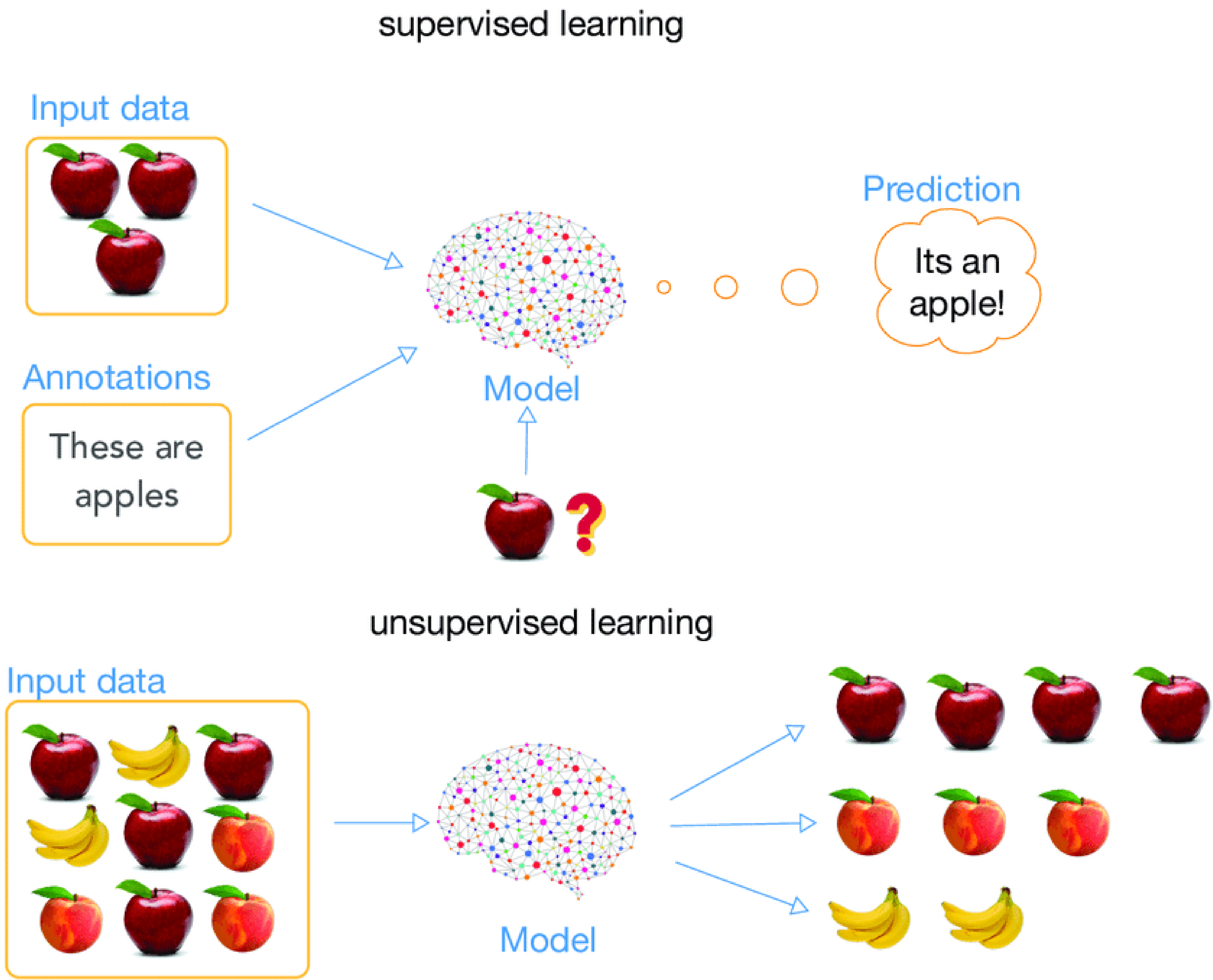
Session 2

Linear Regression & GD

AI2YOU

Virtual Learning


supervised learning
Input data
Annotations
These are apples
Model
Prediction
Its an apple!
unsupervised learning
Input data
Model

