Linear Regression

Regression

It is the statistical technique that relates a dependant variable to one or more independant variables

Notation:

x = "input" variable
feature

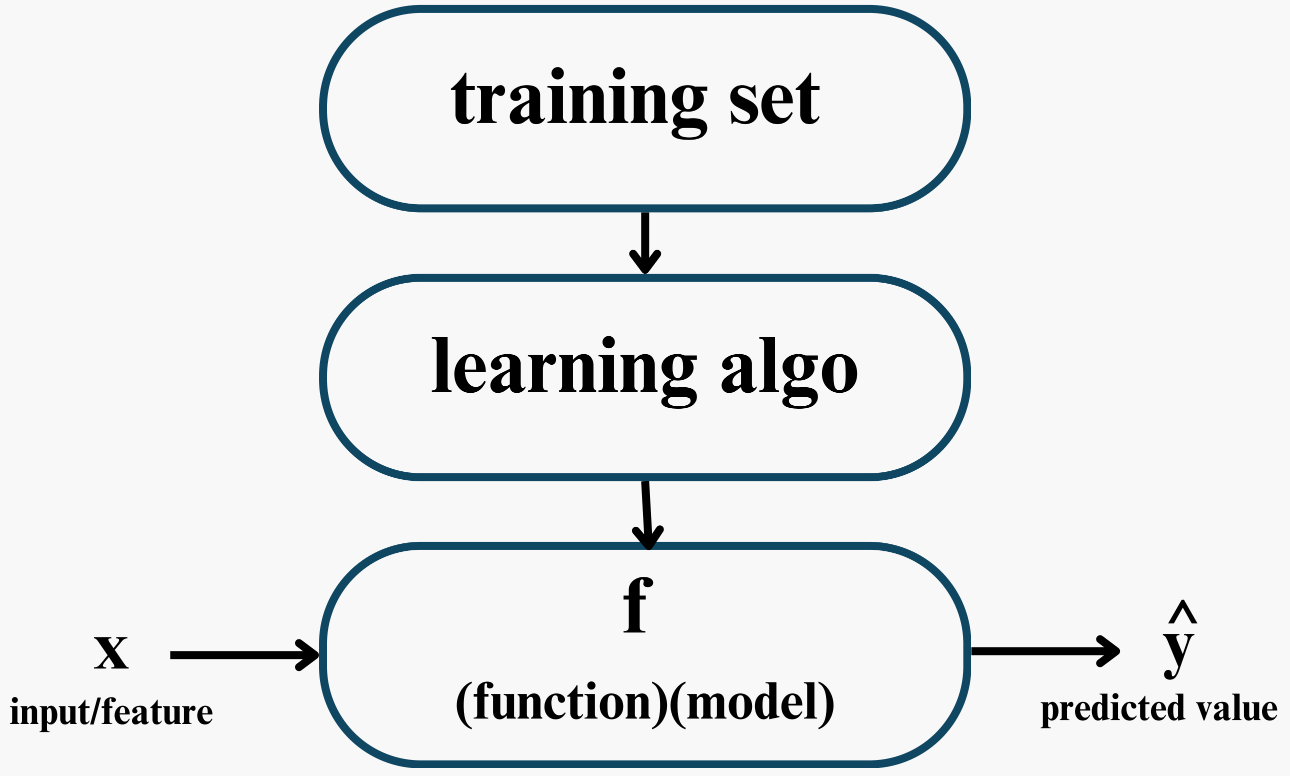
y = "output" variable
"target" variable

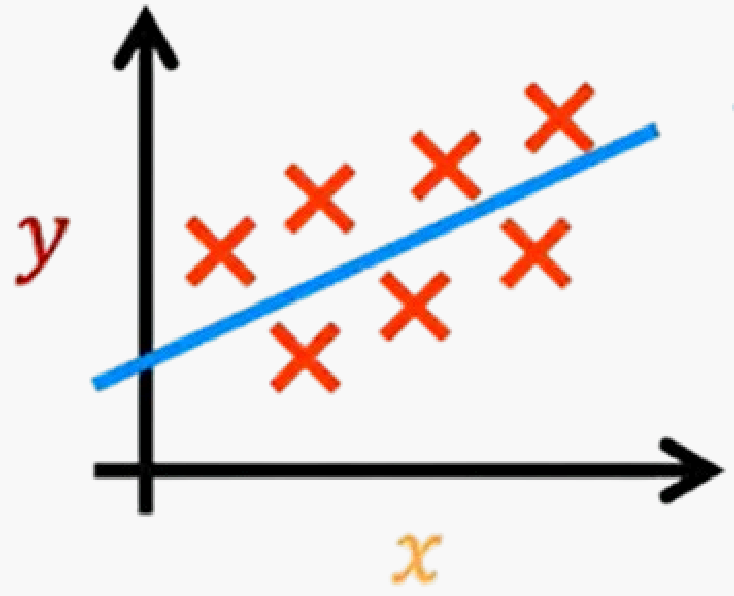
m = number of training examples

(x, y) = single training example

$(x^{(i)}, y^{(i)})$

$(x^{(i)}, y^{(i)})$ = i^{th} training example
(1^{st}, 2^{nd}, 3^{rd} ...)

