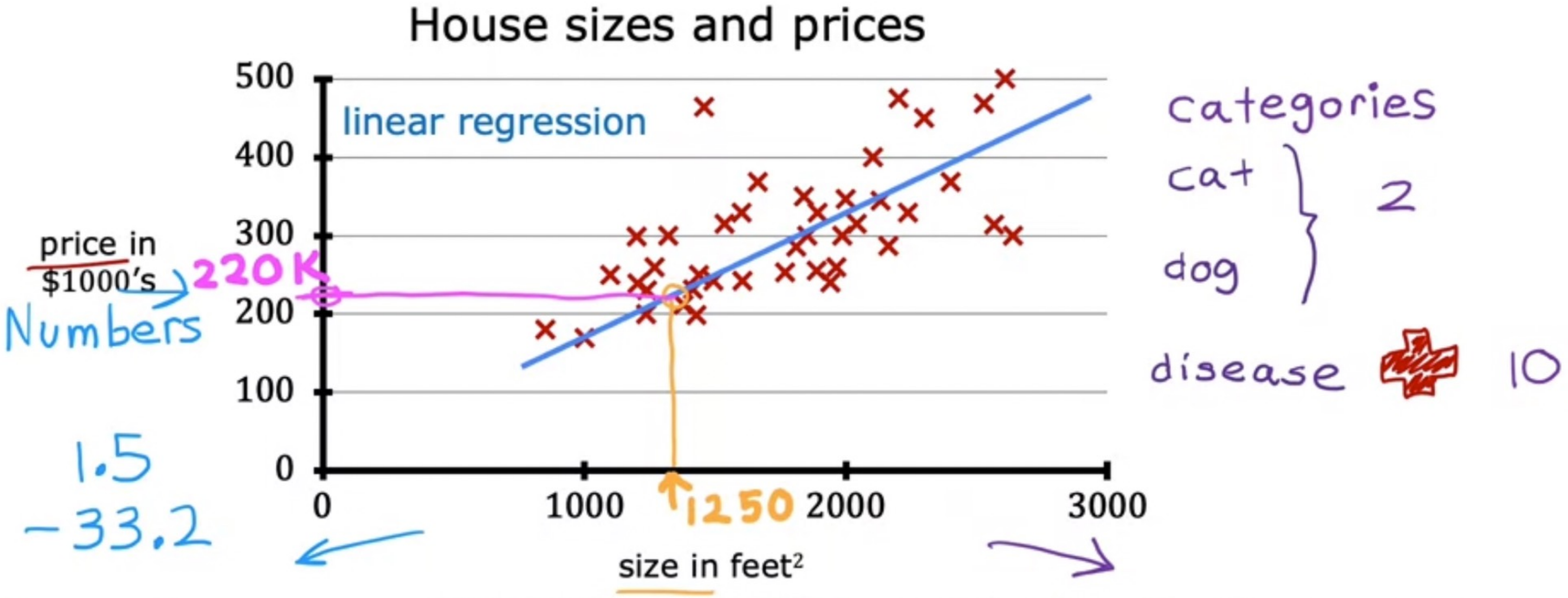
ML – Regression and Cost Function

draft



Regression model
Predicts numbers

Supervised learning model
Data has "right answers"

Classification model
Predicts categories
Small number of possible outputs

House sizes and prices

Data table

size in feet²	price in $1000's
2104	400
1416	232
1534	315
852	178
...	...
3210	870

Terminology

Training set: Data used to train the model

	x size in feet^2	y price in \$1000's
(1)	2104	400
(2)	1416	232
(3)	1534	315
(4)	852	178
...	...	...
(47)	3210	870

$m = 47$

$x^{(1)} = 2104$ $\quad y^{(1)} = 400$

$(x^{(1)}, y^{(1)}) = (2104, 400)$

$x^{(2)} = 1416$ $\quad x^{(2)} \neq x^2$ not exponent

Notation:

x = "input" variable
feature

y = "output" variable
"target" variable

m = number of training examples

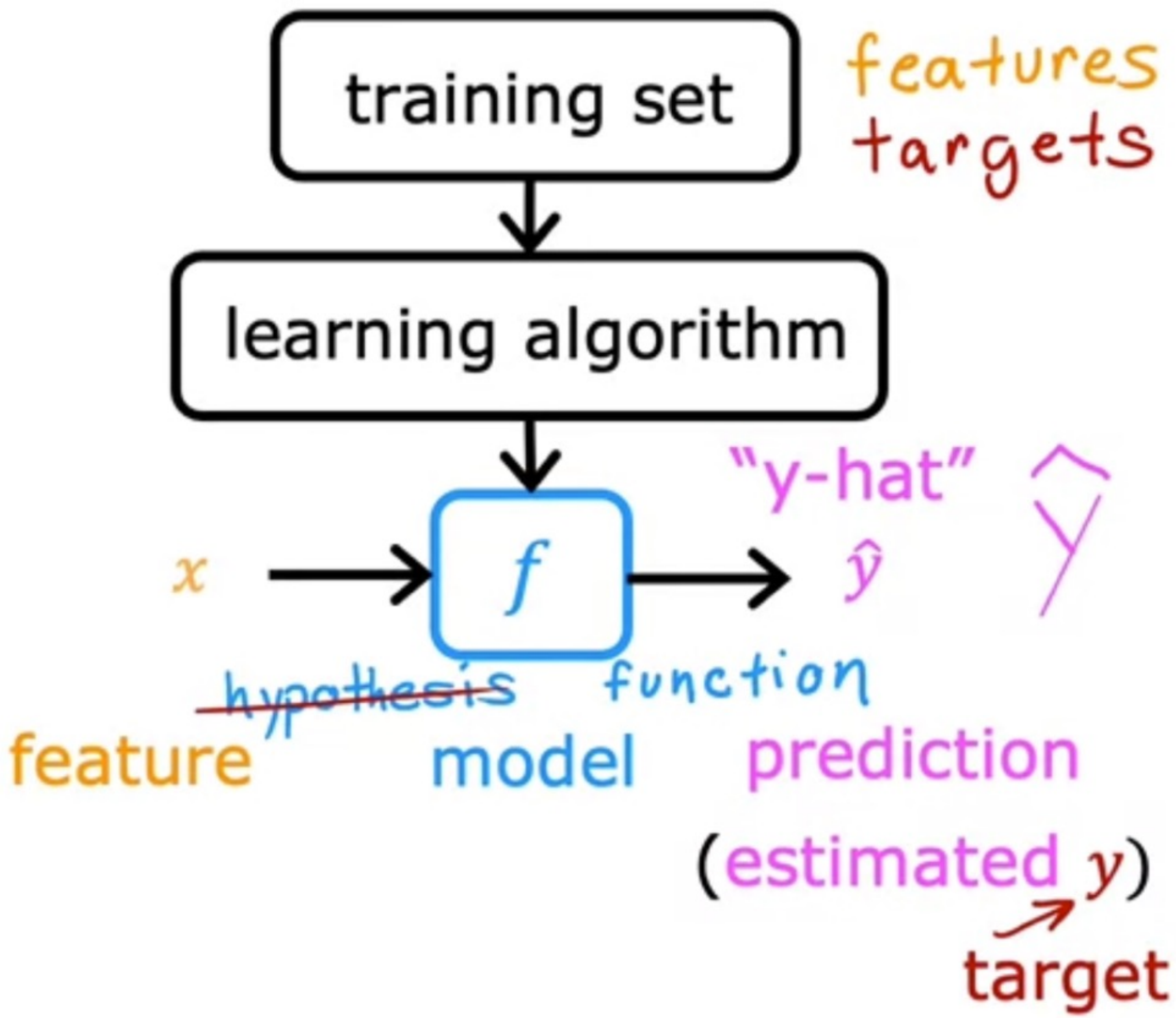
(x, y) = single training example

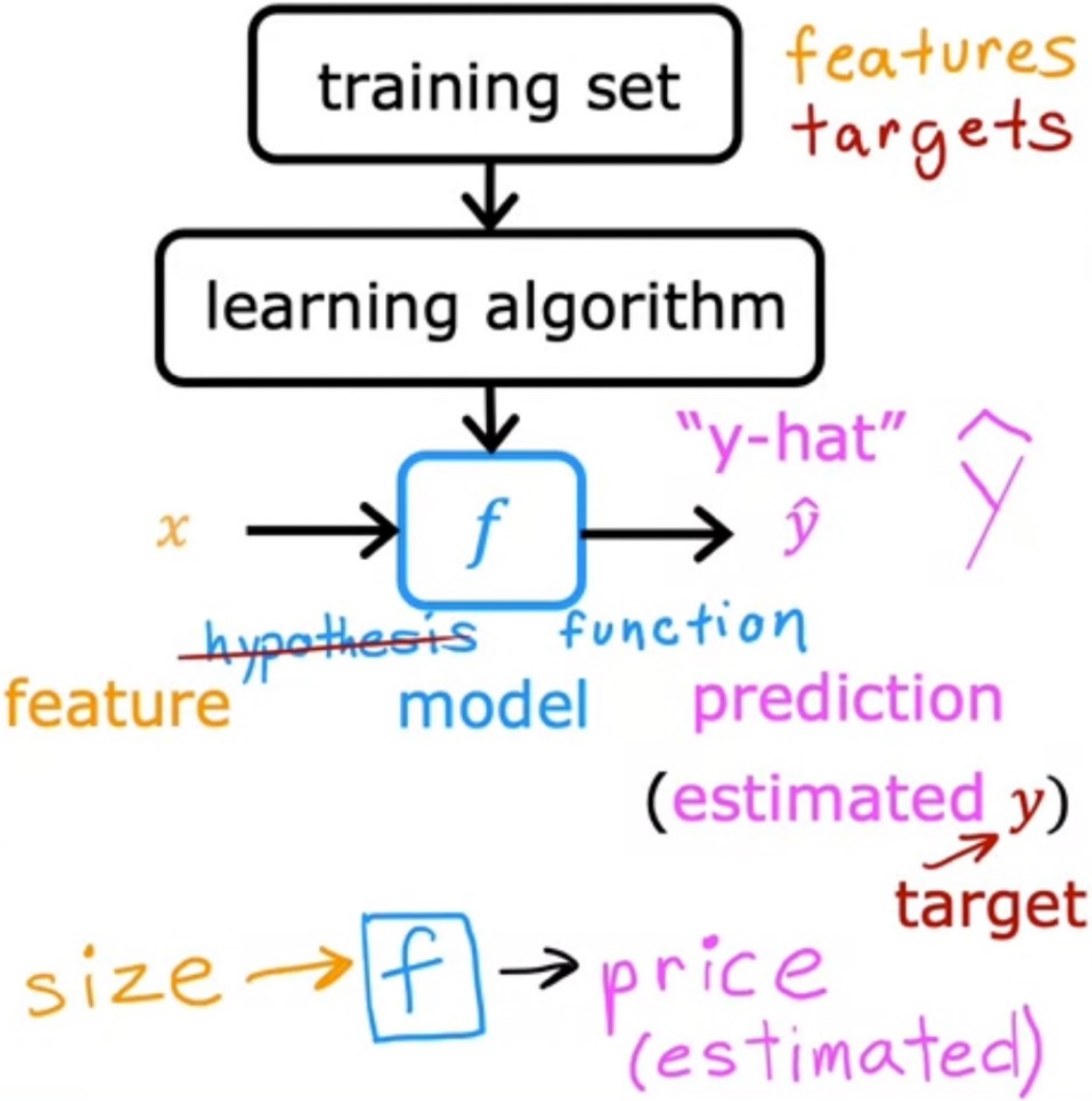
$(x^{(i)}, y^{(i)})$

$(x^{(i)}, y^{(i)})$ = i^{th} training example

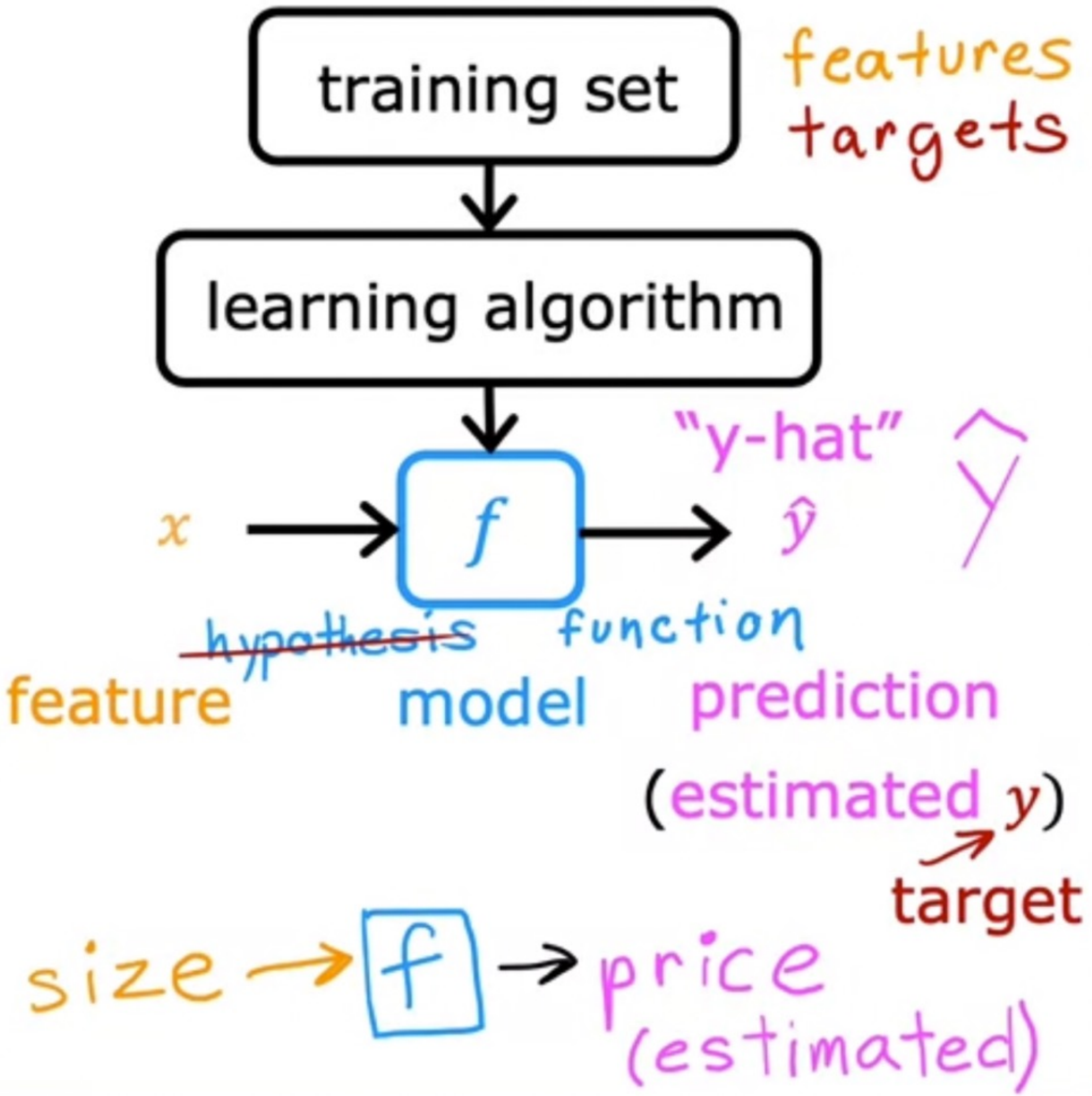
index

(1^{st}, 2^{nd}, 3^{rd} ...)


training set
features
targets
learning algorithm
"y-hat"
x
f
ŷ
~~hypothesis~~ function
feature
model
prediction
(estimated y)
target



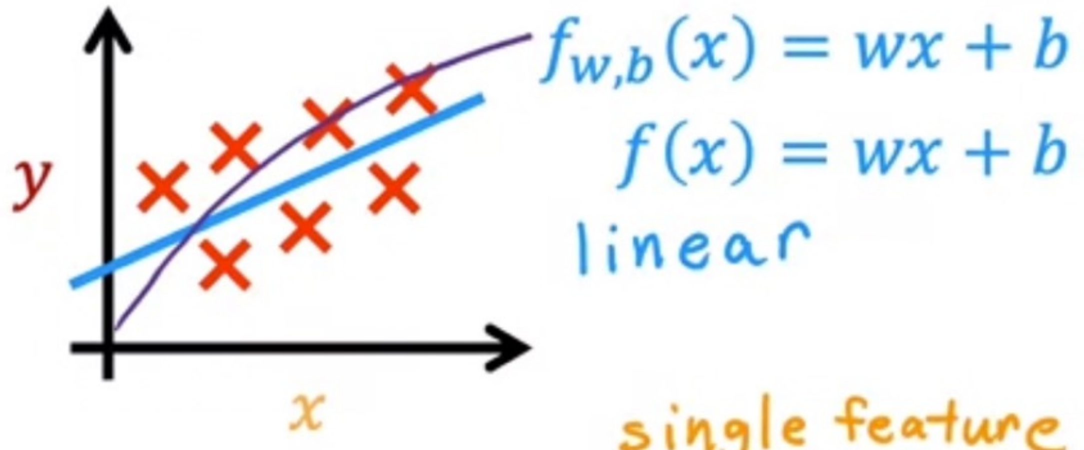
How to represent f?



How to represent f?

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

$$f(x)$$



Linear regression with one variable.
size

Univariate linear regression.
one variable

Cost Function

- How well the model is doing

Training set

features size in feet^2 (x)	targets price \$1000's (y)
2104	460
1416	232
1534	315
852	178
...	...

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

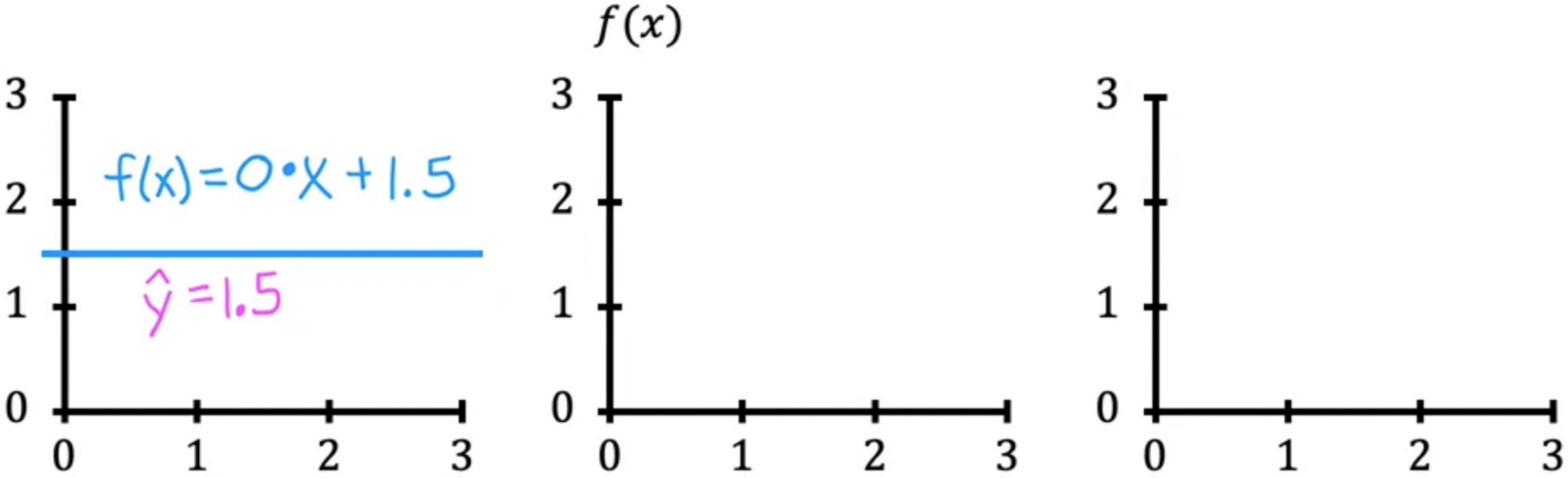
w, b: parameters

coefficients

weights

What do w, b do?