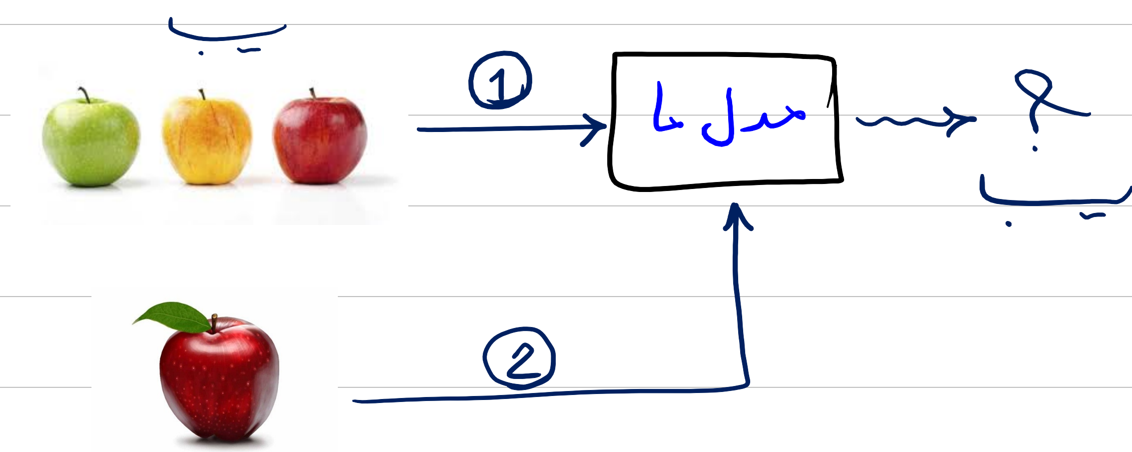
یادگیری ماشین یا Machine Learning

ML چیست؟ یکی از زیر مجموعه های هوش مصنوعی است که به کامپیوتر (سیستم) این امکان را می دهد که به صورت خودکار یاد بگیرد و پیشرفت کند بدون اینکه به برنامه نویسی صریحی نیاز داشته باشد.

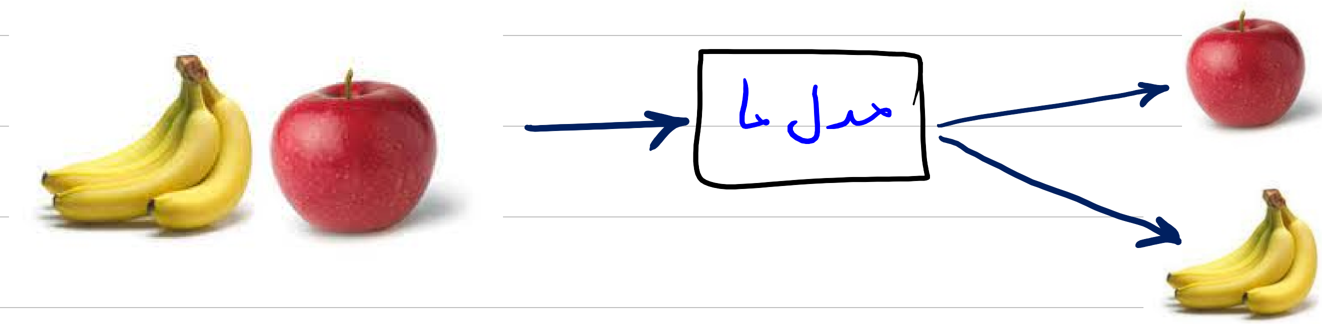
Arthur Samuel

ML

- Supervised — نظارتی ⟵ داده‌های label دارند.



