Session 5&6

Linear Regression | Part01

NLP | Machine Learning | Zahra Amini

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

https://zil.ink/zahraamini

Supervised Learning

$X \longrightarrow y$

input — output, label

- Email Spam (0,1)
- House Price

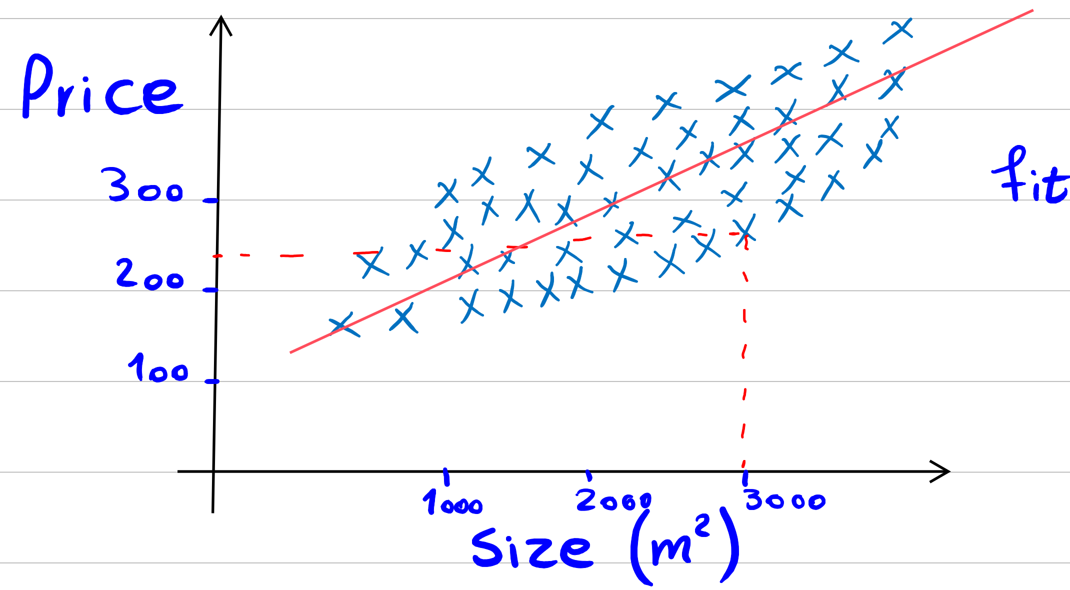
Supervised

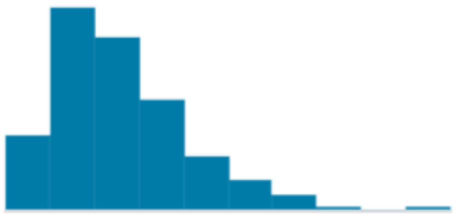
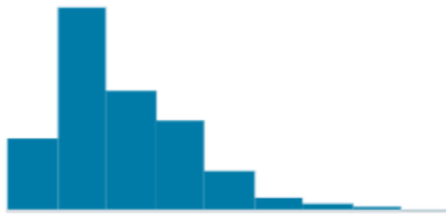
- Regression
- Classification

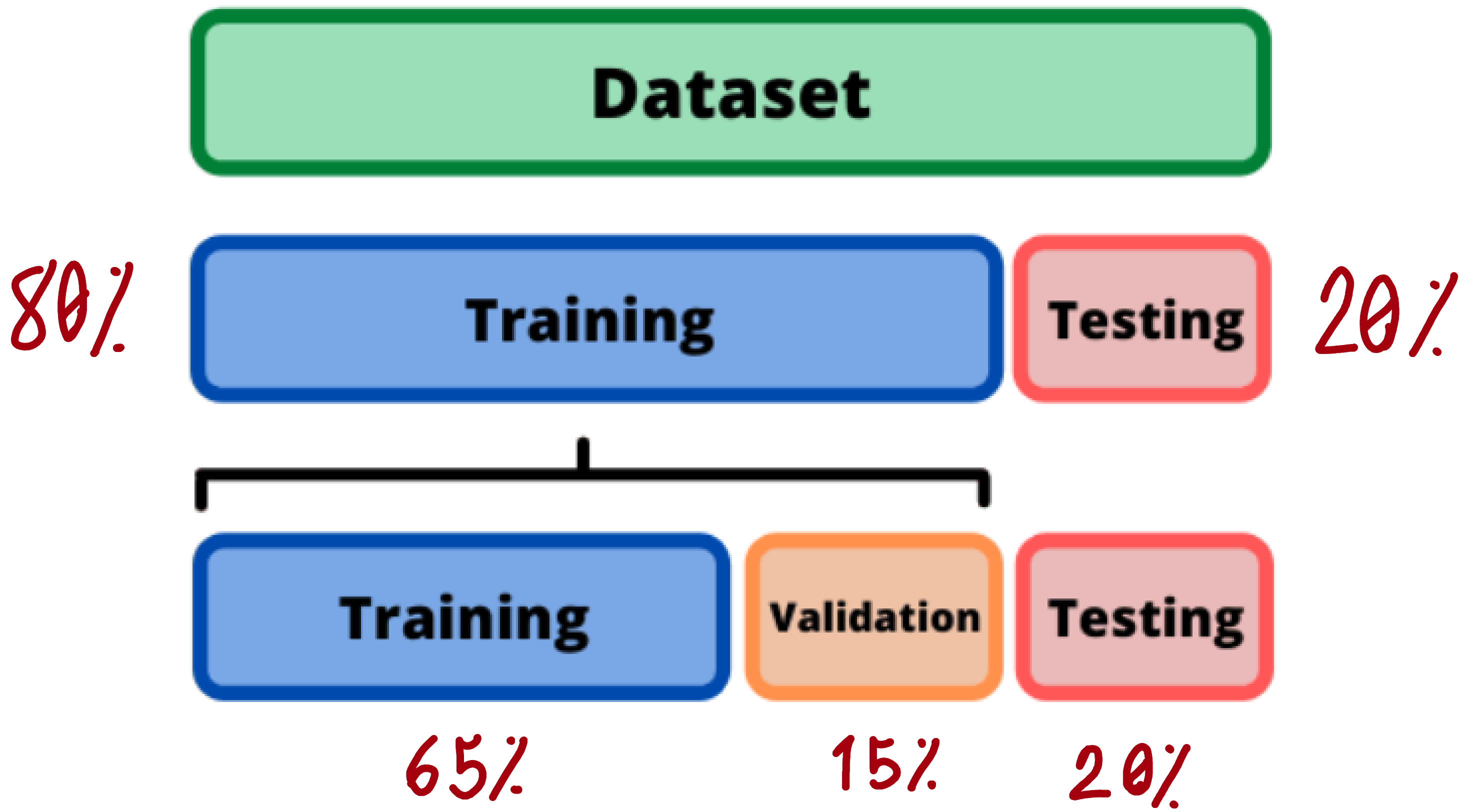
Linear Regression → Housing Price Prediction