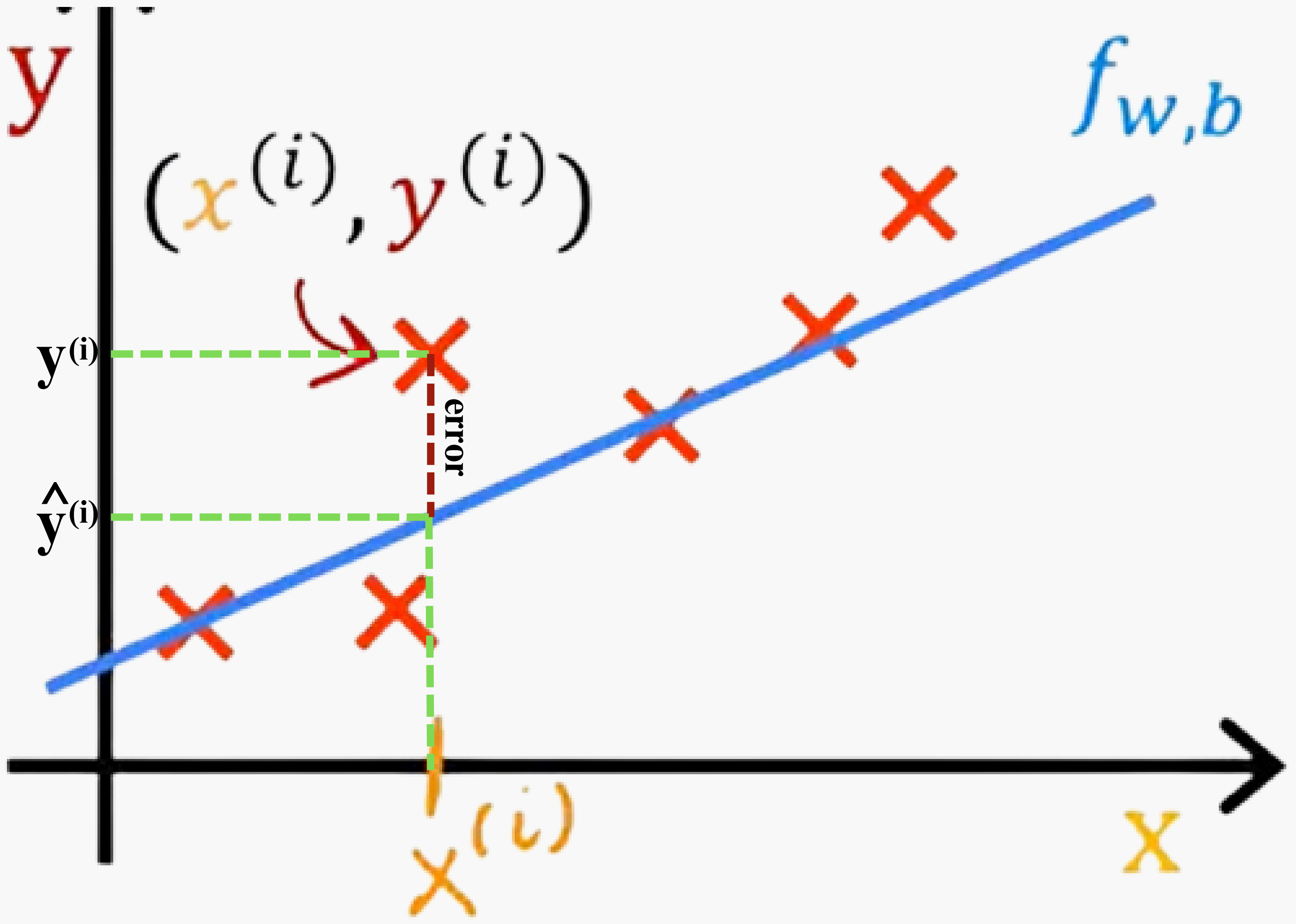




$f_{w,b}(x) = wx + b$

w and b : parameters
coefficients
weights

Cost Function



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