- Unsupervised — غیرنظارتی — داده‌های بدون label



Supervised Learning

$X \longrightarrow y$

input — output label

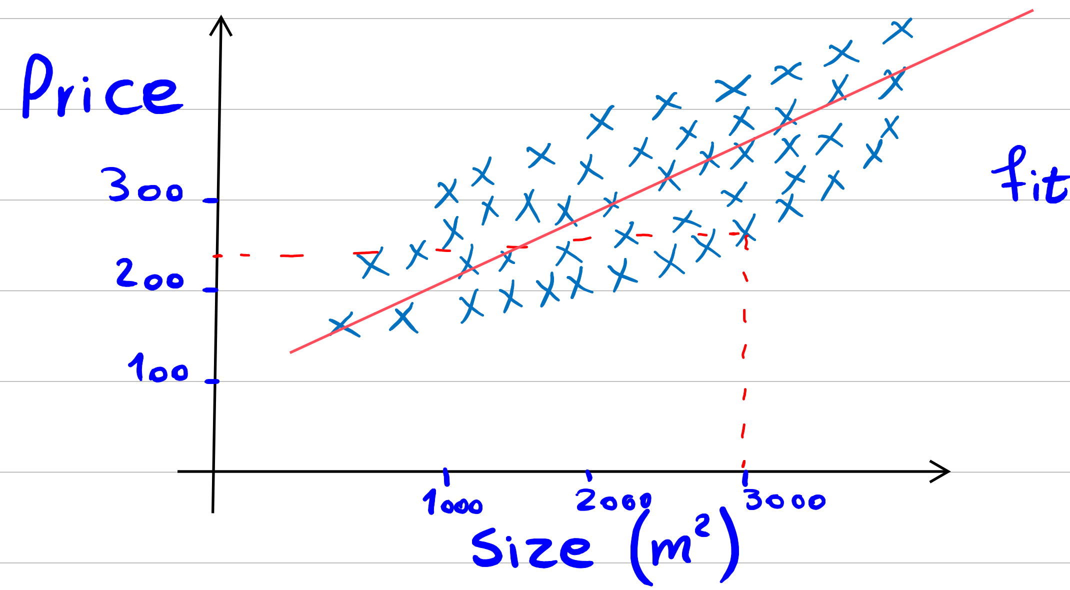
- Email Spam (0,1)
- House Price

Supervised
- Regression
- Classification

Linear Regression ⟶ Housing Price Prediction

می‌خواهیم با توجه به متراژ خانه قیمتش را پیش‌بینی کنیم:

Size (m²)	Price
180	1000
200	1200
300	2000
⋮	⋮



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

Continous

	x (feature)	y (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

$x \longrightarrow \boxed{\text{مدل}} \longrightarrow y$

Notation:

x : Input

y : output

m : # training examples   4

n : # features   1

(x, y)