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

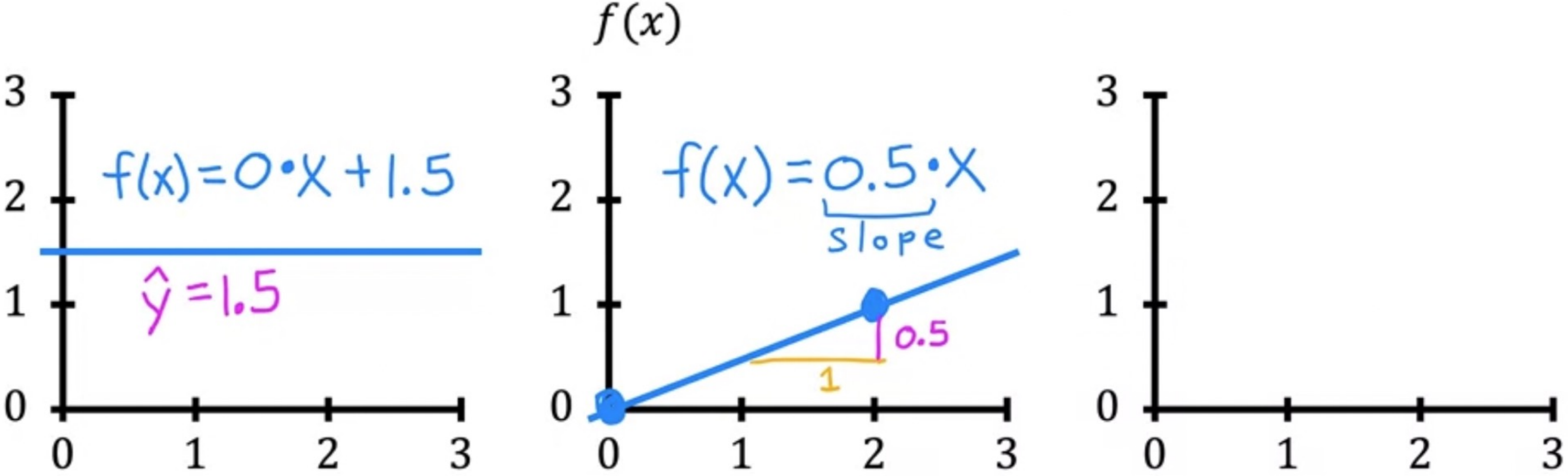


$\rightarrow$ $w = 0$

$\rightarrow$ $b = 1.5$

y-intercept

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



$\rightarrow$ $w = 0$

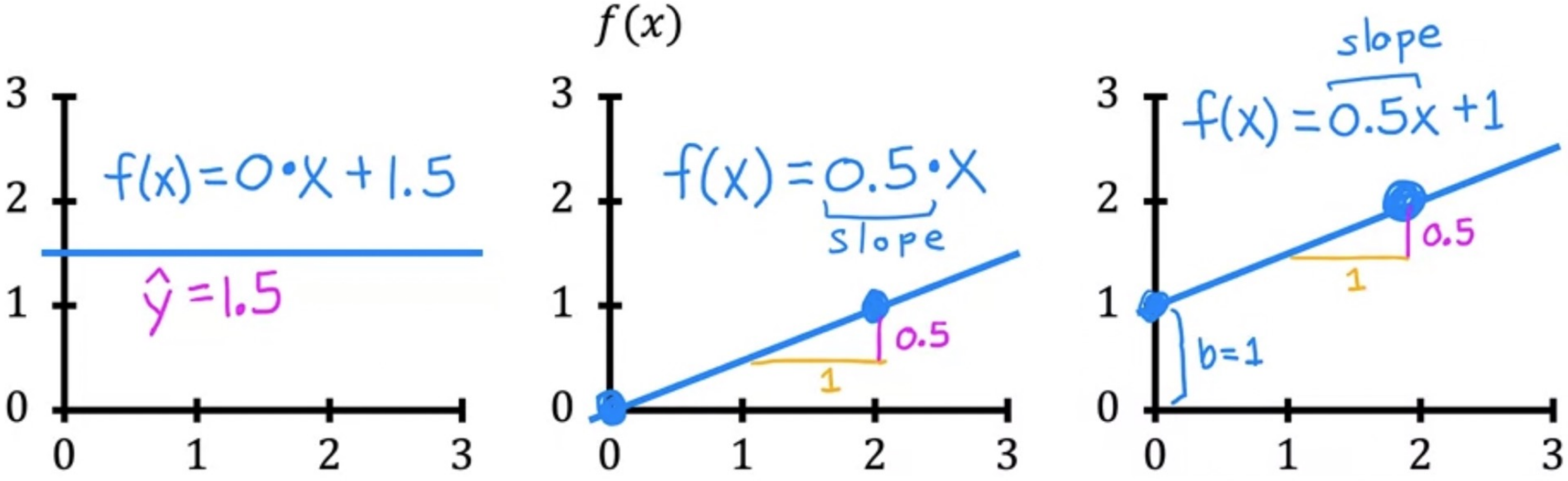
$\rightarrow$ $b = 1.5$

y-intercept

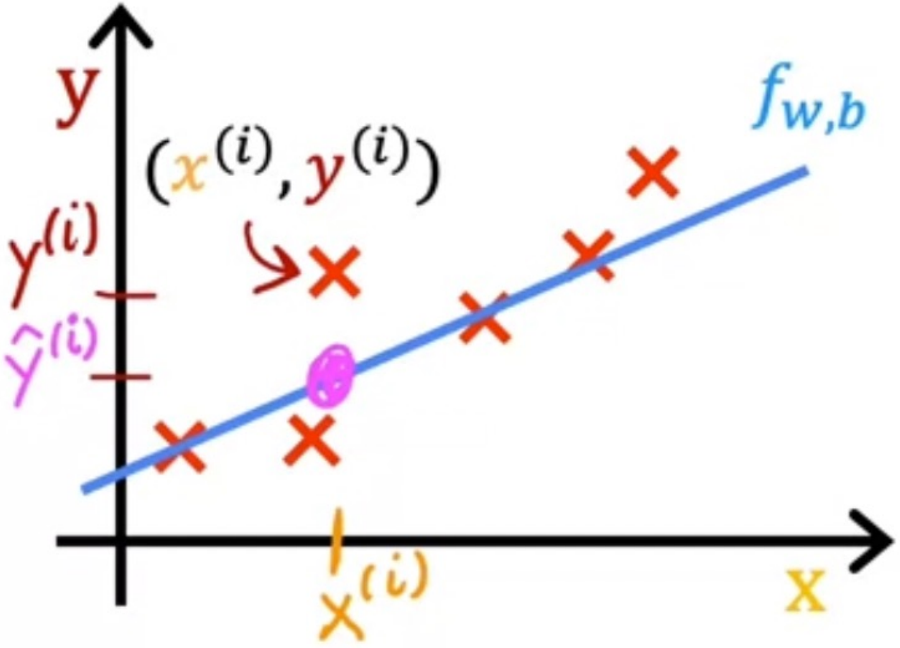
$\rightarrow$ $w = 0.5$

$\rightarrow$ $b = 0$

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



→ $w = 0$	→ $w = 0.5$	→ $w = 0.5$
→ $b = 1.5$	→ $b = 0$	→ $b = 1$
y-intercept		

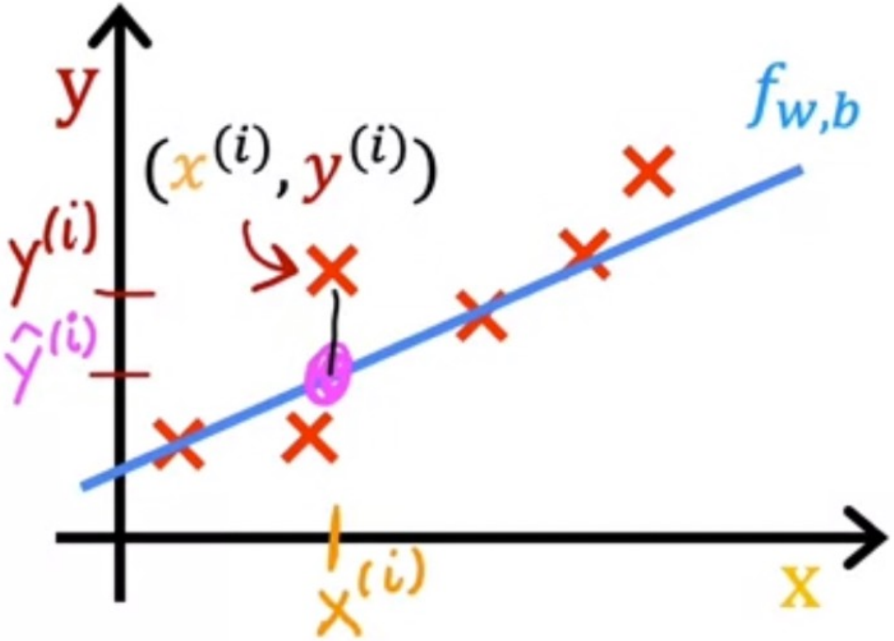


How Can you choose w and b so that our model fits the data well?
ANS: We use a cost function to measure how close is Type equation here.y-hat to y.

$$\hat{y}^{(i)} = f_{w,b}(x^{(i)})$$

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

Find w, b:
$\hat{y}^{(i)}$ is close to $y^{(i)}$ for all $(x^{(i)}, y^{(i)})$.



