Session 9

Gradient Descent (GD)

Math | Hobot | Zahra Amini

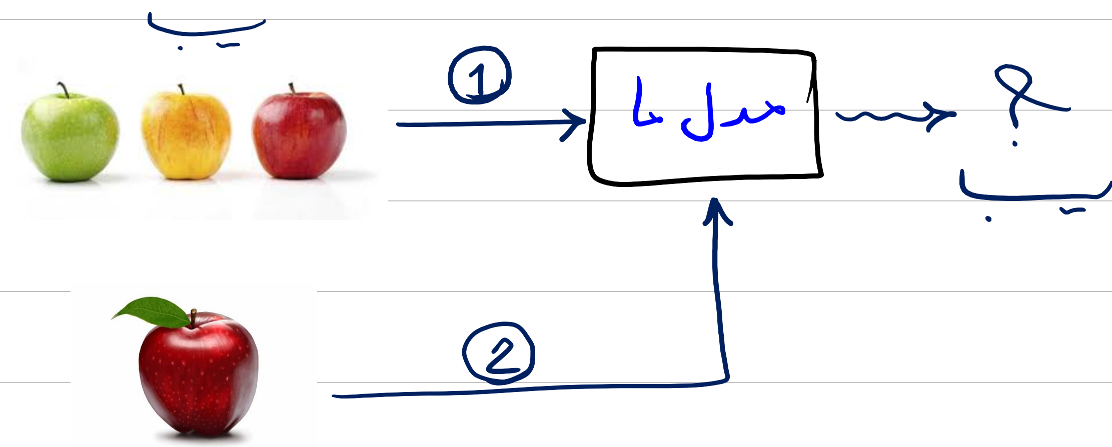
Telegram: @zahraamini_ai & Instagram:@zahraamini_ai & LinkedIn: @zahraamini-ai

https://zil.ink/zahraamini

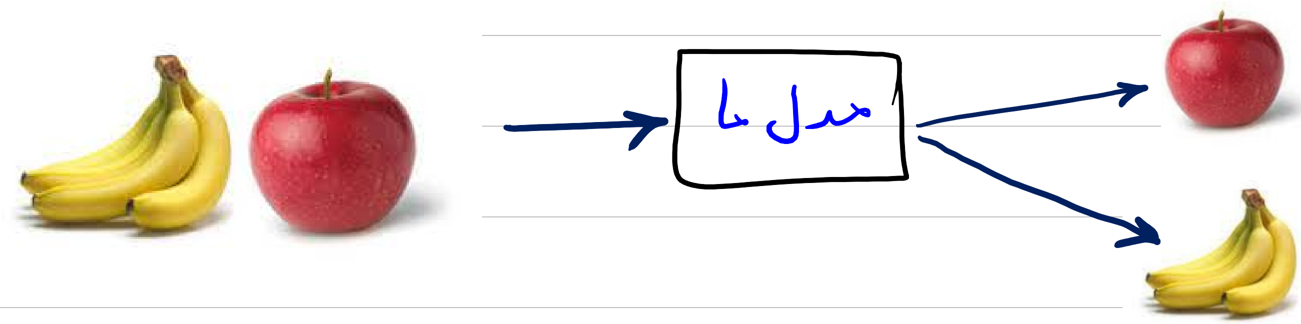
یادگیری ماشین یا Machine Learning

ML چیست؟ یکی از زیرمجموعه 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.



ML → Unsupervised غیرنظارتی داده‌های بدون label



Supervised Learning

$X \rightarrow y$

input — output label

Email Spam (0,1)