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

Test Set

X

Size (m^2)
350
2100
833
120
⋮

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

Size of House $\rightarrow \boxed{F} \rightarrow$ Price

؟ F : چیست ؟

$\boxed{\text{Training Set}}$ ⇝ feature, target

↓

$\boxed{\text{learning Algorithm}}$

↓

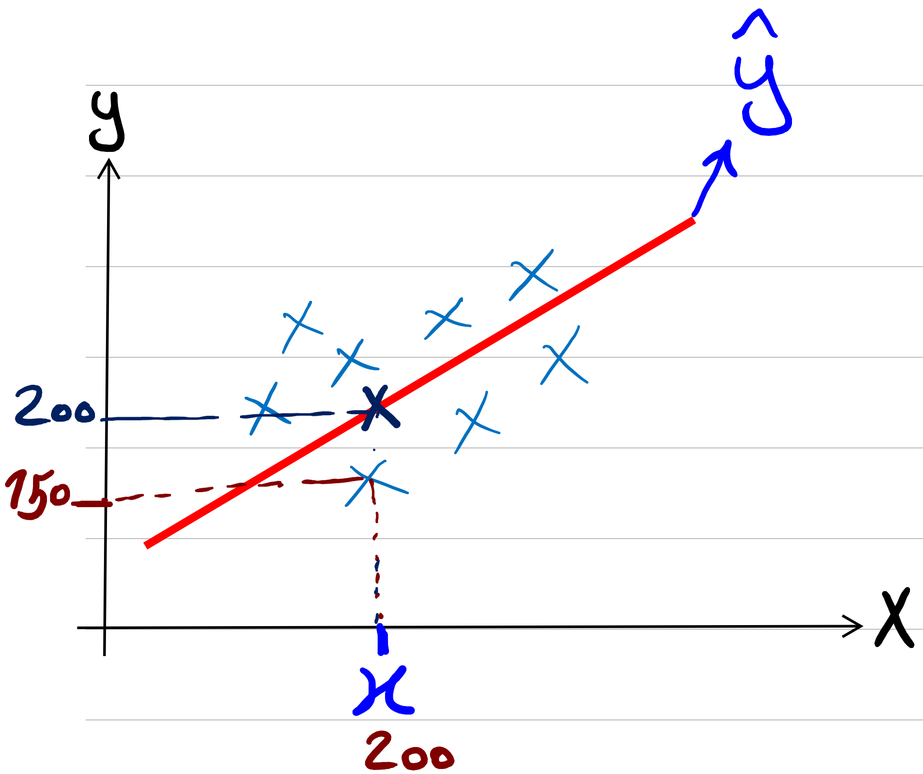
$X \rightarrow \boxed{F} \rightarrow \hat{y}$

X: feature

F: $\underbrace{\text{function/hypothesis}}_{\text{model}}$

$\hat{y}$: $\underbrace{\text{predict (estimated } y\text{)} \quad \text{تخمین } y}_{\text{Target}}$