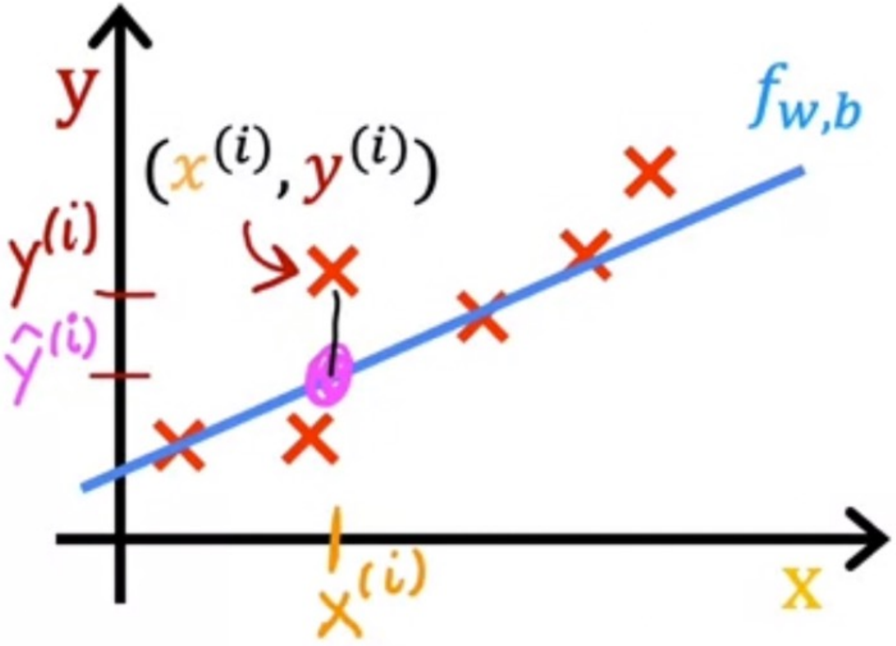
$$\hat{y}^{(i)} = f_{w,b}(x^{(i)})$$

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

Cost function

$$(\underbrace{\hat{y}^{(i)} - y^{(i)}}_{\text{error}})^2$$

Find w, b:
$\hat{y}^{(i)}$ is close to $y^{(i)}$ for all $(x^{(i)}, y^{(i)})$.



$$\hat{y}^{(i)} = f_{w,b}(x^{(i)})$$

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

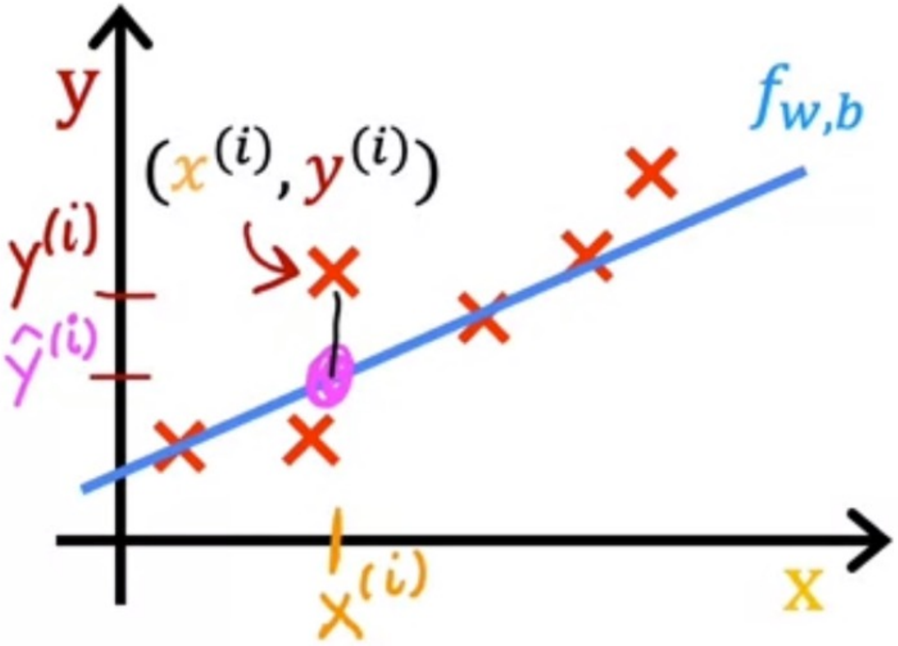
Cost function

$$\sum_{i=1}^{m} \left(\underbrace{\hat{y}^{(i)} - y^{(i)}}_{\text{error}} \right)^2$$

m = number of training examples

Find w, b:

$\hat{y}^{(i)}$ is close to $y^{(i)}$ for all $(x^{(i)}, y^{(i)})$.



$$\hat{y}^{(i)} = f_{w,b}(x^{(i)})$$

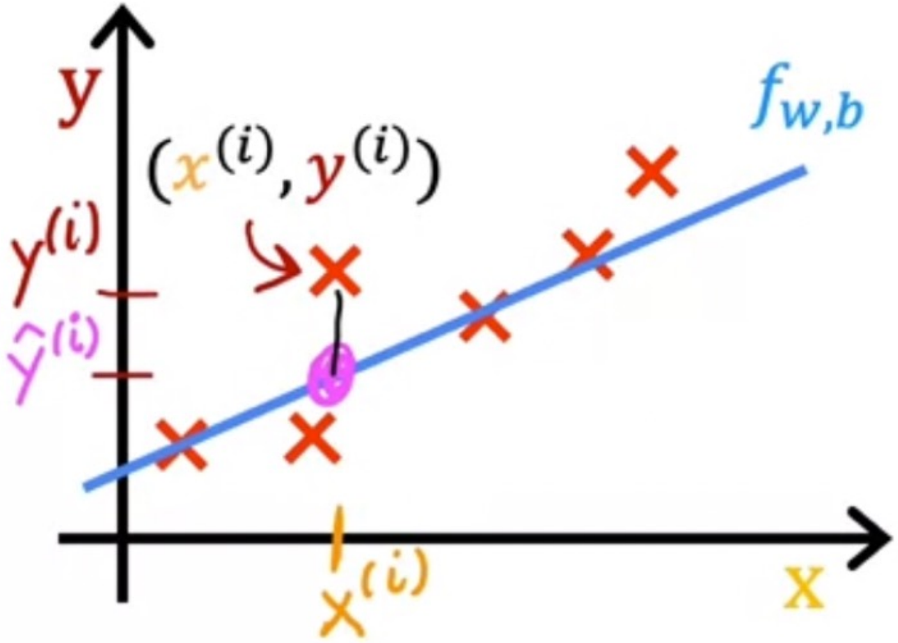
$$f_{w,b}(x^{(i)}) = wx^{(i)} + b$$

Cost function

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

m = number of training examples

Find w, b:
$\hat{y}^{(i)}$ is close to $y^{(i)}$ for all $(x^{(i)}, y^{(i)})$.



$$\hat{y}^{(i)} = f_{w,b}(x^{(i)})$$

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

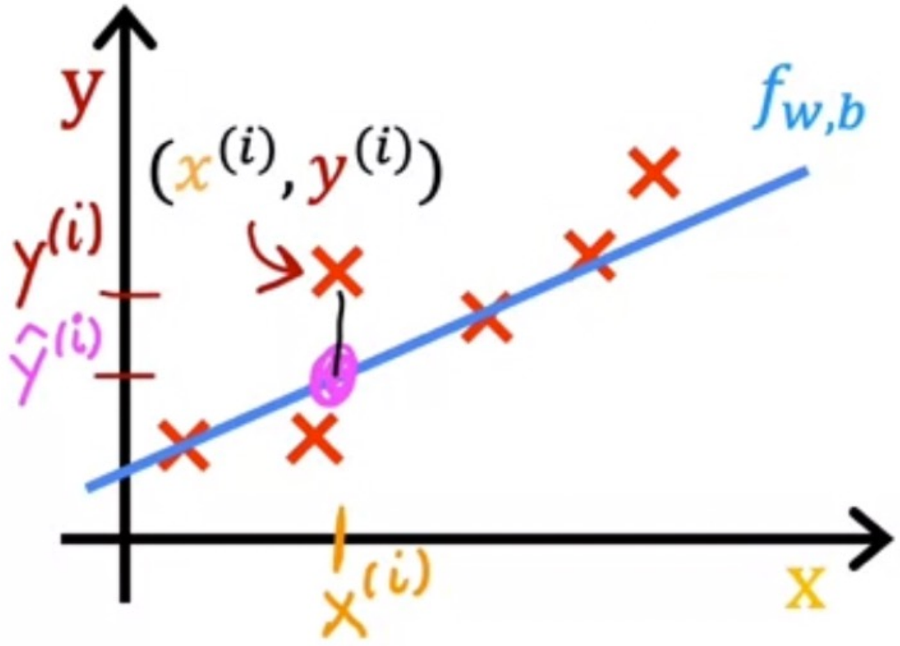
Cost function: Squared error cost function

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

m = number of training examples

There could be other cost function in different ML Algo,
however, this one is the most common one.