error -> $\hat{y}^{(i)} - y^{(i)}$

squared error cost function:

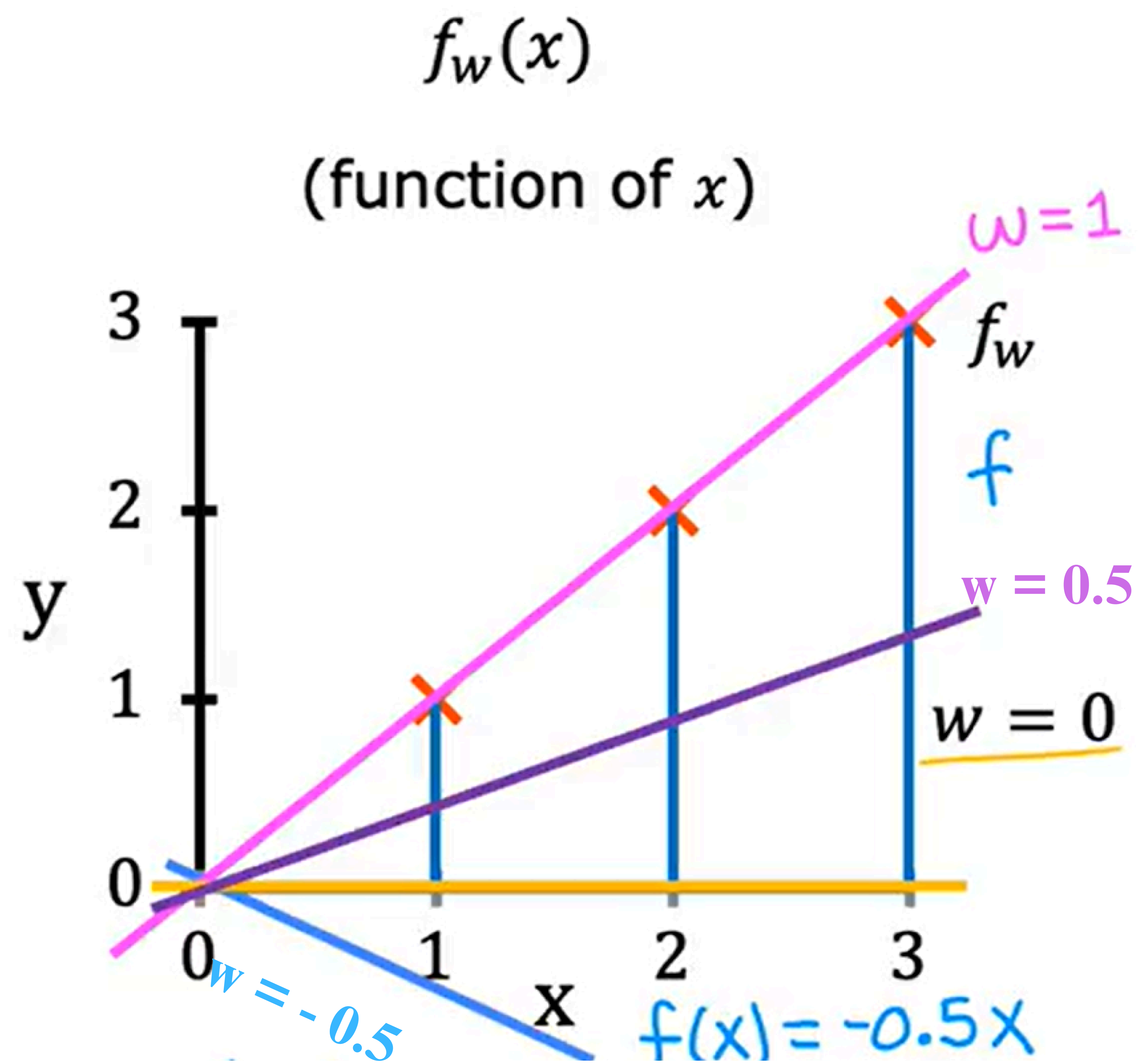
$$J(w,b) = \frac{1}{2m} \sum_{i=1}^{m} \Big(\underbrace{\hat{y}^{(i)} - y^{(i)}}_{\text{error}} \Big)^2$$