House Price

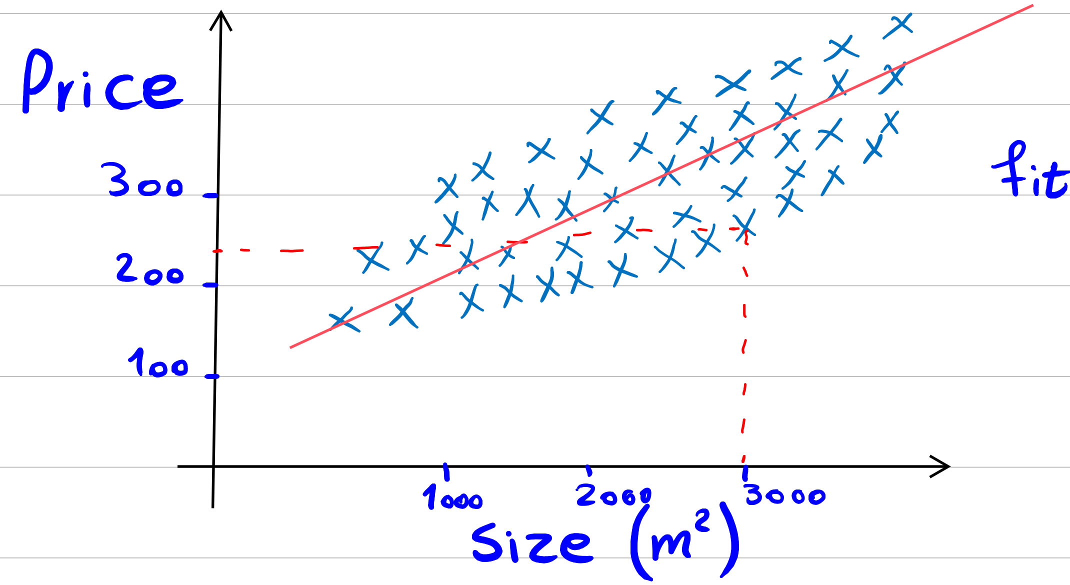
Supervised → Regression

Supervised → Classification

Linear Regression → Housing Price Prediction

میخواهیم با توجه به متراژ خانه قیمتش را پیش بینی کنیم:

Size (m^2)	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
⋮	⋮	⋮

Training Set

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

Notation:

x : Input

y : output

m : # training examples 4

n : # features 1

(x, y)

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

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

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

Test Set

X

Size (m^2)
350
2100
833
120
⋮

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

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

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

↓

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

↓

$x \longrightarrow \boxed{F} \longrightarrow \hat{y}$

x ⟶ feature

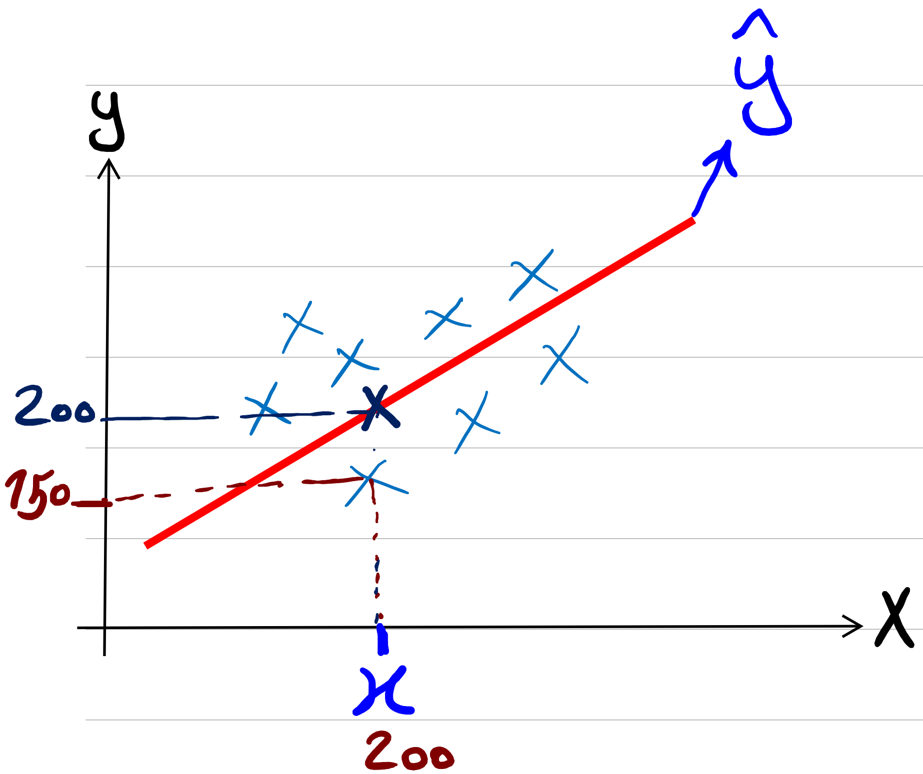
$\underbrace{\text{function/hypothesis}}_{\text{model}}$

