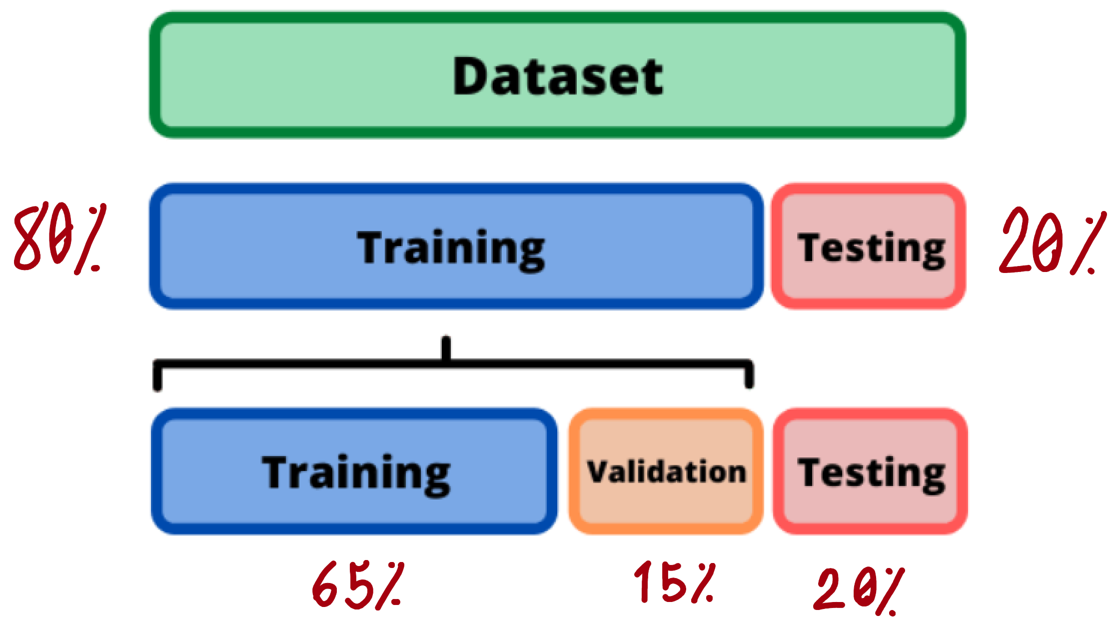



versus
Regression
Classification

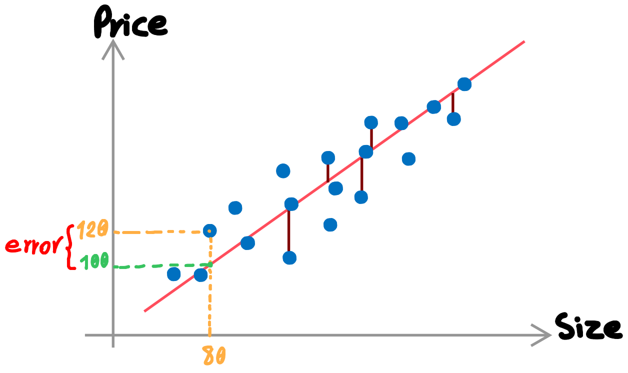

enjoy algorithms
Regression
What will be the temperature tomorrow?
84°
Fahrenheit
Classification
Will it be hot or cold tomorrow?
COLD
HOT
Fahrenheit


Input
Data with Labels
Data without Labels
States + Actions
Error
Supervised Learning
Unsupervised Learning
Reinforcement Learning
Reward
Targets
Evaluations
Mapping
Classes
Action
Output


Model, Training
Model Performance Evaluation
Hyperparameter Tuning
Train Set
Validation Set
Dataset
Split Dataset
Test Set
Final Evaluation
EDA Exploratory Data Analysis
Data Preprocessing
Model Deployment


Dataset
80%
Training
Testing
20%
Training
Validation
Testing
65%
15%
20%

Loss Function ⇝ Error ⇝ به اختلاف بین مقدار واقعی و پیش‌بینی مدل خطا می‌گویند.