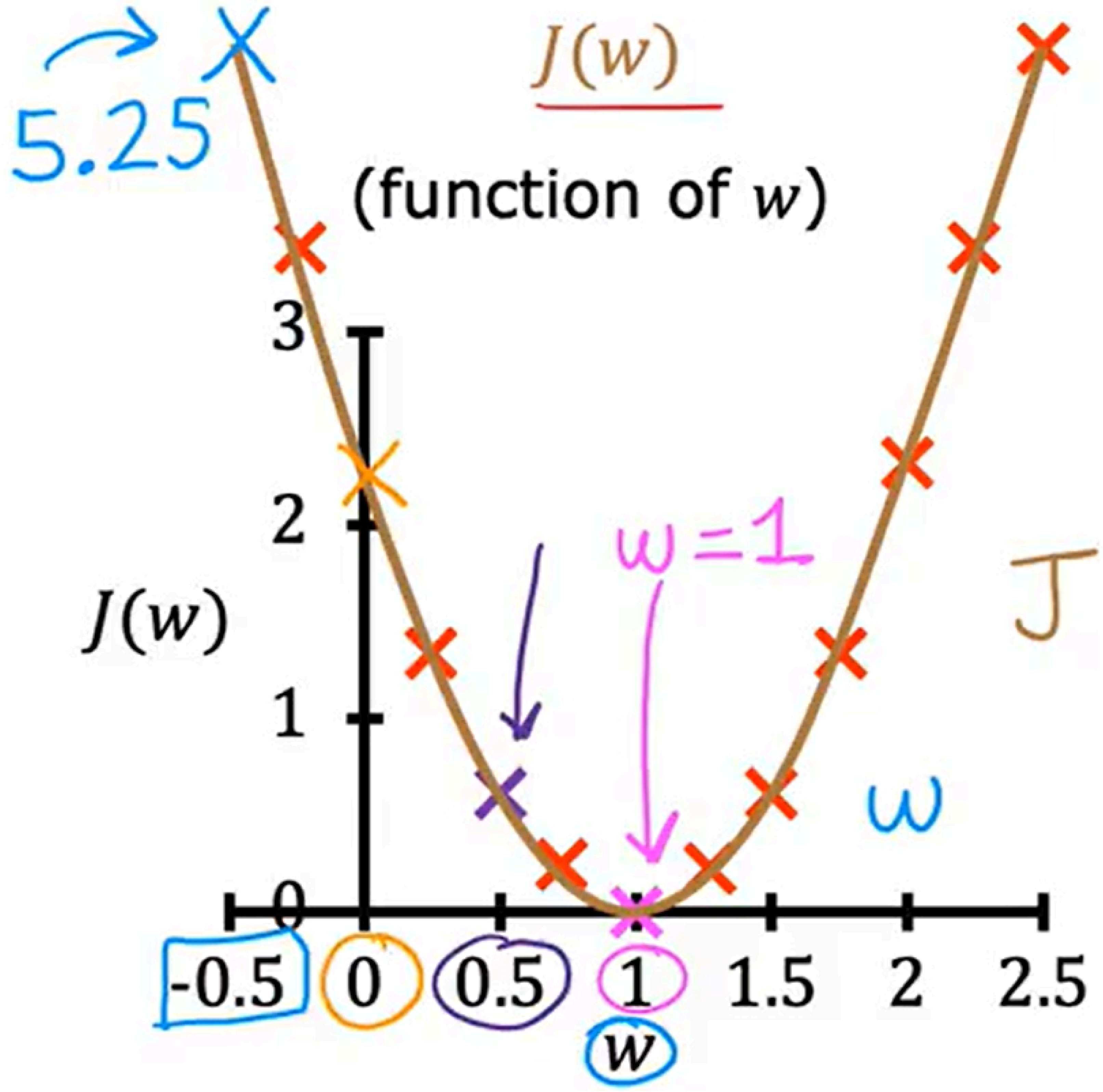
m = number of training examples

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

goal is to minimize J(w,b) hence minimising the error

Cost Function Intuition

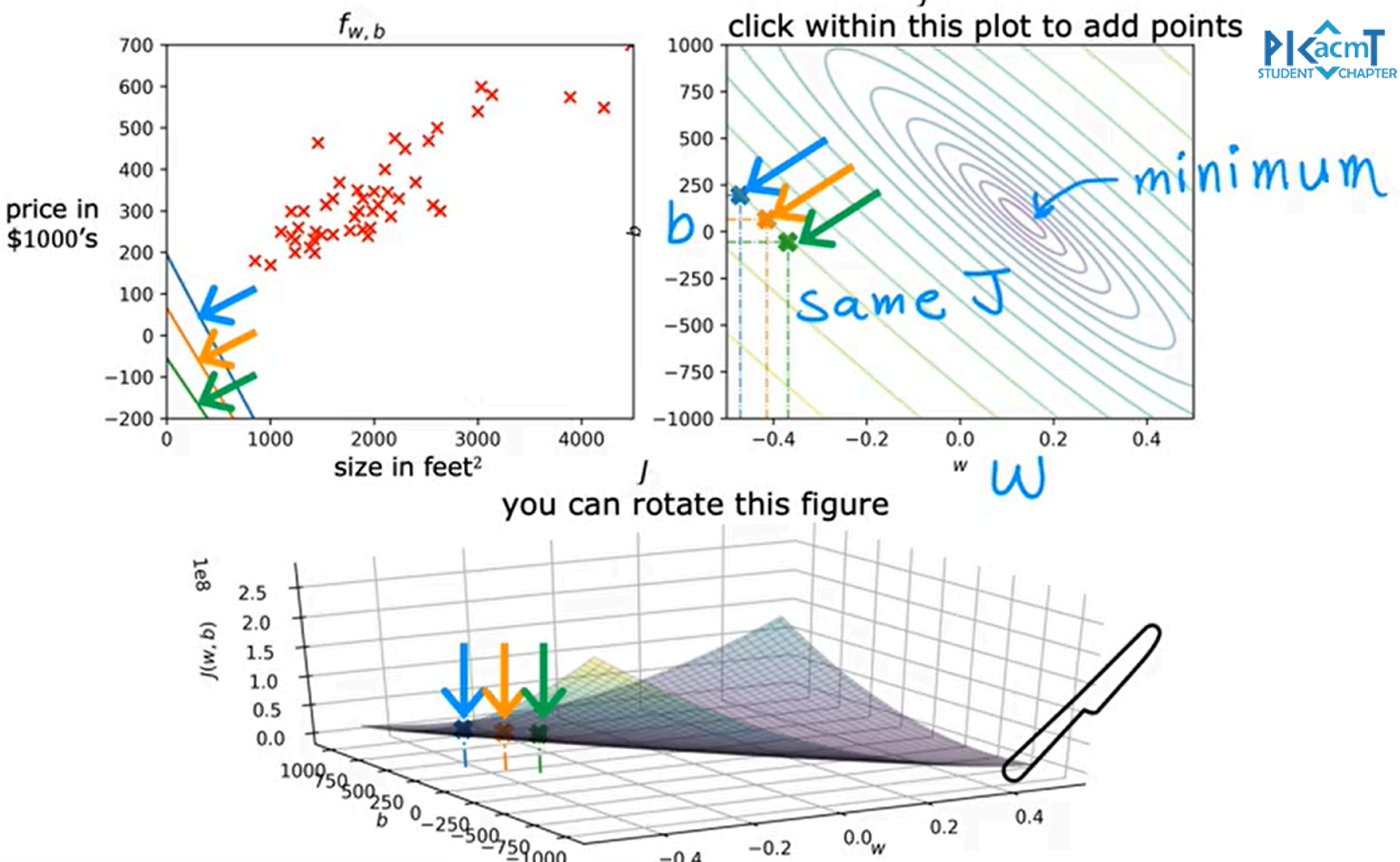




simplified: b=0

f(x) = wx

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



$f_{w,b}$
price in $1000's
size in feet²
0
1000
2000
3000
4000
700
600
500
400
300
200
100
0
−100
−200
click within this plot to add points
1000
750
500
250
0
−250
−500
−750
−1000
−0.4
−0.2
0.0
0.2
0.4
b
w
minimum
same J
J
you can rotate this figure
1e8
$J(w,b)$
2.5
2.0
1.5
1.0
0.5
0.0
1000
750
500
250
0
−250
−500
−750
−1000
−0.4
−0.2
0.0
0.2
0.4