می‌خواهیم با توجه به متراژ خانه قیمتش را پیش‌بینی کنیم:

Size (m^2)	Price
180	1000
200	1200
300	2000
⋮	⋮



# 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


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

نکته مهم: در خروجی ما بی نهایت عدد می توانیم داشته باشیم، بنابراین از رگرسیون برای پیش بینی متغیرهای پیوسته می توان استفاده کرد.

Continous

$x \rightarrow$ feature

$y \rightarrow$ target

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

$x^{(1)} = 2104 \qquad y^{(1)} = 460$

$(x^{(1)}, y^{(1)}) = (2104, 460)$

Training Set

$x \rightarrow$ مدل $\rightarrow y$

Notation:

x : Input

y : output

m : # training examples 4

n : # features 1

(x, y)

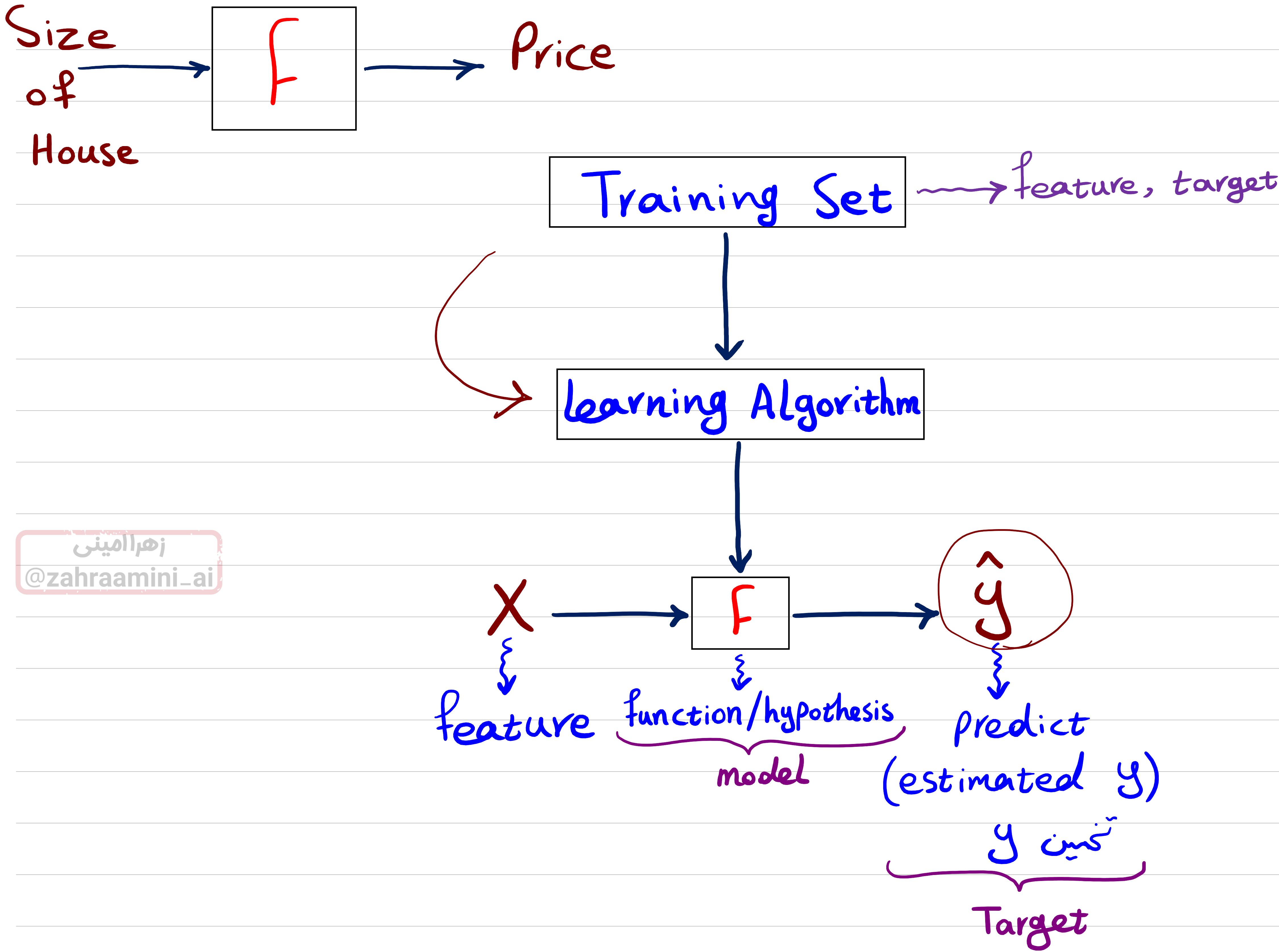
$\rightarrow (x^{(i)}, y^{(i)})$: i^{th} training example