F در LR چیست؟ ؟

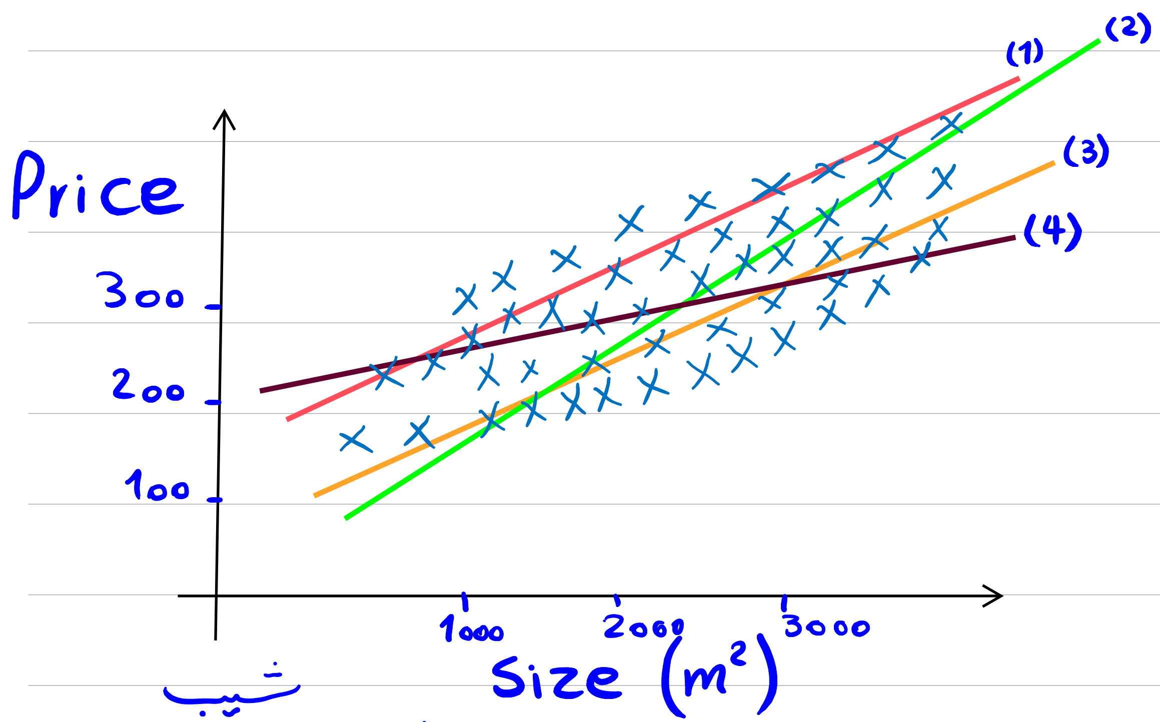


$y = ax + b$ → Line

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

$f(x) = WX + b$

؟ : کدام f بهتر است ؟ از کجا بفهمیم ؟



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

parameters: w, b

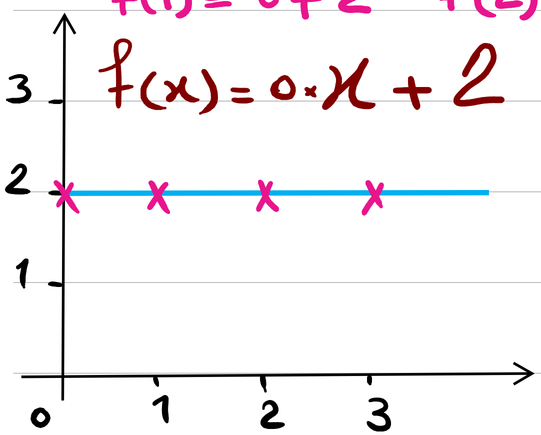
شیب

$f = wx + b$ عرض مبدأ

: Cost Function

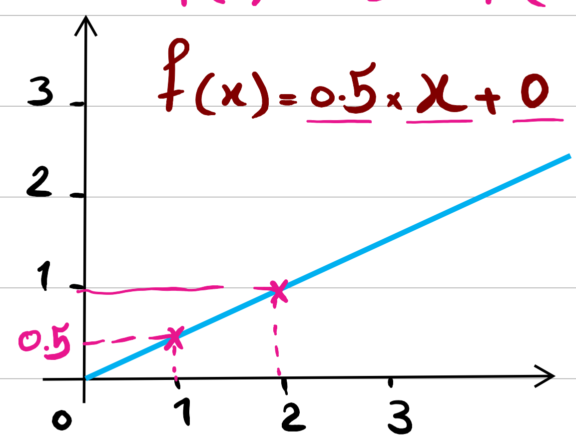
$f(1) = 0 + 2 \quad f(2) = 2$

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



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

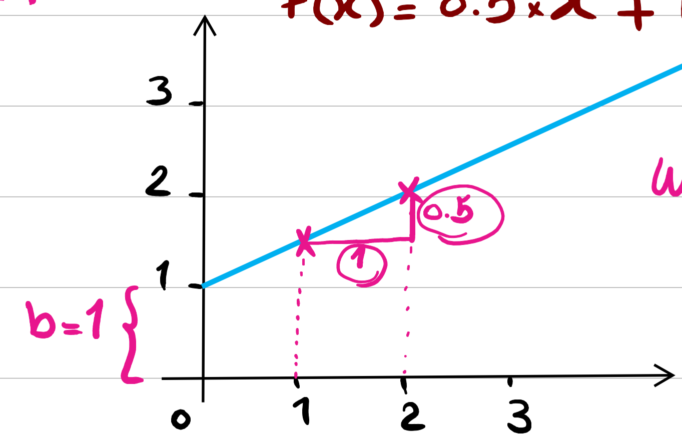
$f(x) = \underline{0.5} \times \underline{x} + \underline{0}$