$\hat{y}$ ⟶ predict (estimated y)

تخمین y

Target

F: چیست؟

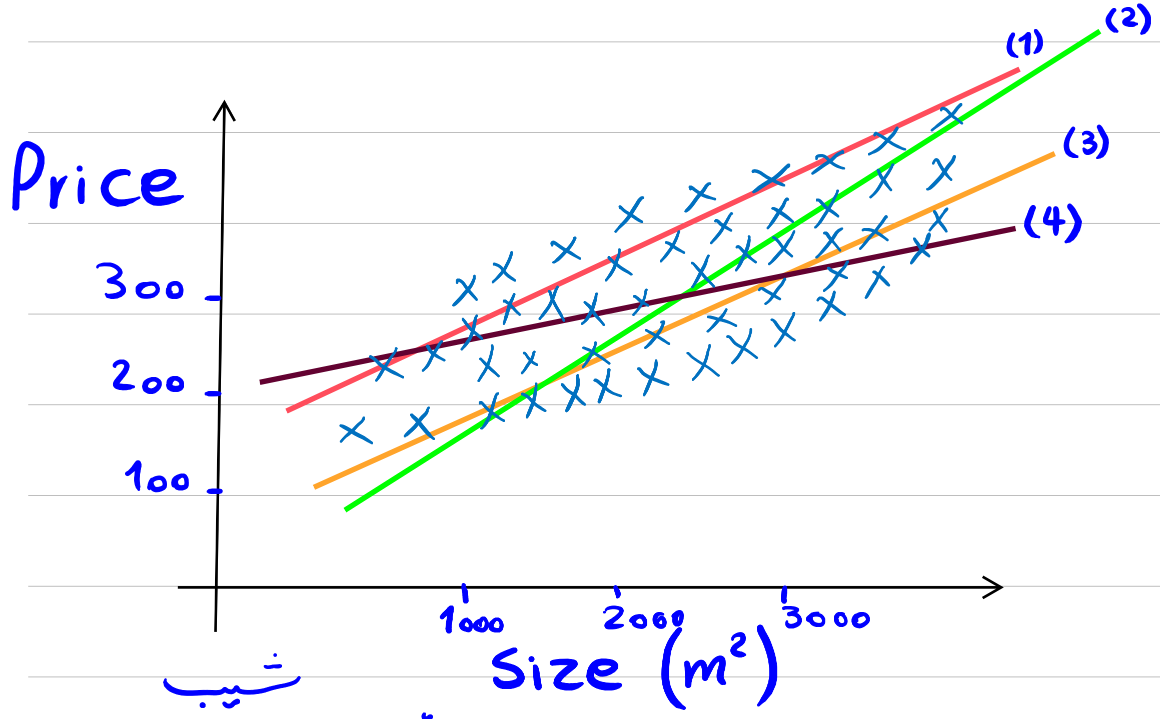


$y = ax + b$ ⟶ Line

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

$f(x) = WX + b$

؟: کدام f بهتر است؟ از کجا بفهمیم؟



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

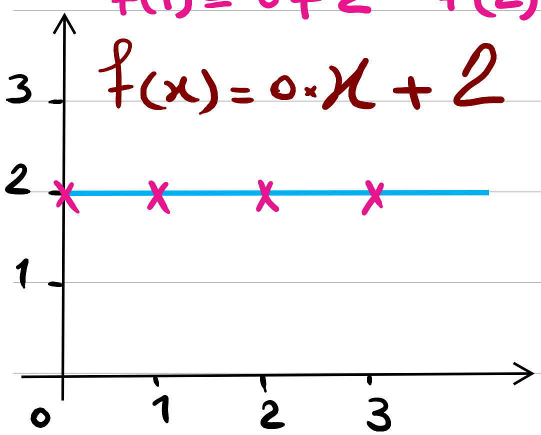
parameters: w, b

: Cost Function

$f = wx + b$ — شیب (w)، عرض از مبدأ (b)