Test Set

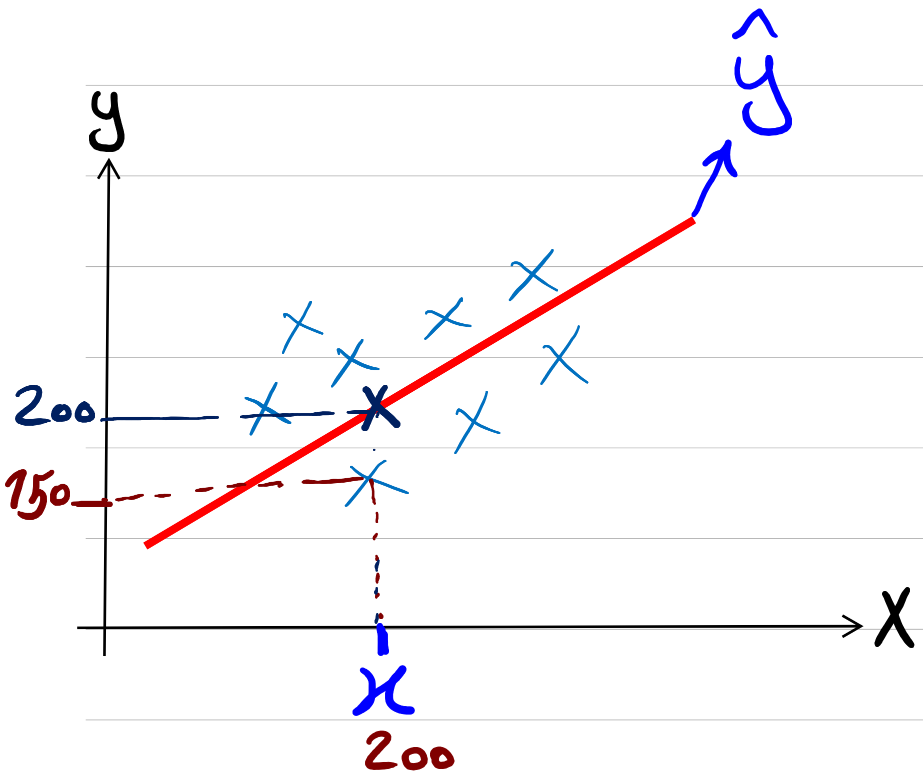
X

$Size\ (m^2)$
350
2100
833
120
⋮

$\hat{y}$
?
?
?
?
?


Size of House
F
Price
Training Set
feature, target
learning Algorithm
X
F
ŷ
feature
function/hypothesis
model
predict
(estimated y)
تخمین y
Target