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

$f(1) = 1,5$

$f(2) = 2$



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

$\frac{2 - 1,5}{2 - 1} = \frac{0.5}{1}$

$W = 0, \quad b = 2$

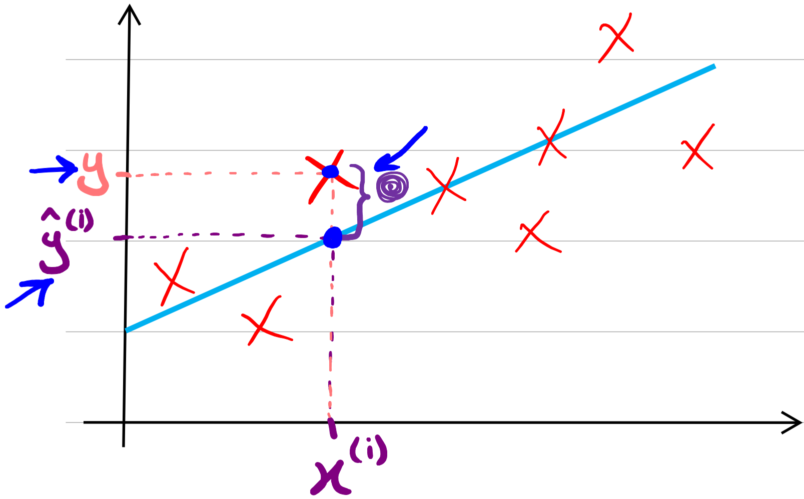
Slope

شیب

$W = 0.5, \quad b = 0$

$W = 0.5, \quad b = 1$

چه 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.

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

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

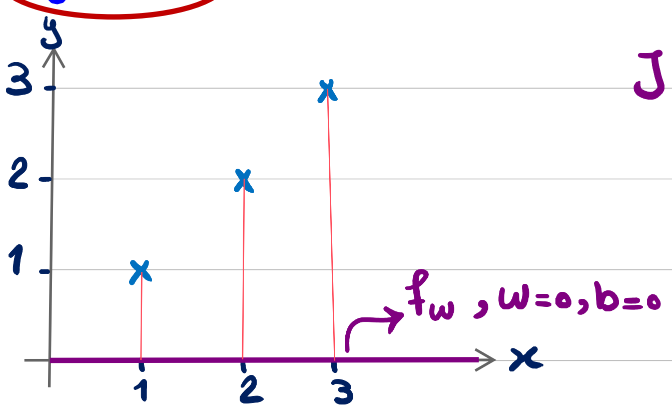
Parameters: w , $b=0$

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

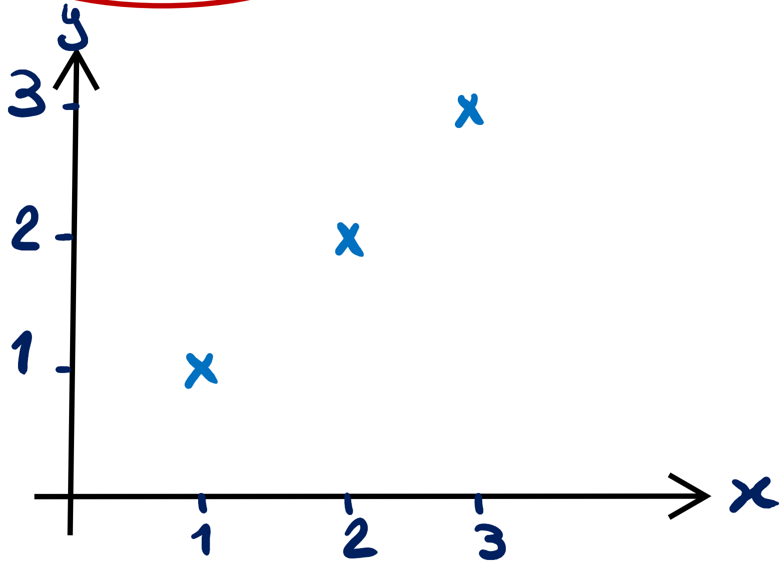
$W=0$



$$J(0)=\frac{1}{2\times 3}\left[\cancelto{1}{(0-1)^2}+\cancelto{4}{(0-2)^2}+\cancelto{9}{(0-3)^2}\right]=\frac{1}{6}\left[1+4+9\right]=\frac{14}{6}=2.3$$

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

$W = 0.5$



x	y
1	1
2	2
3	3

(x, y)