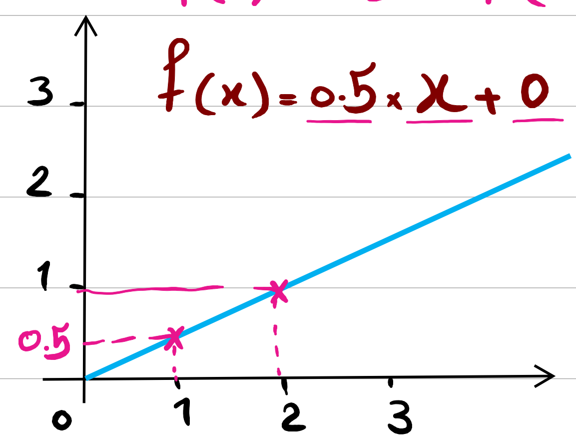
$f(1) = 0 + 2$ $\quad f(2) = 2$

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



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

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

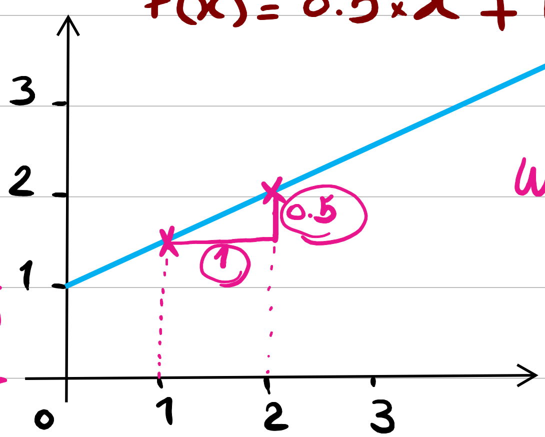


$f(x) = 0.5 \times x + 1$

$f(1) = 1.5$

$f(2) = 2$

$b = 1$



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

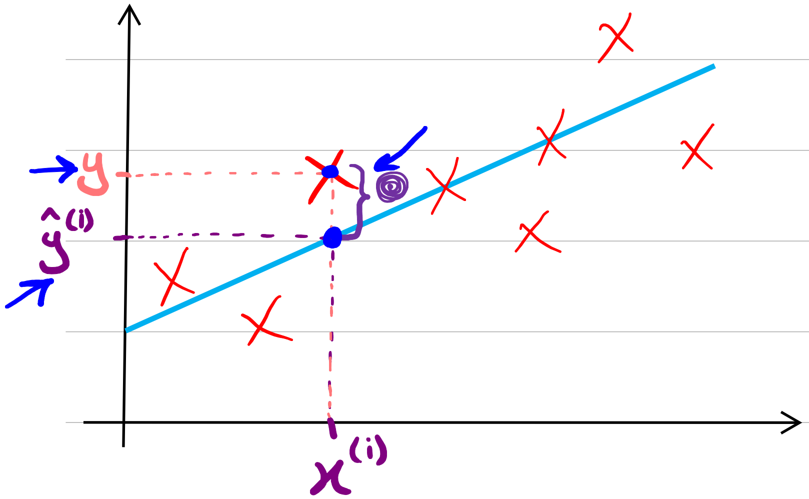
$\frac{2 - 1.5}{2 - 1} = \frac{0.5}{1}$

$W = 0, \quad b = 2$ $\qquad W = 0.5, \quad b = 0$ $\qquad W = 0.5, \quad b = 1$

→ Slope

شیب

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



⊚ خطا ← error

چه w و b خوب هست؟

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

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

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

m: # training examples

یکی از معروف ترین 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 $\longrightarrow f_{w,b}(x)$