؟: f چیست؟

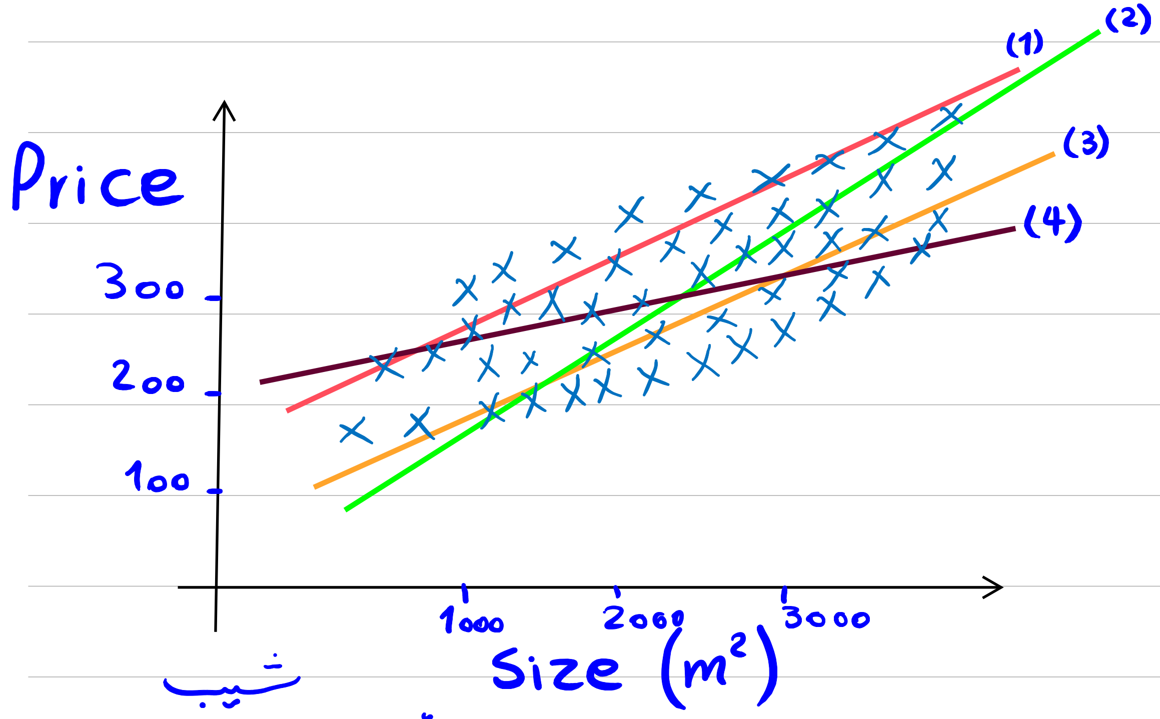


$y = ax + b$ ⟶ Line

$f_{w,b}(x) = WX + b$ ⟶ Linear function

$f(x) = WX + b$

ج: کدام f بهتر است؟ از کجا بفهمیم؟



model: $f_{w,b}(x) = wx + b$

parameters: w, b

$f = wx + b$

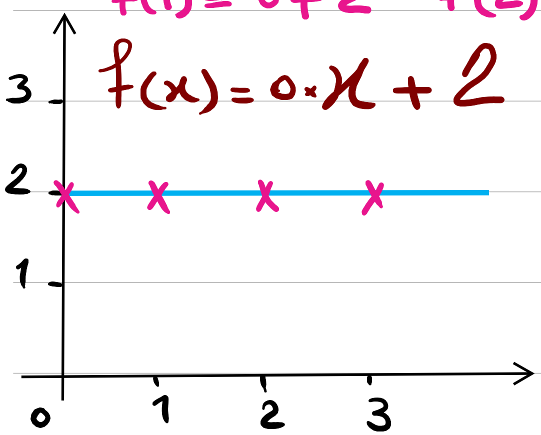
w → شیب

b → عرض از مبدأ

: Cost Function

$f(1) = 0 + 2 \quad f(2) = 2$

$f(x) = 0 \times x + 2$



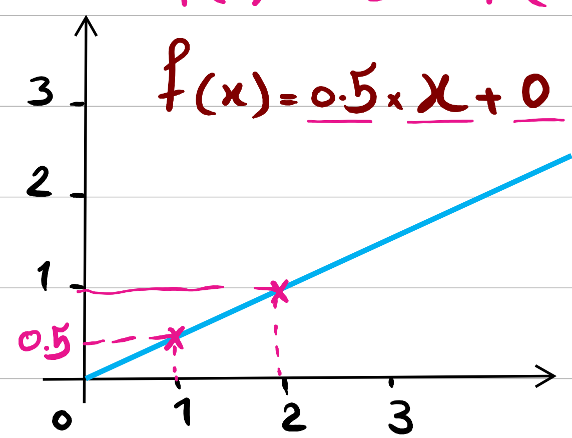
$W = 0, \quad b = 2$

↳ Slope

شیب

$f(1) = 0.5 \quad f(2) = 1$

$f(x) = 0.5 \times x + 0$

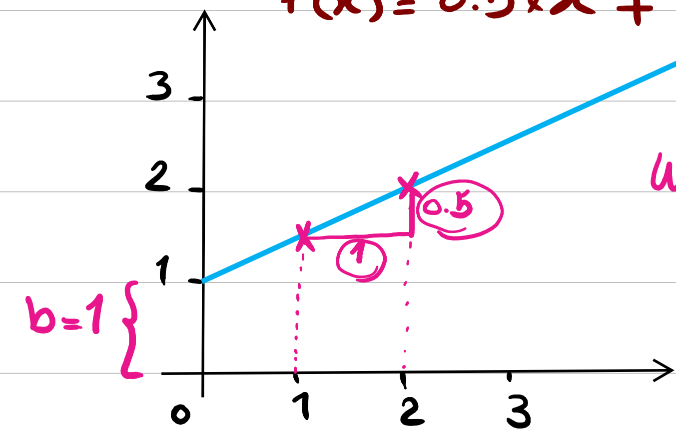


$W = 0.5, \quad b = 0$

$f(x) = 0.5 \times x + 1$

$f(1) = 1.5$

$f(2) = 2$

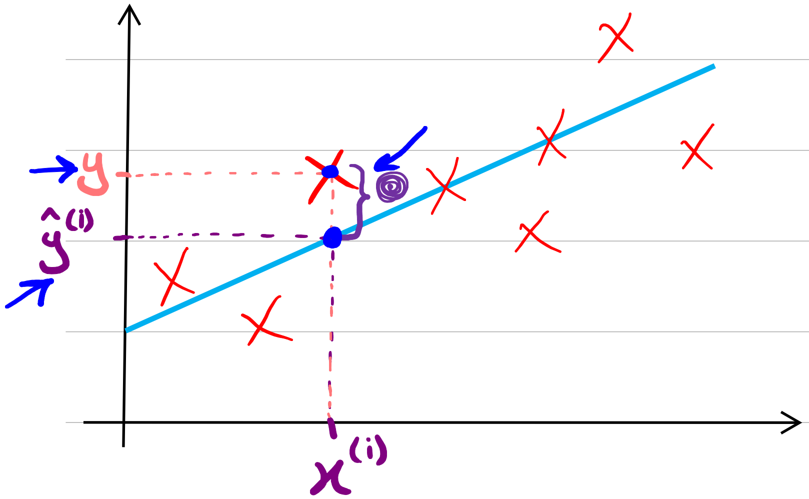


$W = \frac{0.5}{1} = 0.5$

$\frac{2 - 1.5}{2 - 1} = \frac{0.5}{1}$

$W = 0.5, \quad b = 1$

ج: w و b مناسب کدام هستند؟ با .Linear Reg می‌خواهیم مقادیری برای w و b پیدا کنیم.