Find w, b:
$\hat{y}^{(i)}$ is close to $y^{(i)}$ for all $(x^{(i)}, y^{(i)})$.



$$\hat{y}^{(i)} = f_{w,b}(x^{(i)})$$

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

Cost function: Squared error cost function

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

m = number of training examples

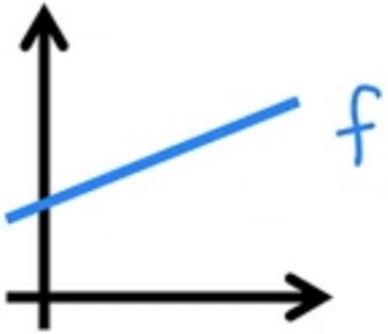
$$J(w,b) = \frac{1}{2m}\sum_{i=1}^{m}(f_{w,b}(x^{(i)}) - y^{(i)})^2$$

Find w, b:
$\hat{y}^{(i)}$ is close to $y^{(i)}$ for all $(x^{(i)}, y^{(i)})$.

RECAP

model:

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



parameters:

$$w, b$$

cost function:

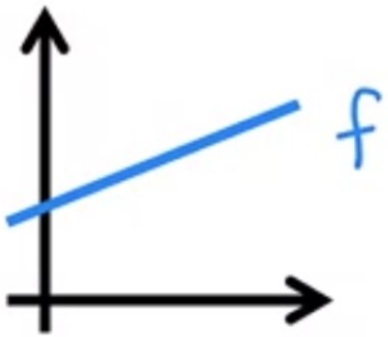
$$J(w,b) = \frac{1}{2m}\sum_{i=1}^{m}\left(f_{w,b}\left(x^{(i)}\right) - y^{(i)}\right)^2$$

goal:

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

model:

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



parameters:

w, b

cost function:

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

goal:

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

simplified

$$f_w(x) = wx$$

$b = 0$

w

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

$wx^{(i)}$

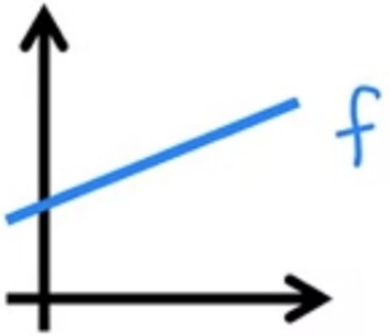
$$\underset{w}{\text{minimize}}\, J(w)$$

model:

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

parameters:

w, b



cost function:

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

goal:

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

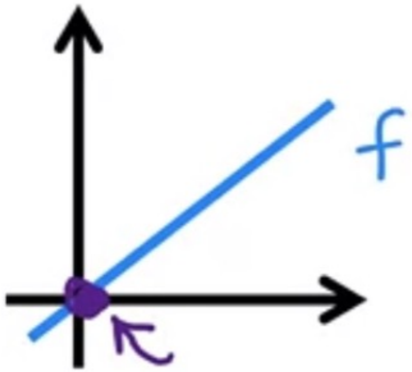
simplified

$$f_w(x) = wx$$

$b = \emptyset$

0 0

w



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

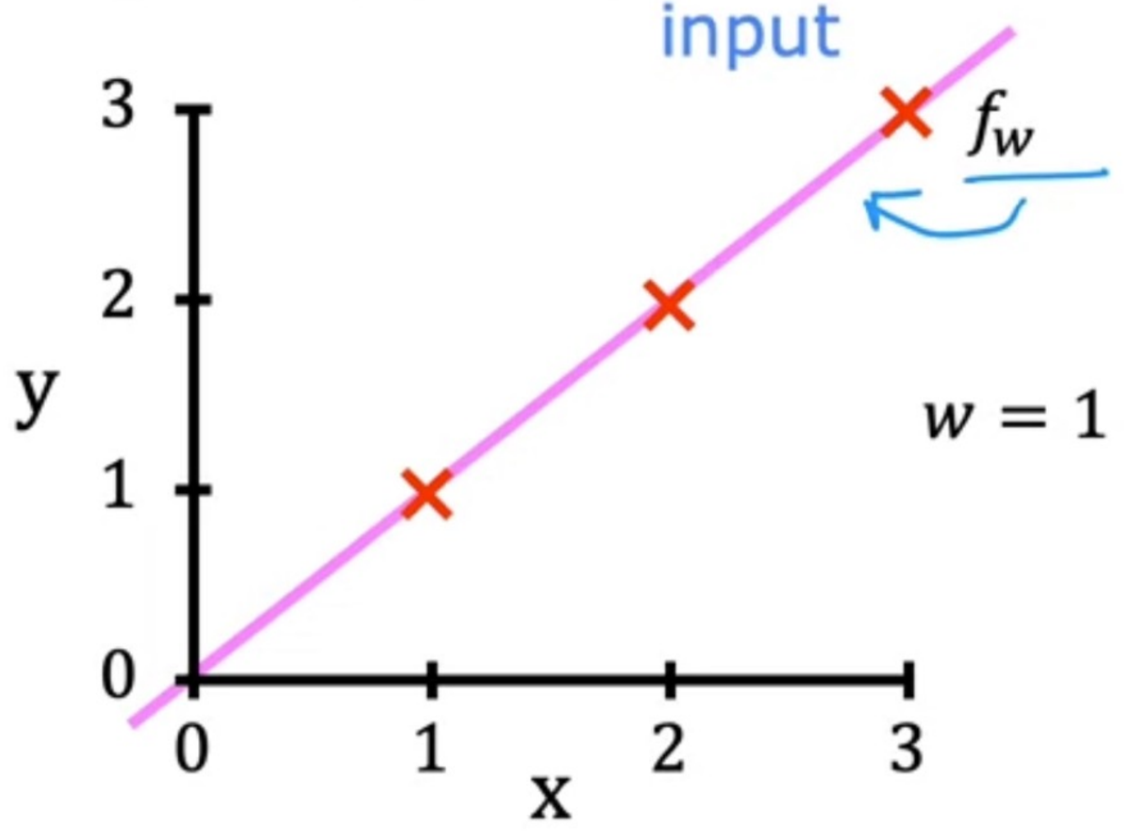
$wx^{(i)}$

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

$f_w(x)$

(for fixed w, function of x)

input



$J(w)$

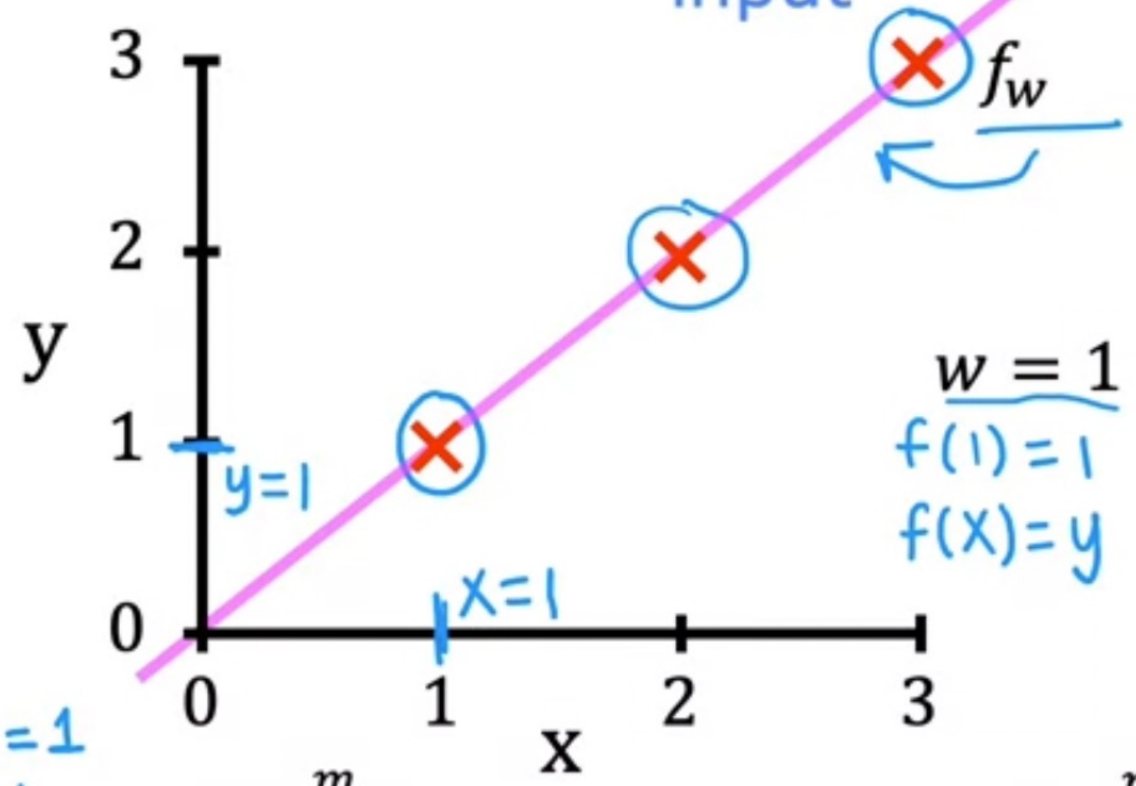
(function of w)

parameter

$f_w(x)$

(for fixed w, function of x)

input



$J(w)$

(function of w)