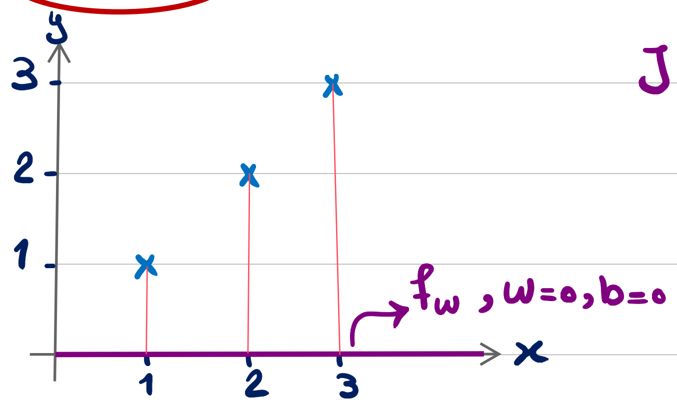
Parameters: w , $b=0$

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

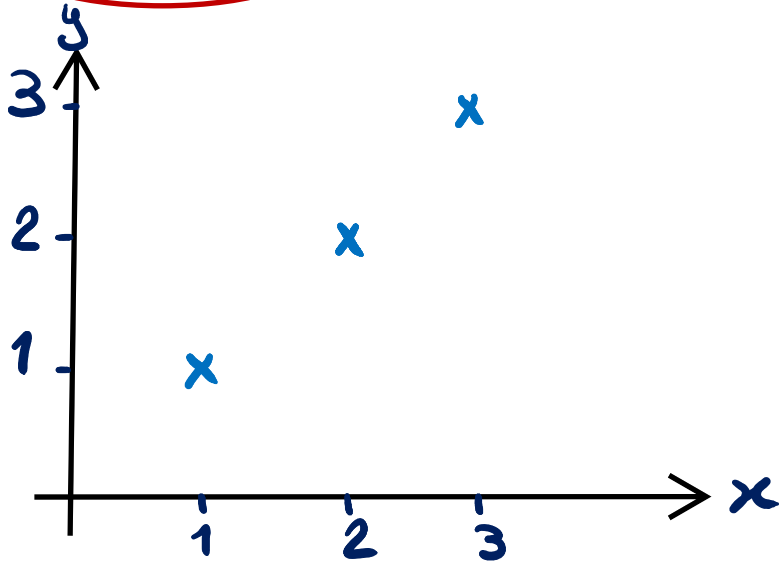
$w=0$



$$J(0)=\frac{1}{2\times3}\left[\overset{1}{\cancel{(0-1)^2}}+\overset{4}{\cancel{(0-2)^2}}+\overset{9}{\cancel{(0-3)^2}}\right]=\frac{1}{6}\left[1+4+9\right]=\frac{13}{6}=2.3$$

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

$W = 0.5$



x	y
1	1
2	2
3	3

(x, y)