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. $R2\text{-}Score = 1 - \dfrac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$

$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \rightarrow SS_R$ (SUM of Squares for Regression)

$\sum_{i=1}^{n}(y_i - \bar{y})^2 \rightarrow SS_m$ (Sum of Squares for Model)

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

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

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

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

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

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

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

R2 چه عددی باشه عملکرد مدل خوبه؟

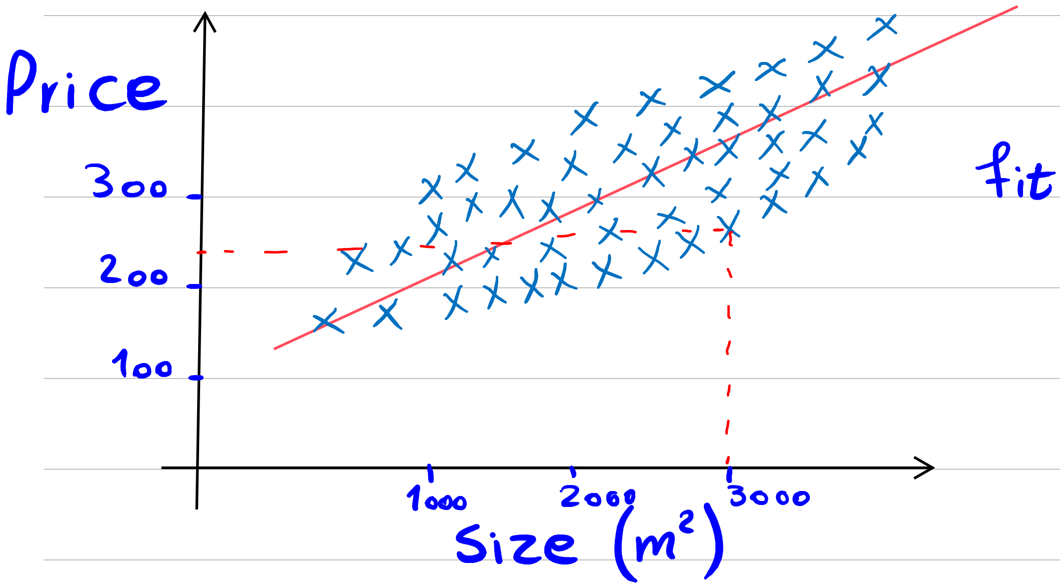
R2 =1: Perfect fit

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) Size (m²)	y (target) 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 $\longrightarrow J(w,b)$

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

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

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

GD Algorithm:

$W=0$, $b=0$

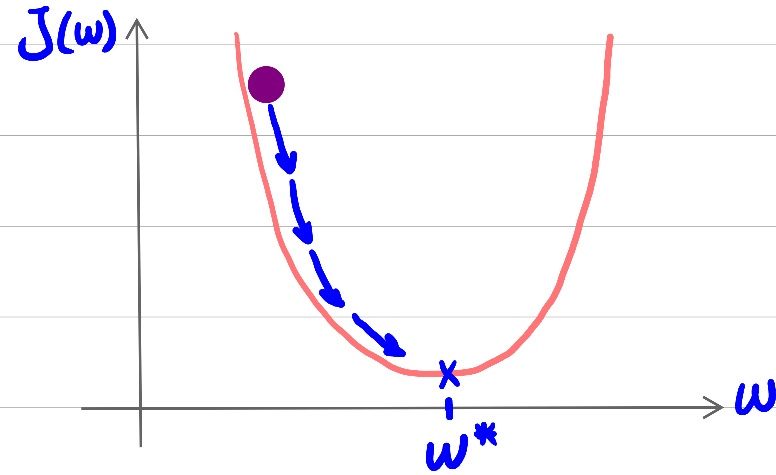
repeat until convergence

{

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

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

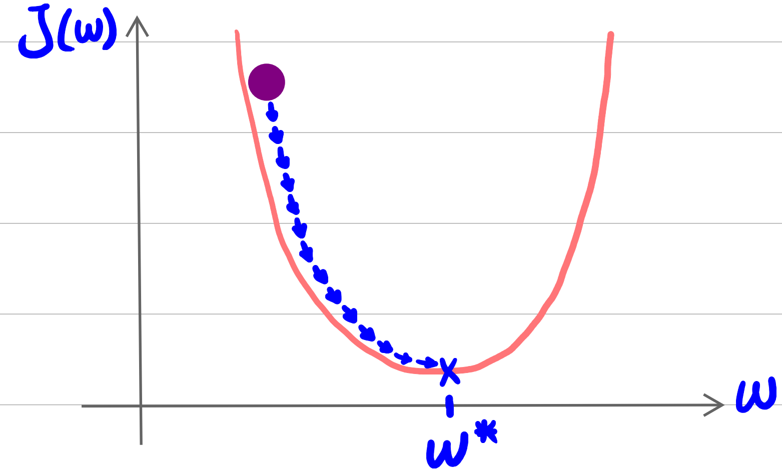
}



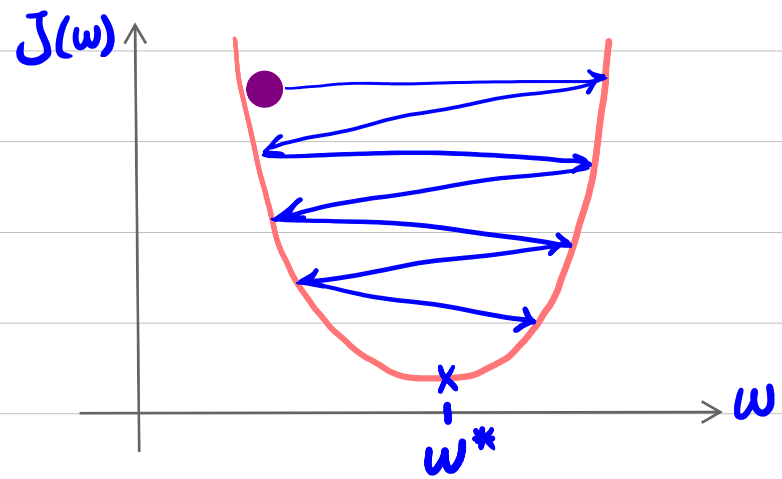
$\longrightarrow$ Minimum

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
$f_{w,b}(x) = wx + b$	$f_{w,b}(x) = w_1 x_1 + w_2 x_2 + \dots + w_n x_n + b$

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

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

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

dot product

Multiple Linear Regression:

$n=3$

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

$$f_{\vec{w},b}(\vec{x}) = w_1x_1 + w_2x_2 + w_3x_3 + b$$

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

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

vectorization

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

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

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

$$(\underset{10}{\cancel{1\times10}}) + (\underset{50}{\cancel{2.5\times20}}) + (\underset{-99}{\cancel{-3.3\times30}}) + \overset{b}{4} = -35$$

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	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

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

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

$$f_{w,b}(X) = W_1 X_1 + W_2 X_2 + \dots + W_n X_n + b$$

یا

$$f_{w,b} = W \cdot X + b$$

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

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

$a_1 =$ $\qquad a_2 =$

Feature Size and Parameter Size:

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

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

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 $w_1 = 50$, $w_2 = 0.1$, $b = 50$

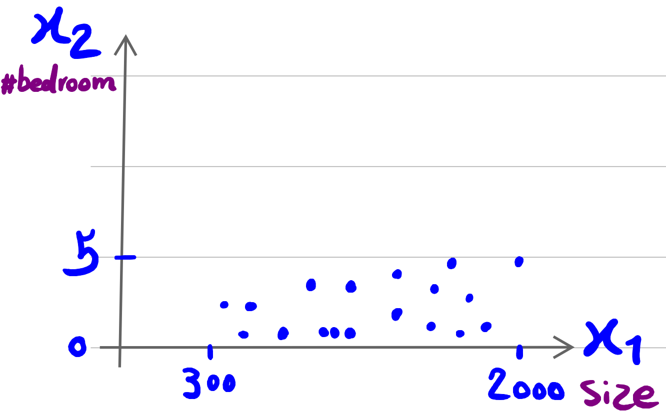
$$Price\_Predict = \overset{w_1}{50} * \overset{x_1}{2000} + \overset{w_2}{0.1} * \overset{x_2}{5} + \overset{b}{50} = 100{,}050.5k$$