چه w و b خوب هست؟

w و b خوب است که $\hat{y}$ به ما بدهد که به y نزدیک‌تر باشد.

$$\hat{y}^{(i)} \simeq y^{(i)} \rightsquigarrow \left(x^{(i)}, y^{(i)}\right)$$

◎ خطا ← error

Cost function: $J(w,b) = \frac{1}{2m} \sum_{i=1}^{m} \underbrace{\left(\hat{y}^{(i)} - y^{(i)}\right)^2}_{\text{error}}$

m: # training examples

یکی از معروف‌ترین cost function ها برای .Reg مجذور خطا است.

Mean Square Error

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

خب حالا بریم تنظیم پارامترهای w, b برای بهبود عملکرد مدل:

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

$$\hat{y}=f(x)=wx+b$$

Ex:

Model $\longrightarrow f_{w,b}(x)$

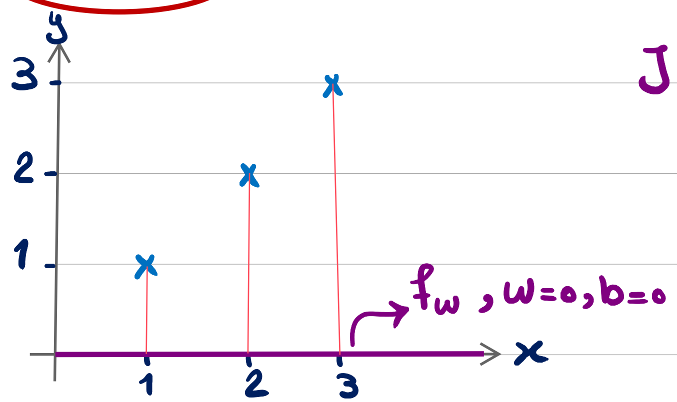
Parameters: w , $b=0$

x	y
1	1
2	2
3	3

Cost function $\longrightarrow J(w)=\frac{1}{2m}\sum_{i=1}^{m}\left(f_w(x^{(i)})-y^{(i)}\right)^2$

Goal $\longrightarrow \underset{w,b}{\text{minimize}}\ J(w,b)$

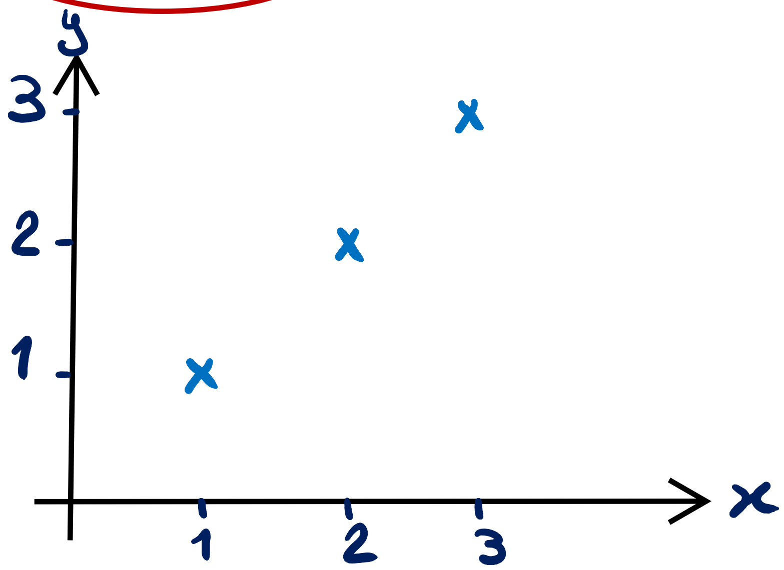
W=0



$$J(0)=\frac{1}{2\times 3}\left[\overset{1}{\cancel{(0-1)^2}}+\overset{4}{\cancel{(0-2)^2}}+\overset{9}{\cancel{(0-3)^2}}\right]=\frac{1}{6}\left[1+4+9\right]=\frac{13}{6}=2.3$$

$$\rightarrow \hat{y}=\begin{bmatrix}0\\0\\0\end{bmatrix}$$

$W = 0.5$



(x, y)