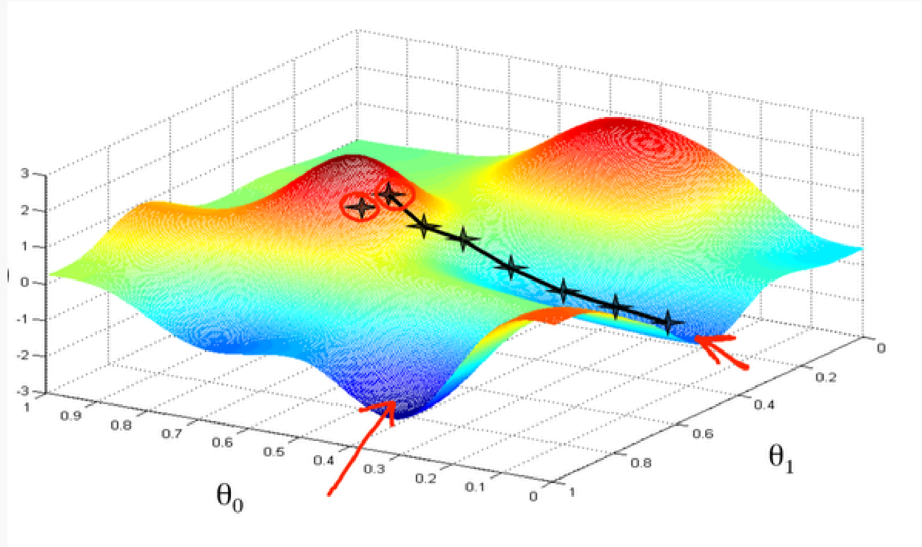
Gradient Descent

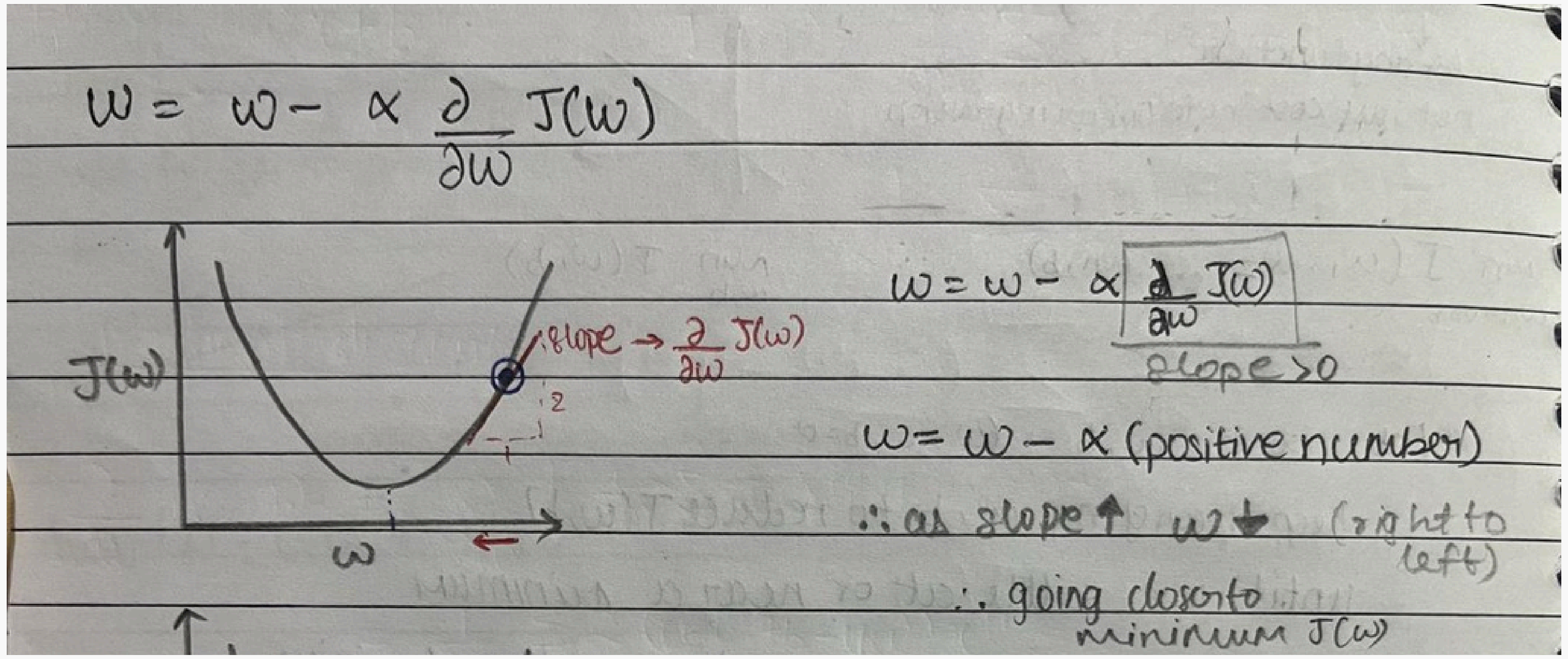
$$\begin{cases} w = w - \alpha \boxed{\frac{\partial}{\partial w} J(w,b)} \\ b = b - \alpha \frac{\partial}{\partial b} J(w,b) \end{cases}$$