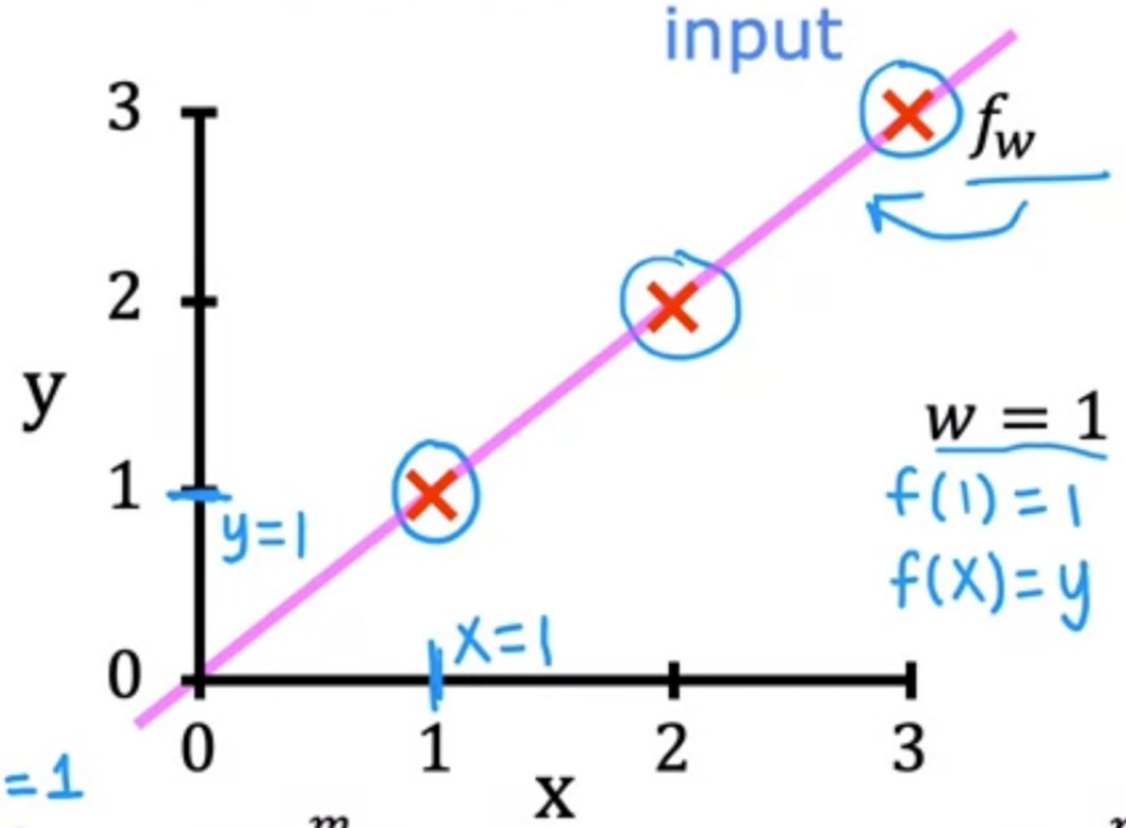
parameter

w=1

$$J(w) = \frac{1}{2m}\sum_{i=1}^{m}(\underbrace{f_w(x^{(i)})}_{wx^{(i)}} - y^{(i)})^2 = \frac{1}{2m}\sum_{i=1}^{m}(\overbrace{wx^{(i)} - y^{(i)}}^{0})^2 = \frac{1}{2m}(0^2+0^2+0^2)$$

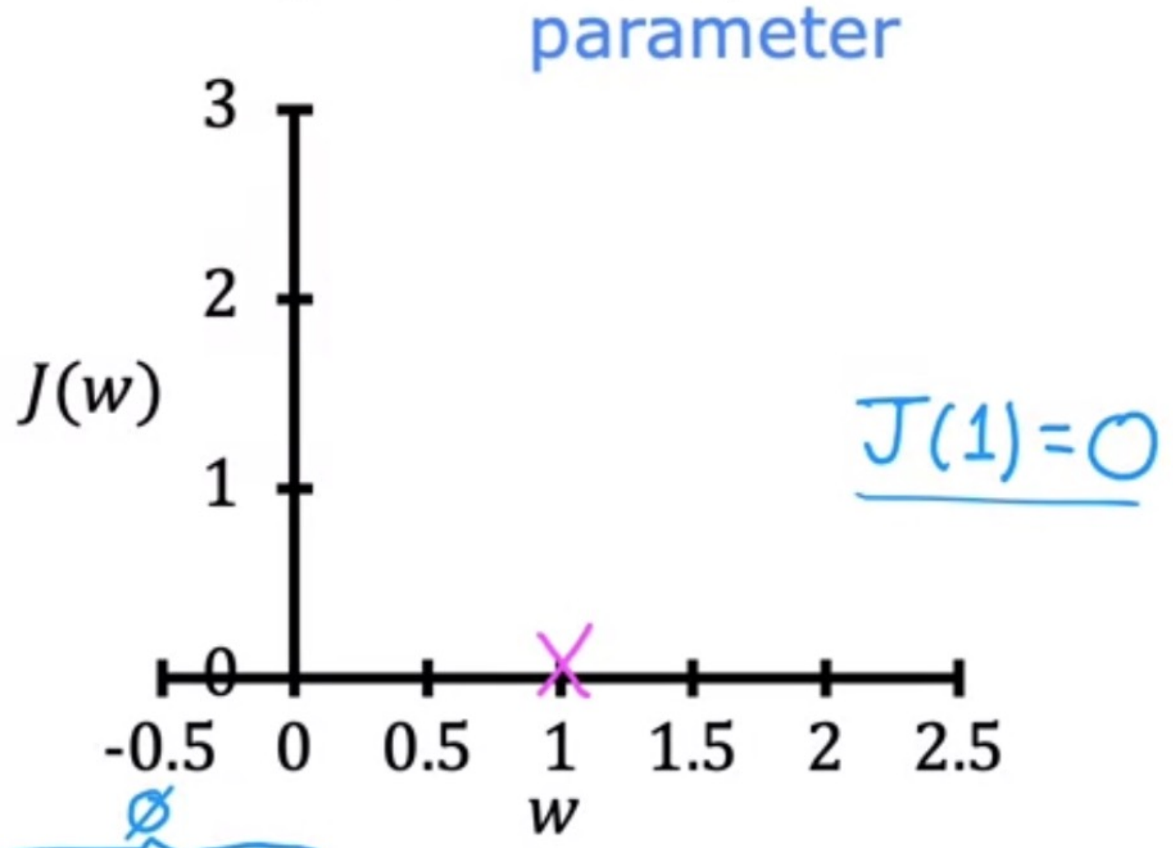
$f_w(x)$

(for fixed w, function of x)



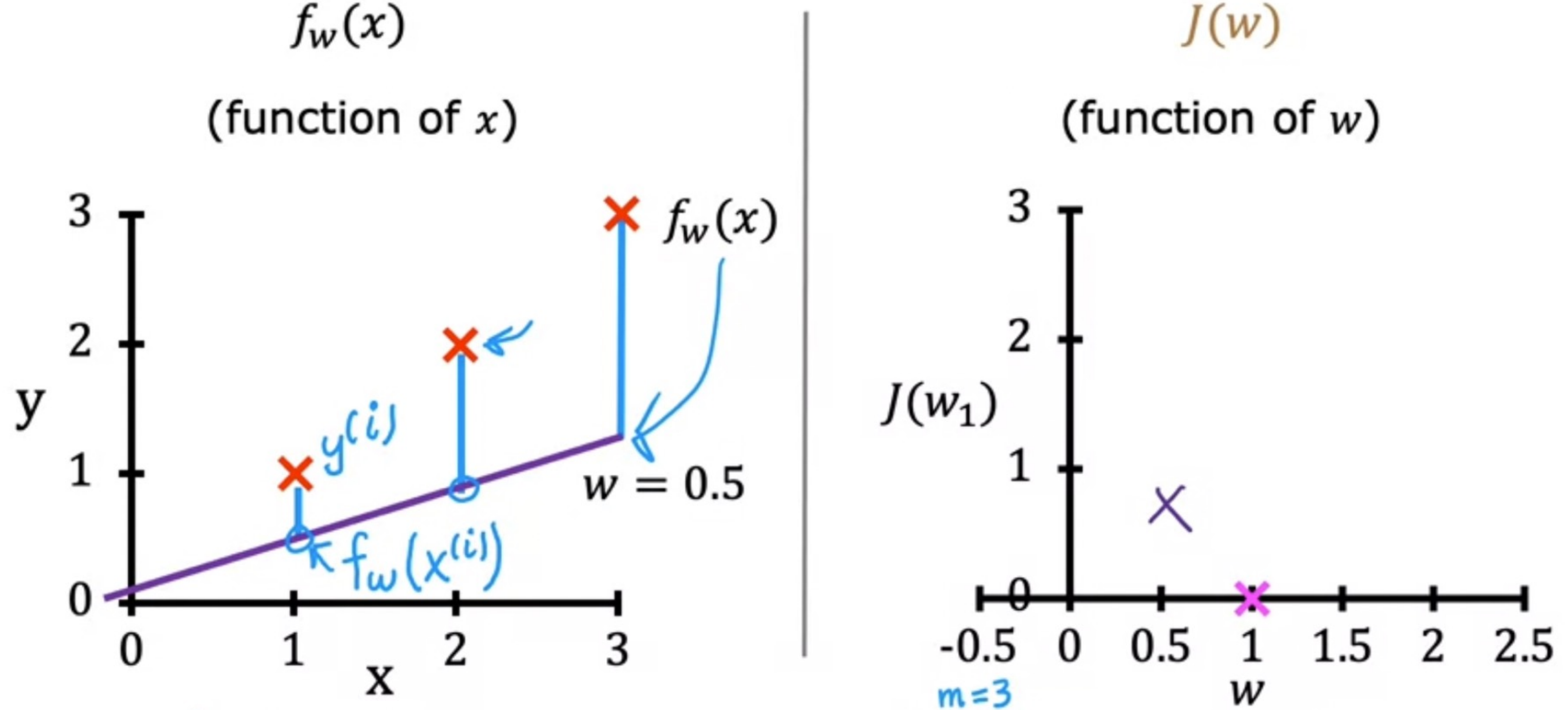
$J(w)$

(function of w)



$$\underbrace{J(w)}_{} \;\; \text{(w=1)} = \frac{1}{2m}\sum_{i=1}^{m}(\underbrace{f_w(x^{(i)})}_{wx^{(i)}} - y^{(i)})^2 = \frac{1}{2m}\sum_{i=1}^{m}(\overbrace{wx^{(i)} - y^{(i)}}^{\emptyset})^2 = \frac{1}{2m}(0^2 + 0^2 + 0^2) = 0$$