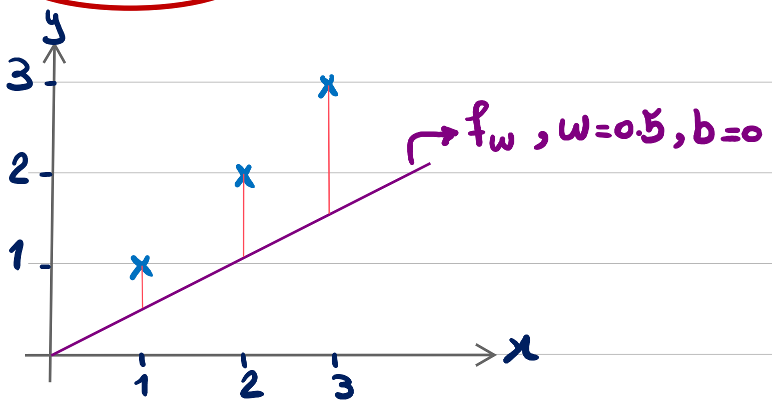
$m = 3$

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

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

$$J(0.5) = ?$$

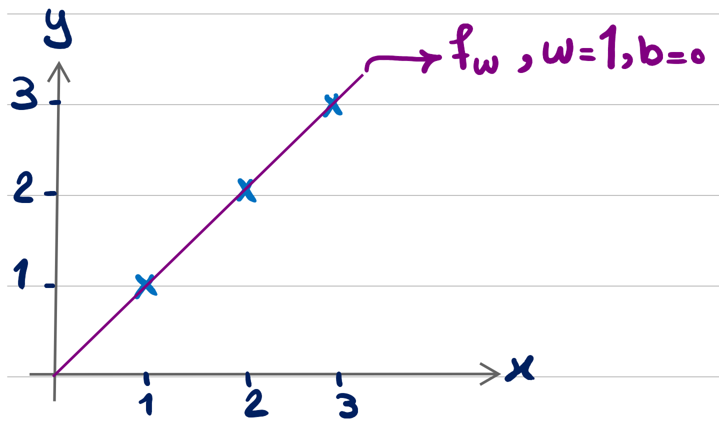
$w=0.5$



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

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

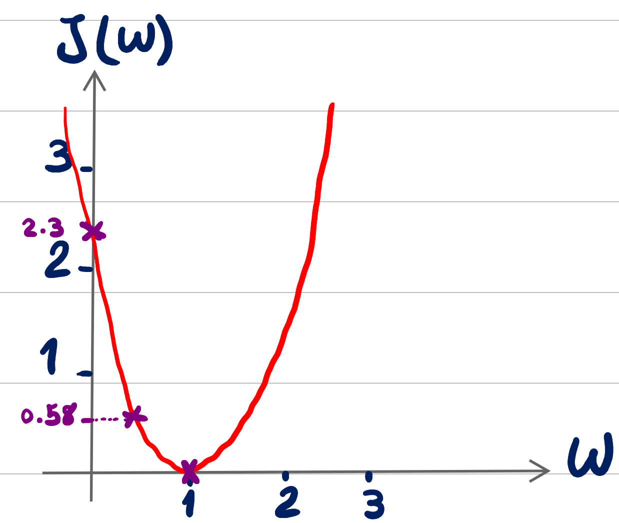
$w=1$



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

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

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

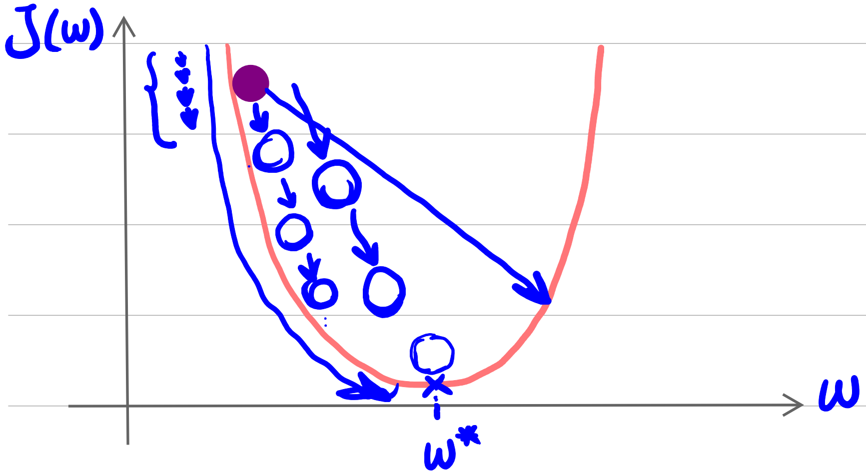


خب حالا می خواهیم بریم سراغ الگوریتمی که تابع J را به حداقل برساند. ← Gradient Descent (GD)

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

Function → $J(w,b)$

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 بعد از 1 بار آپدیت به دست آورید.

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

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

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

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

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

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