x	y
1	1
2	2
3	3

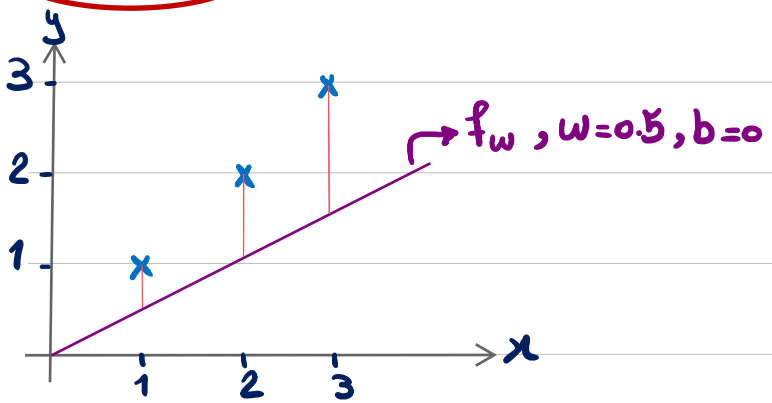
$m = 3$

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

$$f_w(x) = wx + b$$

$$J(0.5) = ?$$

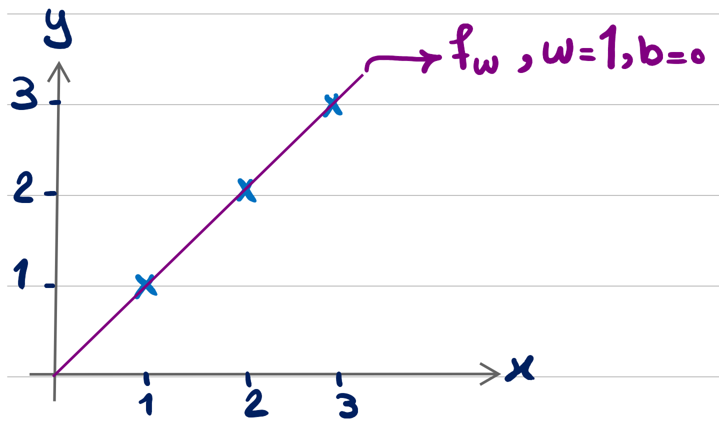
$w=0.5$



$$J(0.5)=\frac{1}{6}\left[\overset{0.25}{\cancel{(0.5-1)^2}}+\overset{1}{\cancel{(1-2)^2}}+\overset{2.25}{\cancel{(1.5-3)^2}}\right]=\frac{1}{6}(3.5)=0.58$$

$$\hat{y}=\begin{bmatrix}0.5\\1\\1.5\end{bmatrix}$$

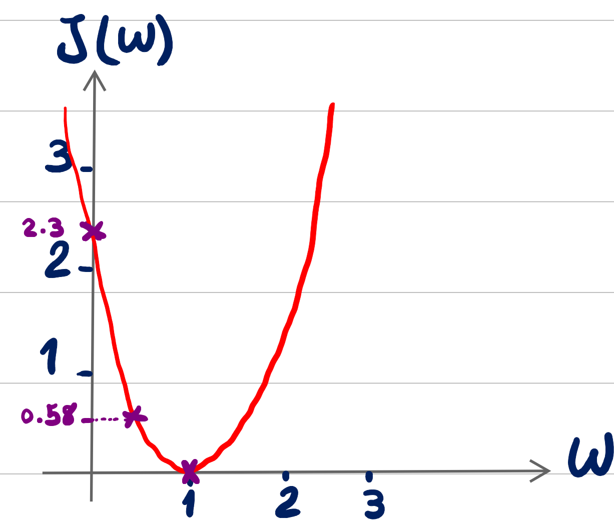
$w=1$



$$J(1)=\frac{1}{6}\left[(1-1)^2+(2-2)^2+(3-3)^2\right]=0$$

$$\hat{y}=\begin{bmatrix}1\\2\\3\end{bmatrix}$$

ما به دنبال کمترین خطا بودیم، کدام w کمترین خطا را به ما می دهد؟ $w=1$

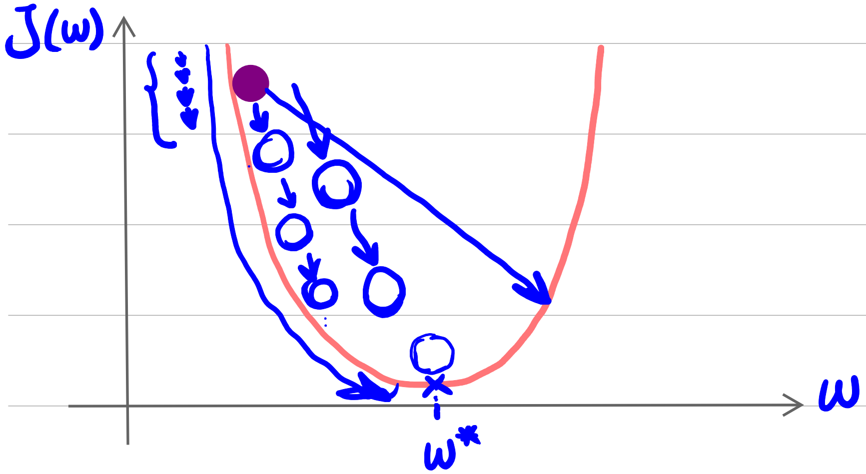


Gradient Descent (GD) ← خب حالا می خواهیم بریم سراغ الگوریتمی که تابع J را به حداقل برساند.

function → $J(w,b)$

از GD برای به حداقل رساندن هر تابعی می توان استفاده کرد.

Goal → min $J(w,b)$



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

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

$$f_{w,b}(x) = wx + \cancel{b}^{\,0} \xrightarrow{\frac{\partial}{\partial w}} x \qquad * \; (f^2)' = 2f'f$$

$$\frac{\partial}{\partial w} J(w,b) = \frac{\partial}{\partial w} \frac{1}{2m} \sum_{i=1}^{m} \left(f_{w,b}(x^{(i)}) - y^{(i)}\right)^2 = \frac{\partial}{\partial w} \frac{1}{2m} \sum_{i=1}^{m} \left(wx^{(i)} + b - y^{(i)}\right)^2$$

$$= \frac{1}{\cancel{2}m} \sum_{i=1}^{m} \left(wx^{(i)} + b - y^{(i)}\right) \times \cancel{2}\, x^{(i)} = \frac{1}{m} \sum_{i=1}^{m} \left(f_{w,b}(x^{(i)}) - y^{(i)}\right) x^{(i)}$$

$$f_{w,b}(x) = wx + b \xrightarrow{\partial/\partial b} 1$$

$$\frac{\partial}{\partial b} J(w,b) = \frac{\partial}{\partial b} \frac{1}{2m} \sum_{i=1}^{m} \left(f_{w,b}(x^{(i)}) - y^{(i)}\right)^2 = \frac{\partial}{\partial b} \frac{1}{2m} \sum_{i=1}^{m} \left(wx^{(i)} + b - y^{(i)}\right)^2$$

$$= \frac{1}{\cancel{2}m} \sum_{i=1}^{m} \left(wx^{(i)} + b - y^{(i)}\right) \times \cancel{2} = \frac{1}{m} \sum_{i=1}^{m} \left(f_{w,b}(x^{(i)}) - y^{(i)}\right)$$