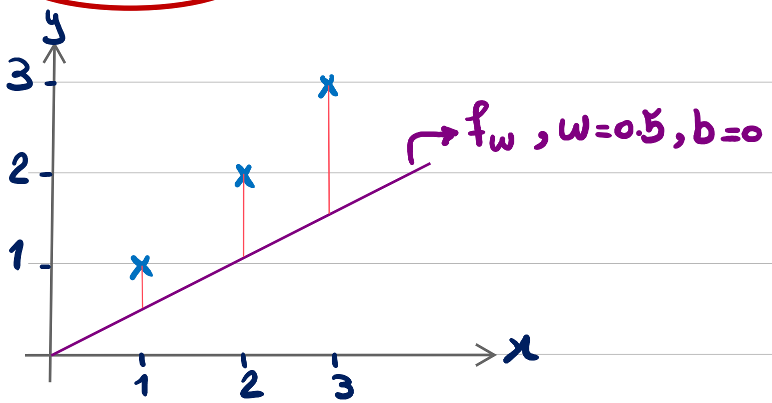
$m = 3$

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

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

$$J(0.5) = ?$$

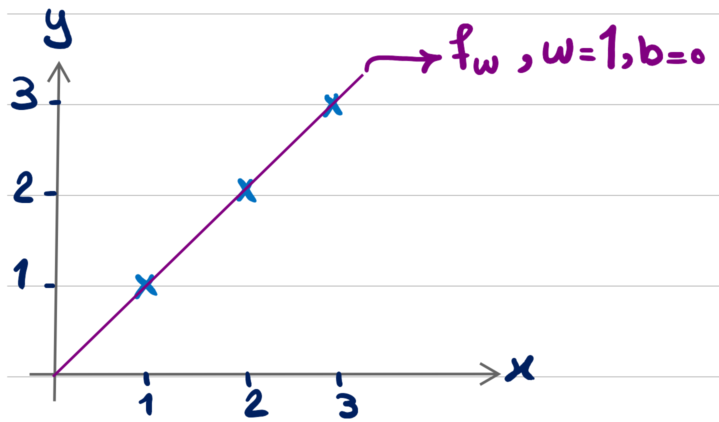
$W=0.5$



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

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

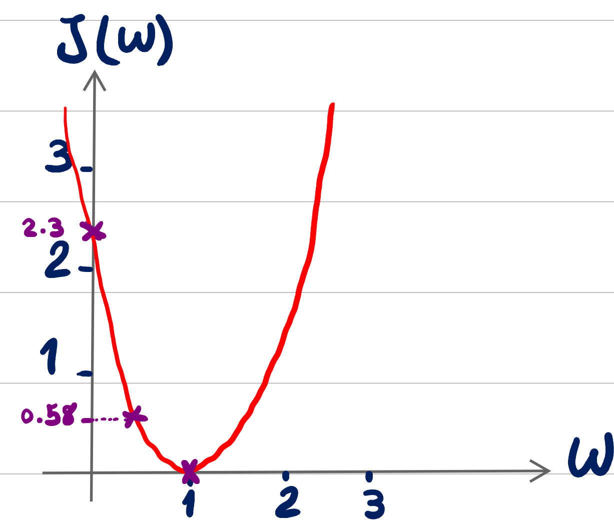
$W=1$



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

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

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



؟: حالا چطور مقدار w و b را پیدا کنیم ؟

$$W = \frac{m(\sum xy) - (\sum x)(\sum y)}{m(\sum x^2) - (\sum x)^2}$$

$$b = \frac{\sum y - w(\sum x)}{m}$$

	x	y	xy	x^2
1	1	3	3	1
2	2	4	8	4
3	3	6	18	9
4	4	8	32	16
$\sum$	10	21	61	30

$m = 4$

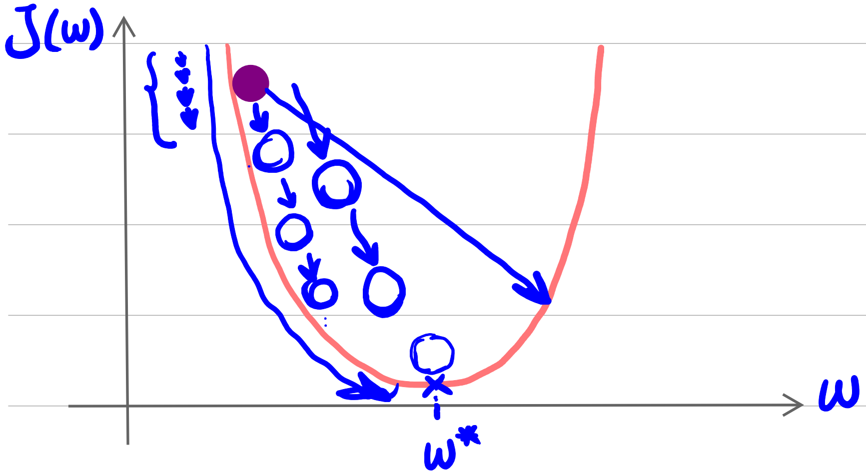
$$w = \frac{m(\sum xy) - (\sum x)(\sum y)}{m(\sum x^2) - (\sum x)^2} = \frac{\overset{244}{\cancel{4 \times 61}} - \overset{210}{\cancel{10 \times 21}}}{\underset{120}{\cancel{4 \times 30}} - \underset{100}{\cancel{10^2}}} = \frac{34}{20} = 1.7$$

$$b = \frac{\sum y - w(\sum x)}{m} = \frac{21 - 1.7 \times 10}{4} = \frac{\overset{4}{\cancel{21 - 17}}}{4} = 1$$

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

function → $J(w,b)$ از GD برای به حداقل رساندن هر تابعی می توان استفاده کرد.

Goal → min $J(w,b)$



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

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

$$f_{w,b}(x) = wx + \overset{0}{\cancel{b}} \xrightarrow{\frac{\partial}{\partial w}} x$$

$$* \quad f^2 = 2f'f$$

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

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

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

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

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

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 به دست آورید.

$(\alpha = 0.1,\ b = 1,\ w = 0)$

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