② $w_1 = 0.1$, $w_2 = 50$, $b = 50$

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

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

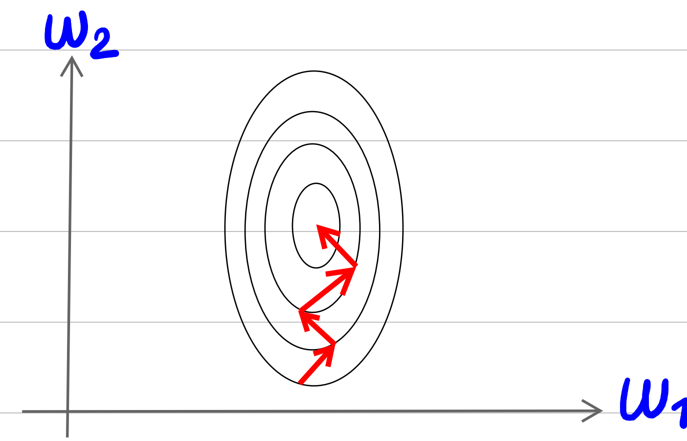
Features:

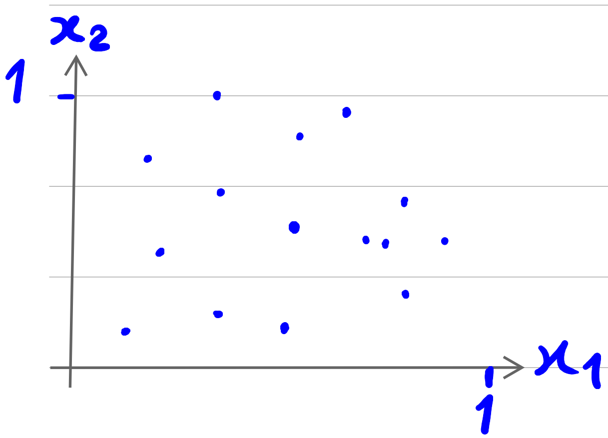


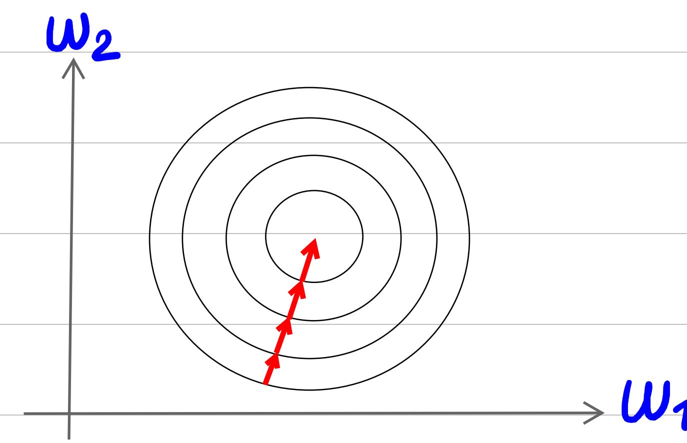
$300 \leq x_1 \leq 2000$

$0 \leq x_2 \leq 5$

Parameters:







Feature Scaling:

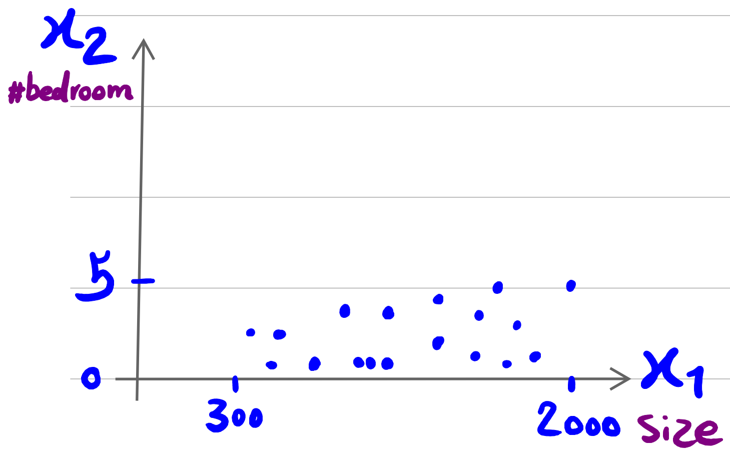
$$x_{1\text{-Scaled}} = \frac{x_1}{\underset{\text{max}}{2000}} \qquad 0 \leq x_{1\text{-Scaled}} \leq 1$$