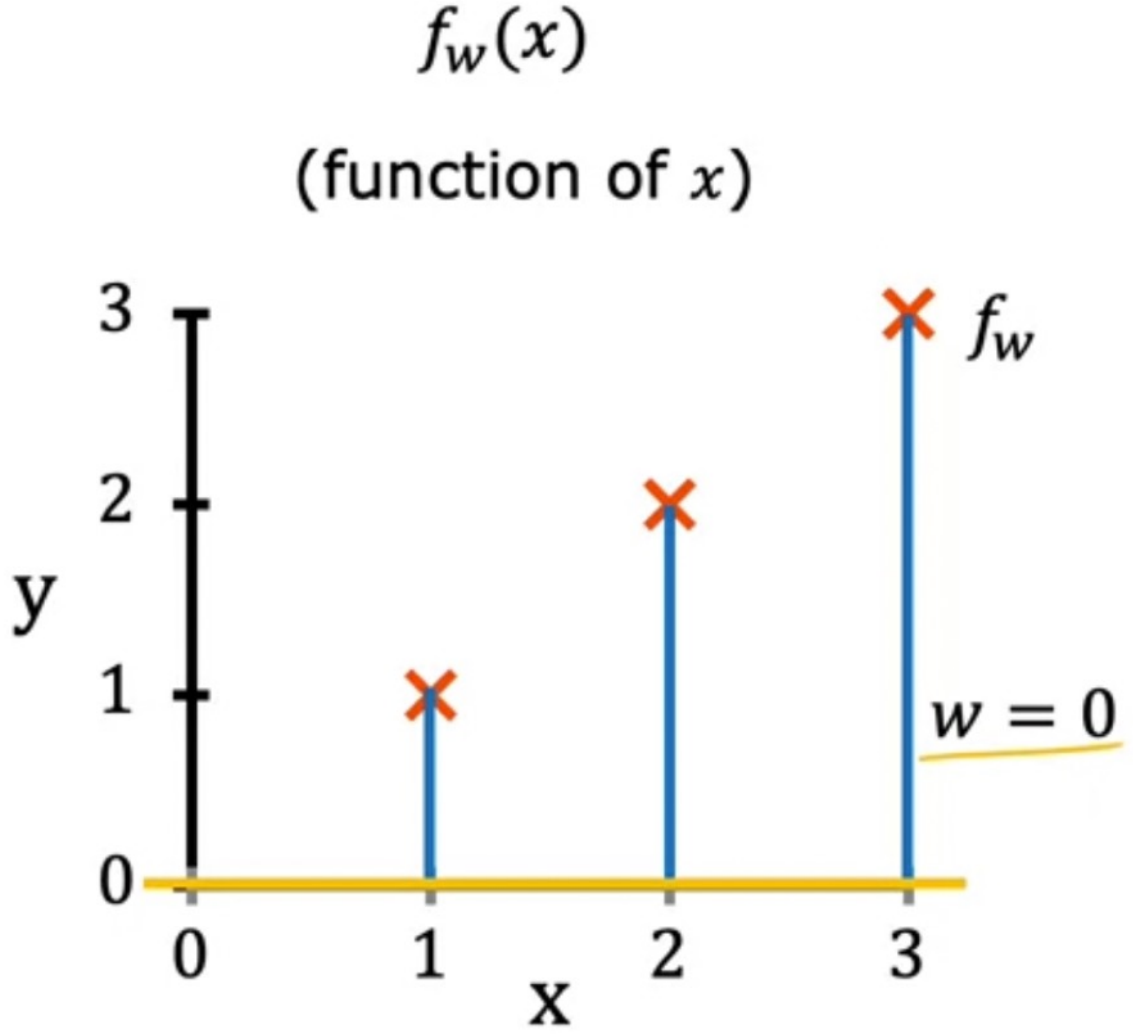
Lets see how w = 0.5 looks like

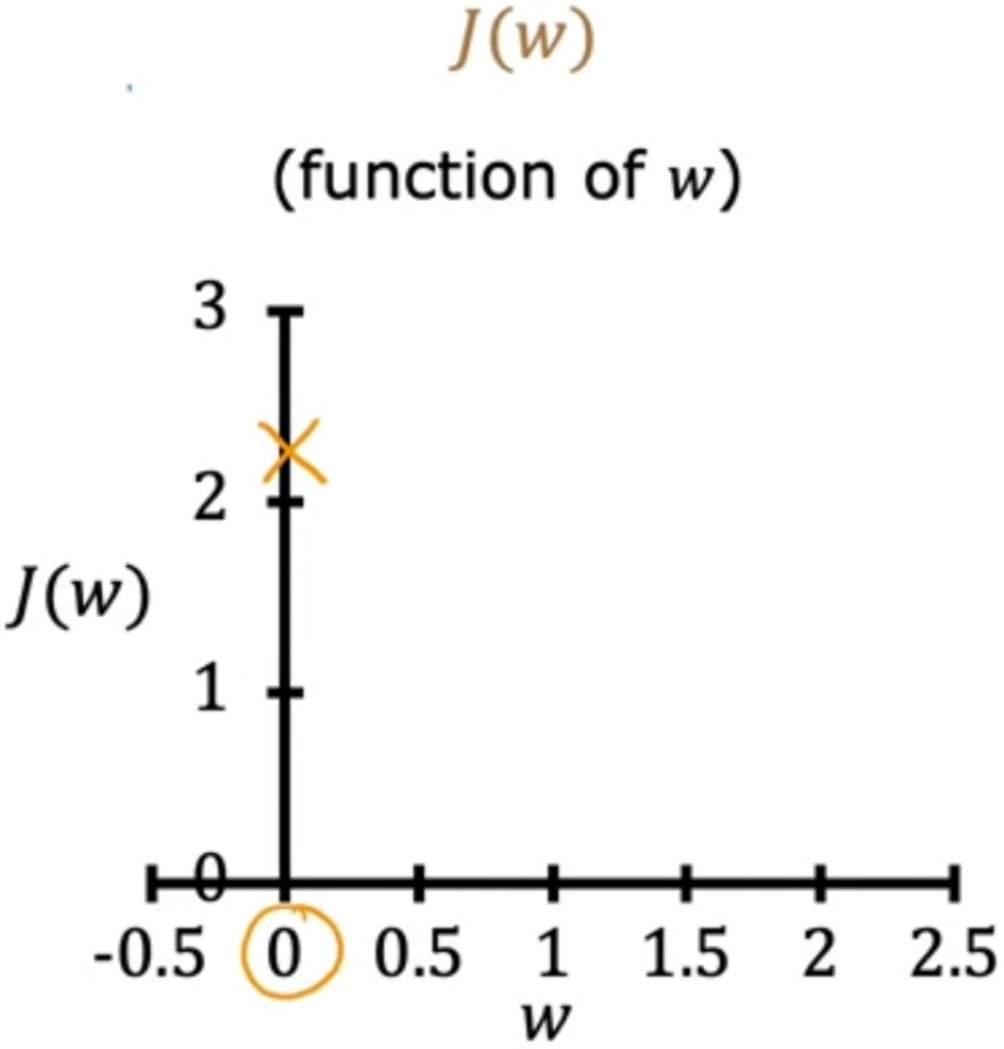


$m=3$

$$J(0.5) = \frac{1}{2m}\left[(0.5-1)^2 + (1-2)^2 + (1.5-3)^2\right] = \frac{1}{2 \times 3}[3.5] = \frac{3.5}{6} \approx 0.58$$