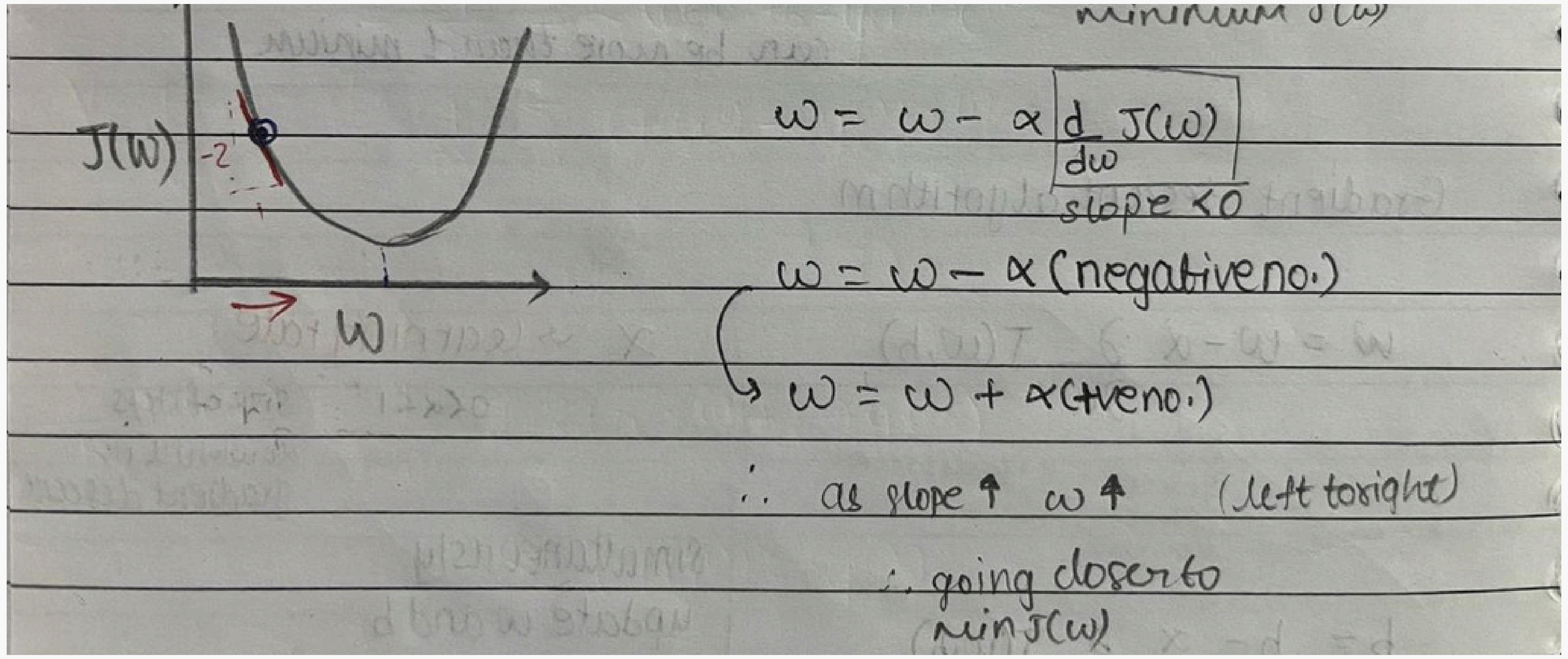
Learning rate 0<alpha<1 number of steps downhill

Derivative gives direction

Simultaneously update w and b







if alpha is too small :

gradient descent may be too slow

if alpha is too large:

gradient descent may never reach min & overshoot

Gradient Descent Implementation

Linear regression model

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

Cost function

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

Gradient descent algorithm

repeat until convergence {

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

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

}

Multi Variable Linear Regression

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

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

parameters of the model

b is a number

vector $\vec{x} = [x_1 \ x_2 \ x_3 \ldots x_n]$

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

dot product

multiple linear regression

	Previous notation	Vector notation
Parameters	$w_1, \cdots, w_n$ b	vector of length n $\vec{\mathrm{w}} = [w_1 \quad \cdots \quad w_n]$ b still a number