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

Mean Normalization:

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

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



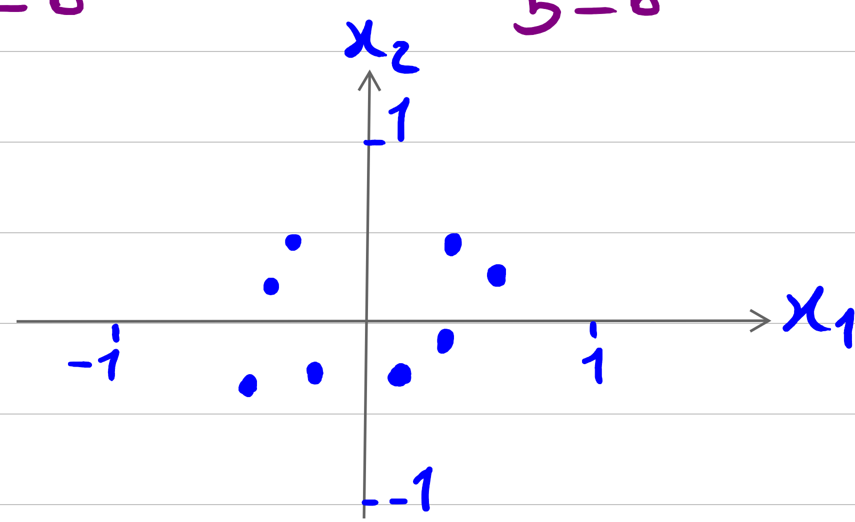
$$300 \leq x_1 \leq 2000$$

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

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

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

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



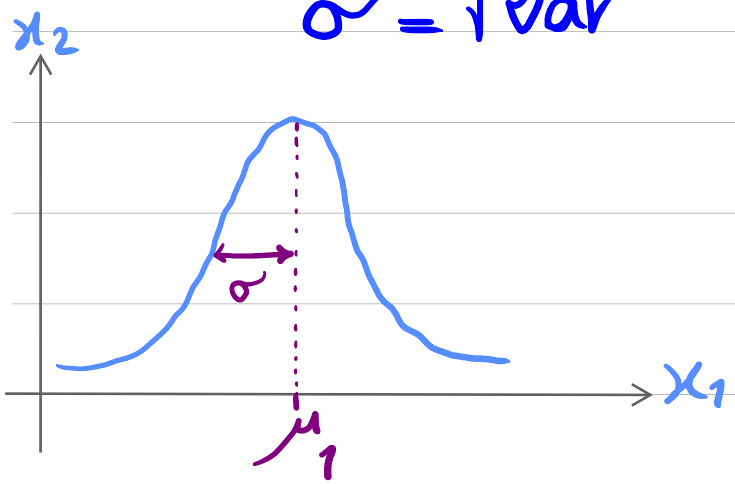
Z-Score Normalization :