GD Algorithm:

$w = 0$, $b = 0$

repeat until convergence

$\{$

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

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

$\}$

x	y	$\hat{y}$
1	1	1
2	3	1
3	3	1

Ex: برای x و y مدل رگرسیون خطی را بعداز 1 بار آپدیت به‌دست آورید.
$(\alpha = 0.1,\ b = 1,\ w = 0)$

$$w := w - \alpha \boxed{\frac{\partial}{\partial w} J(w,b)} \longrightarrow w_{new} = 0 - \left(\tfrac{1}{10}\right) \times \left(-\tfrac{10}{3}\right) = 0 + \tfrac{1}{3} = 0.33$$

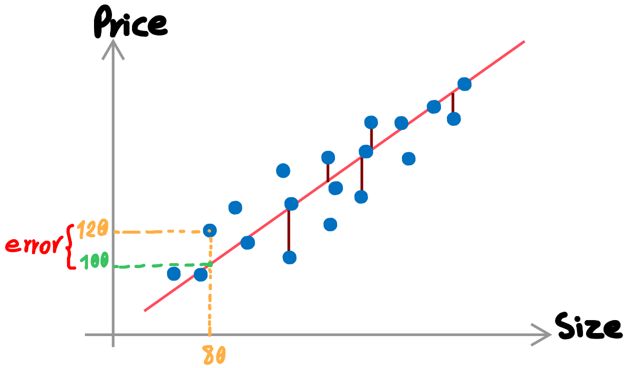
$$\frac{1}{m} \sum_{i=1}^{m} \left(f_{w,b}(x^{(i)}) - y^{(i)}\right) x^{(i)} = \frac{1}{3} \times \left[\cancel{(1-1)}^{0} \times 1 + \cancel{(1-3)}^{-4} \times 2 + \cancel{(1-3) \times 3}^{-6}\right] = -\frac{10}{3}$$

$$b := b - \alpha \frac{\partial}{\partial b} J(w,b) \longrightarrow b_{new} = \underline{1} - \underline{\left(\tfrac{1}{10}\right) \times \left(-\tfrac{4}{3}\right)} = 1.13$$

$$\frac{1}{m} \sum_{i=1}^{m} \left(f_{w,b}(x^{(i)}) - y^{(i)}\right) = \frac{1}{3} \times \left[\cancel{(1-1)}^{0} + \cancel{(1-3)}^{-2} + \cancel{(1-3)}^{-2}\right] = -\frac{4}{3}$$

$\hookrightarrow w_{new} = 0.33$ $\qquad$ $b_{new} = 1.13$ $\qquad$ $f_{w,b}(x) = 0.33x + 1.13$

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



Regression Loss

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

$$2.\ MAE = \frac{1}{m}\sum_{i=1}^{m}|\hat{y}_i - y_i|$$

$$3.\ RMSE = \sqrt{MSE}$$

$$4.\ R2\text{-}Score = 1 - \frac{\overbrace{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}^{SS_R}}{\underbrace{\sum_{i=1}^{n}(y_i - \bar{y})^2}_{SS_M}}$$

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

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

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

$$R2 = 1 - \frac{SS_R}{SS_m} = 1 - \frac{30}{20} = -0.5 \ ?$$ 😳 -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 باشد عملکرد مدل خوبه؟

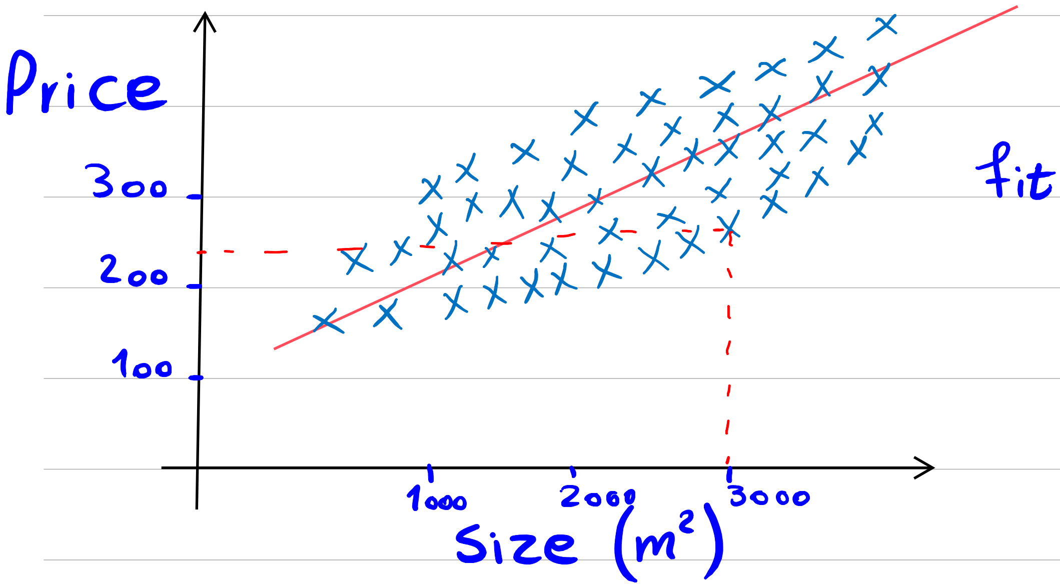
R2 =1: Perfect fit

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

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

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

GD Algorithm:

$w=0$, $b=0$

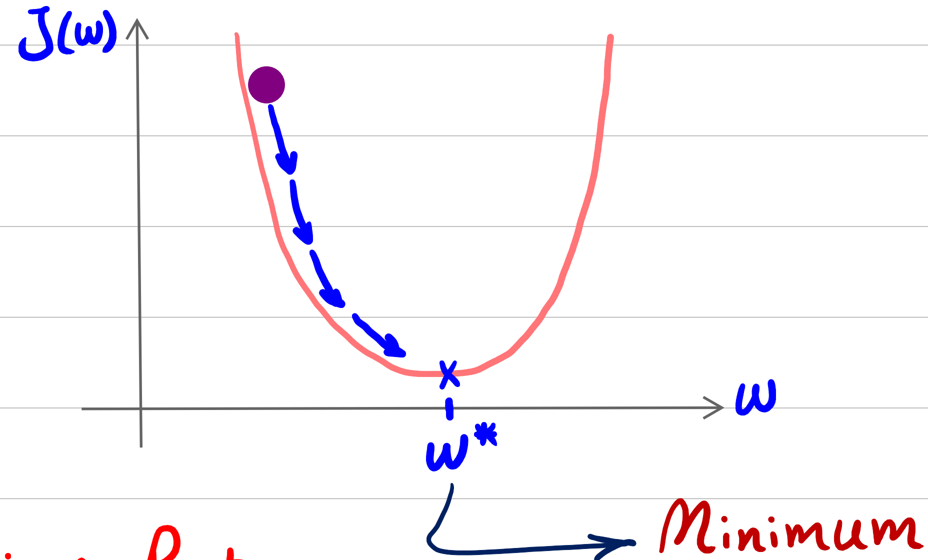
repeat until convergence

{

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

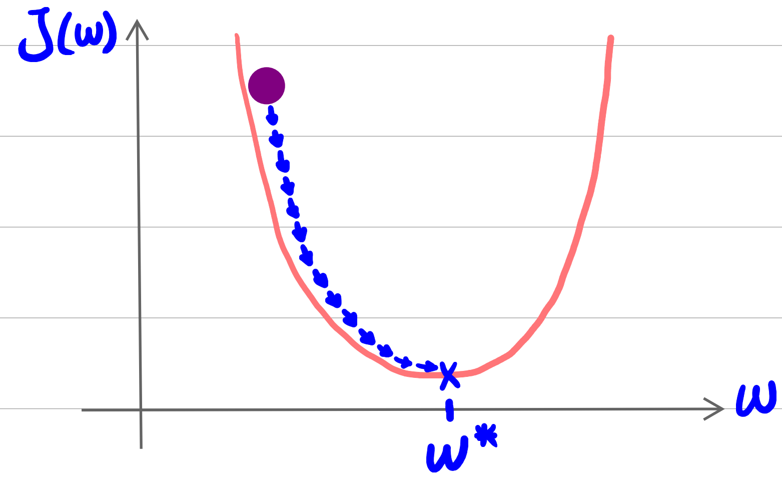
$\quad b = b - \alpha \frac{\partial}{\partial b} J(w,b)$

}

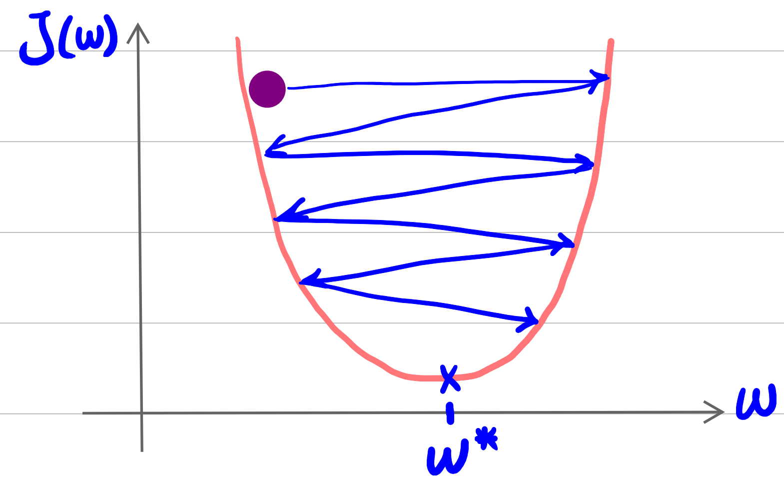


α : Learning Rate

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

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

$b =$ number

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

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

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

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

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

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

vectorization

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

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

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

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

with labels x_0, x_1, x_2 over $10, 20, 30$ and w_0, w_1, w_2 over $1, 2.5, -3.3$.

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

GD:

One feature

repeat

{

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

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

}

n features

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

$X_{3\times 4}$ $\quad y_{3\times 1}$

$$W = \begin{bmatrix} 0.3 \\ 18.5 \\ -53.3 \\ -26.4 \end{bmatrix}_{4\times 1} \qquad b = 785$$

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

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

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

$a_1 =$ $\qquad$ $a_2 =$