Model	$f_{\vec{\mathrm{w}},b}(\vec{\mathrm{x}}) = w_1x_1 + \cdots + w_nx_n + b$	$f_{\vec{\mathrm{w}},b}(\vec{\mathrm{x}}) = \vec{\mathrm{w}} \cdot \vec{\mathrm{x}} + b$ dot product
Cost function	$J(\underbrace{w_1, \cdots, w_n}, b)$	$J(\vec{\mathrm{w}}, b)$

Gradient descent

Previous notation	Vector notation
repeat $\{$ $w_j = w_j - \alpha \frac{\partial}{\partial w_j} J(\underbrace{w_1, \cdots, w_n}, b)$ $b = b - \alpha \frac{\partial}{\partial b} J(\underbrace{w_1, \cdots, w_n}, b)$ $\}$	repeat $\{$ $w_j = w_j - \alpha \frac{\partial}{\partial w_j} J(\vec{\mathrm{w}}, b)$ $b = b - \alpha \frac{\partial}{\partial b} J(\vec{\mathrm{w}}, b)$ $\}$

Gradient Descent

One feature

repeat {

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

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

simultaneously update w, b

}

n features $(n \geq 2)$

repeat {

$j=1$

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

$\vdots$

$j=n$

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

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

simultaneously update w_j (for $j = 1, \cdots, n$) and b

}