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

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

standard deviation یا σ یا انحراف معیار

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

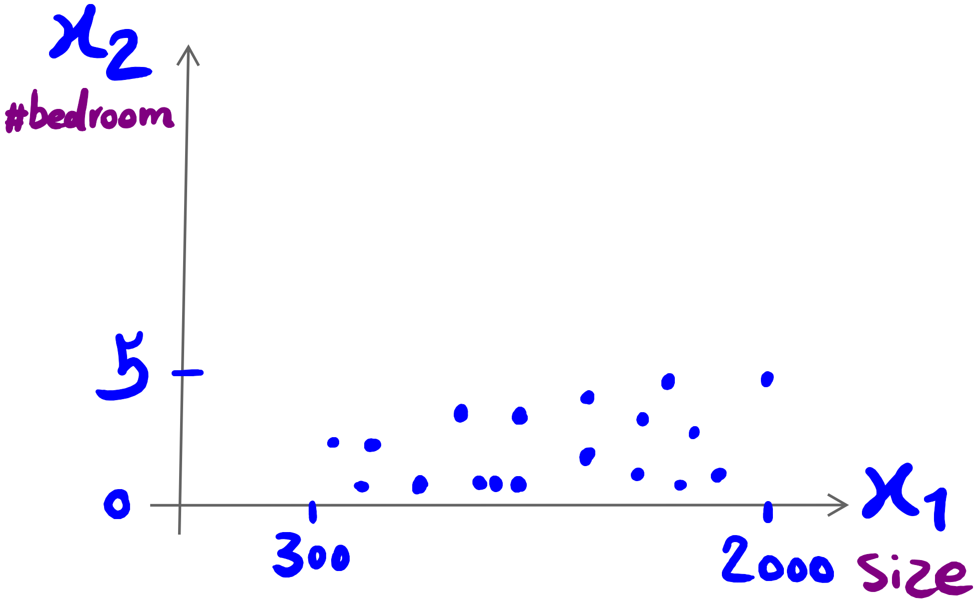


StandardScaler()

در Sklearn

Z-Score Normalization:

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



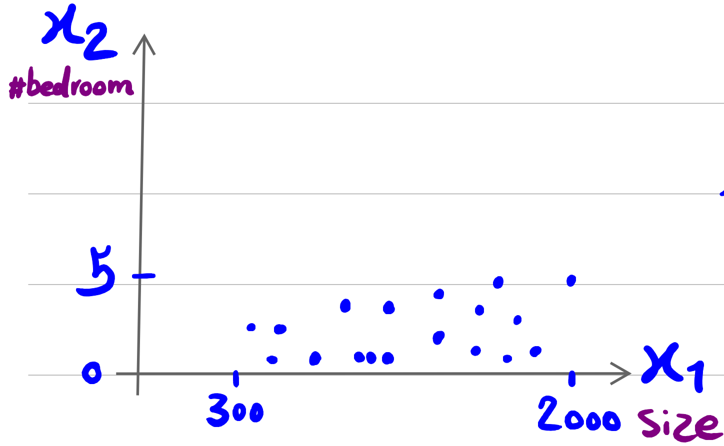
$\mu_1 = 600$ $\sigma_1 = 450$

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

$a \leqslant x_1 \leqslant b$

$c \leqslant x_2 \leqslant d$

$a, b, c, d = ?$



$$\mu_1 = 600 \qquad \sigma_1 = 450$$
$$\mu_2 = 2.3 \qquad \sigma_2 = 1.4$$

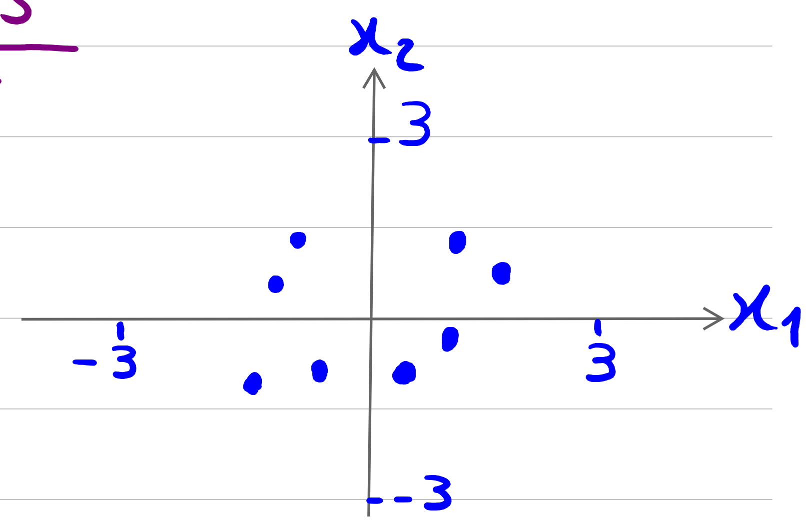
$$x = \frac{x - \mu}{\sigma}$$

$$300 \leqslant x_1 \leqslant 2000$$

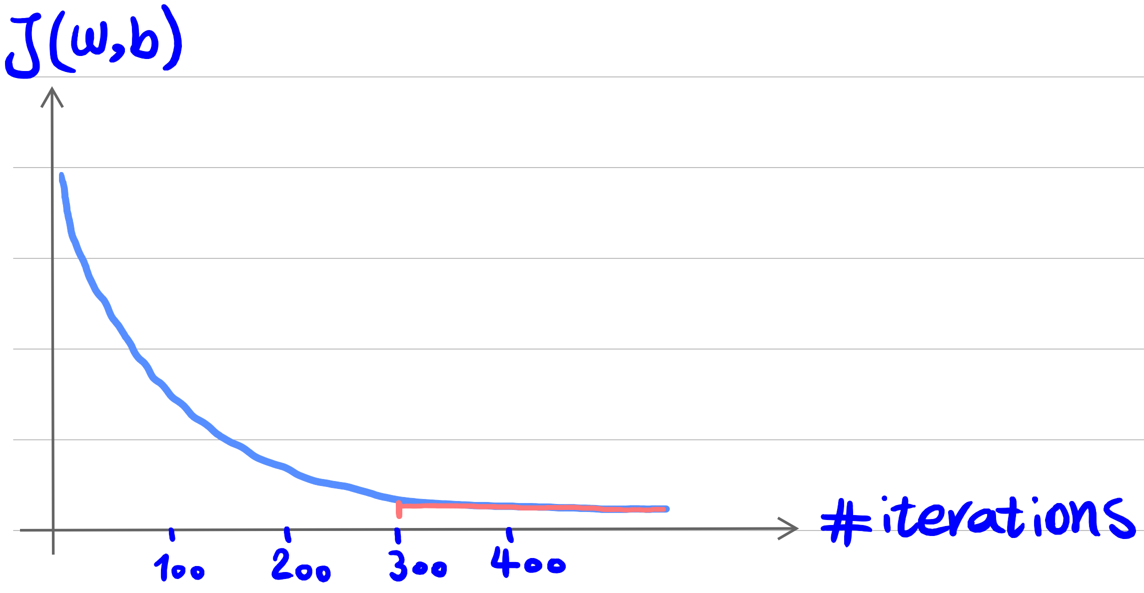
$$x_1 = \frac{x_1 - \mu_1}{\sigma_1} \leadsto \overset{-0.67}{\frac{300-600}{450}} \leqslant x_1 \leqslant \overset{3.1}{\frac{2000-600}{450}}$$

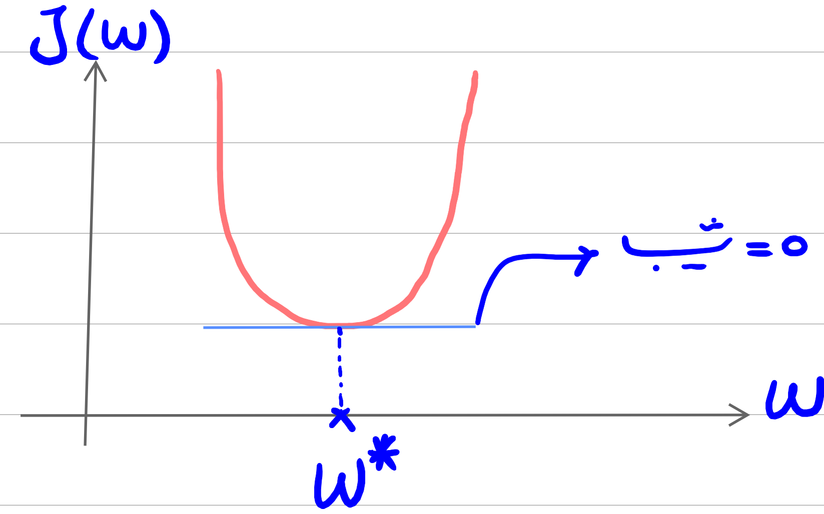
$$0 \leqslant x_2 \leqslant 5$$

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



؟: چه زمانی GD همگرا می شود؟





Goal: Converge to Global minimum

① #iterations

1. تعیین مقدار Iteration

2. تعیین مقدار J(w,b) ← تست همگرایی

for i in range (1000):

{

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

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

}

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