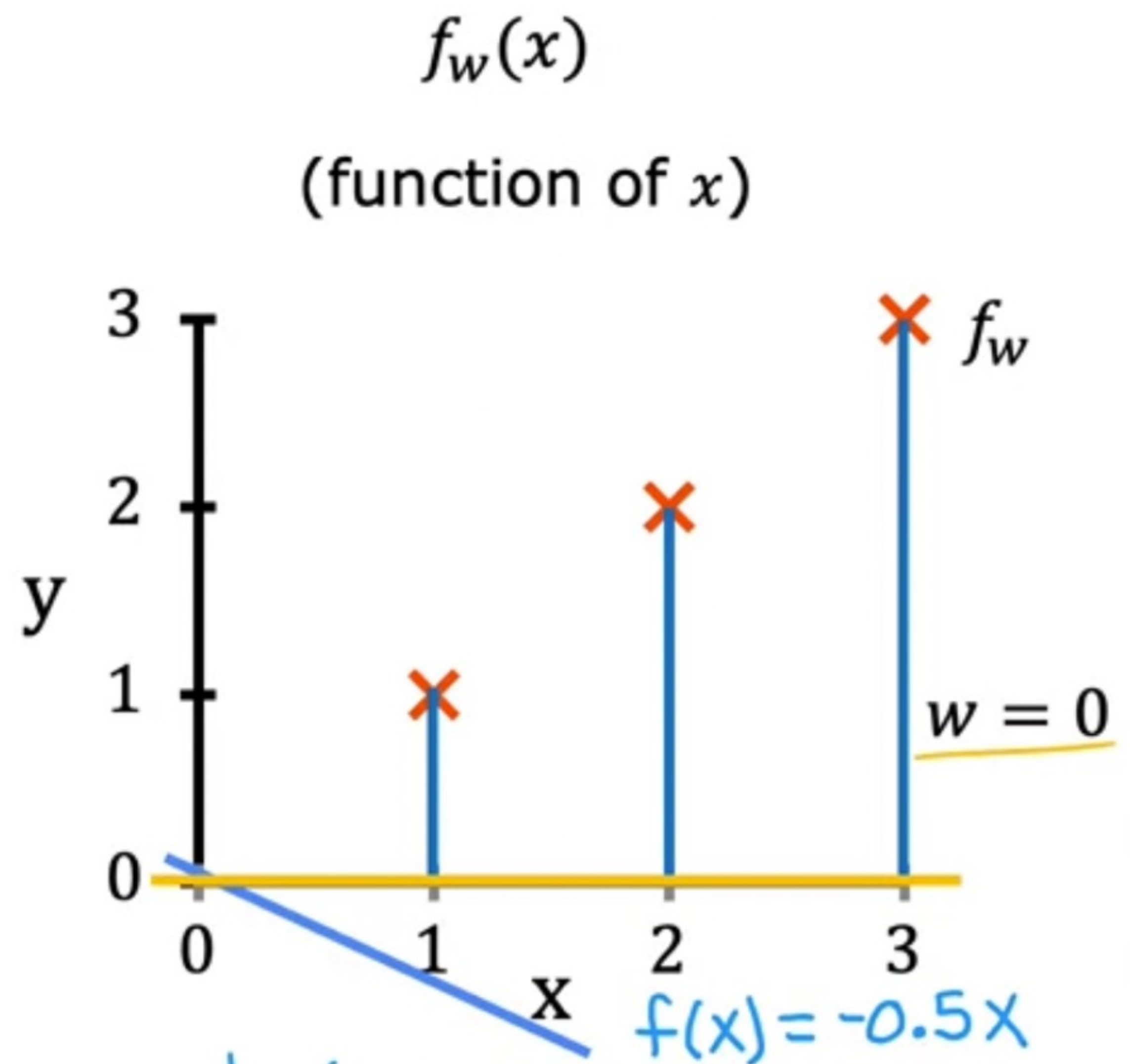
Lets see how w = 0 looks like

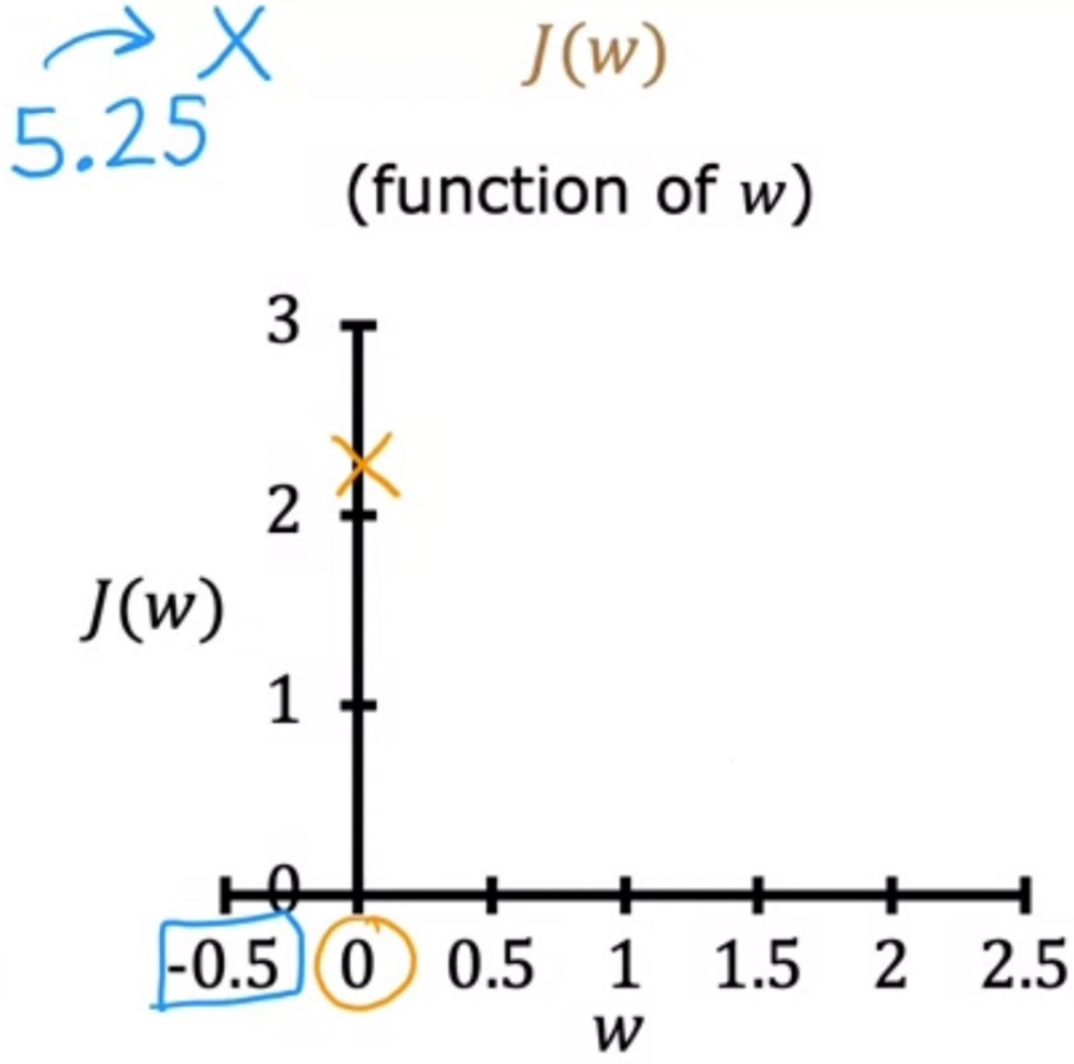




$$J(0) = \frac{1}{2m}\left(1^2 + 2^2 + 3^2\right) = \frac{1}{6}[14] \approx 2.3$$

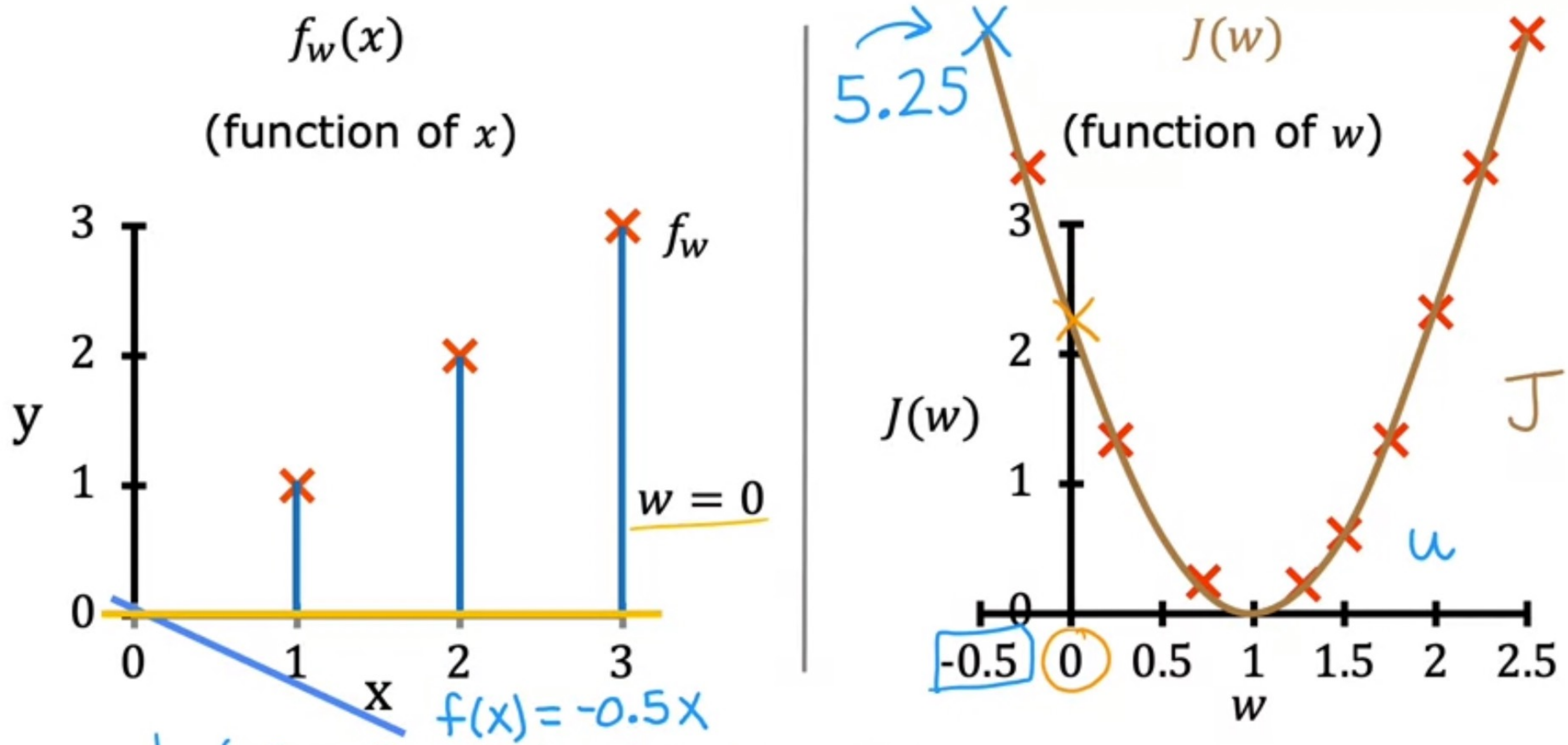
Lets see how w = -0.5 looks like = very high cost!





$$J(0) = \frac{1}{2m}\left(1^2 + 2^2 + 3^2\right) = \frac{1}{6}[14] \approx 2.3$$

We can put different value of w and find the cost function



$$J(0) = \frac{1}{2m}\left(1^2 + 2^2 + 3^2\right) = \frac{1}{6}[14] \approx 2.3$$

How to choose a w such that it minimizes the cost???

goal of linear regression:

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

general case:

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