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

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

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

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

Multiple Linear Regression:

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

X_1 Size(m^2)	X_2 #bedrooms	X_3 #floors	X_4 Age	y 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
$f_{w,b}(x) = wx + b$	$f_{w,b}(x) = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n + b$
	$\vec{w} = [w_1 \quad w_2 \quad w_3 \quad \cdots \quad w_n]_{1\times n}$ $b =$ number
	$\vec{x} = [x_1 \quad x_2 \quad x_3 \quad \cdots \quad x_n]_{1\times n}$
	$f_{\vec{w},b}(\vec{x}) = \vec{w} \cdot \vec{x} + b =$
	dot product
	$w_1 x_1 + w_2 x_2 + \cdots + w_n x_n + b$

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

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

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

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

vectorization

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

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

$$x = \begin{array}{ccc} x_0 & x_1 & x_2 \\ \end{array} \quad w = \begin{array}{ccc} w_0 & w_1 & w_2 \end{array}$$

$$x = [10, 20, 30] \qquad w = [1, 2.5, -3.3] \qquad b = 4$$

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

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

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

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

$$f_{w,b}(x) = W_1X_1 + W_2X_2 + \dots + W_nX_n + b$$

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

$$\begin{bmatrix} 2104 & 5 & 1 & 45 \\ 1416 & 3 & 2 & 40 \\ 852 & 2 & 1 & 35 \end{bmatrix}_{3\times4} \begin{bmatrix} 0.3 \\ 18.5 \\ -53.3 \\ -26.4 \end{bmatrix}_{4\times1} = \begin{bmatrix} -517.6 \\ -682.3 \\ -684.7 \end{bmatrix}_{3\times1}$$

$4\times1 \quad 3\times4$

$$a_{00} = [(2104\times0.3) + (5\times18.5) + (1\times(-53.3)) + (45\times(-26.4))] = -517.6$$