cost function

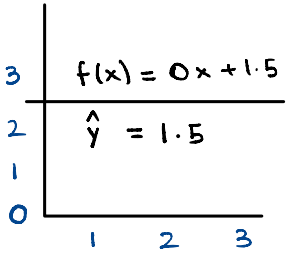
$y = mx + b$ (linear)

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

w, b :- parameters

Training set

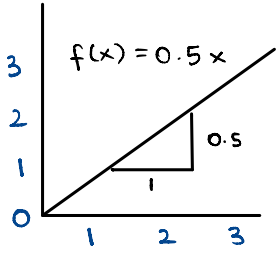
features size in feet^2 (x)	targets price \$1000's (y)
2104	460
1416	232
1534	315
852	178
...	...



→ $w = 0$

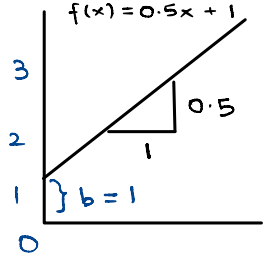
→ $b = 1.5$

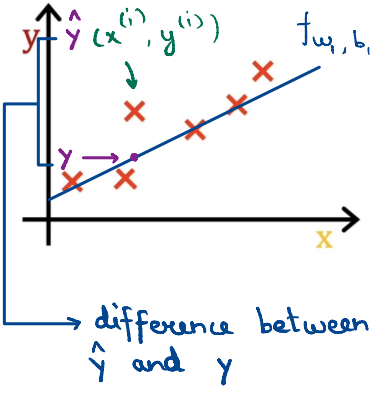
↳ y-intercept since $\hat{y} = b$



→ $w = 0.5$

$b = 0$





In application

← In this dataset, we have bunch of points marked

Our job is to choose the closest function $f(x) = wx + b$ such that it comprises most of the points (best case :- all points)

Q. How can we decide which function best measures the dataset ?

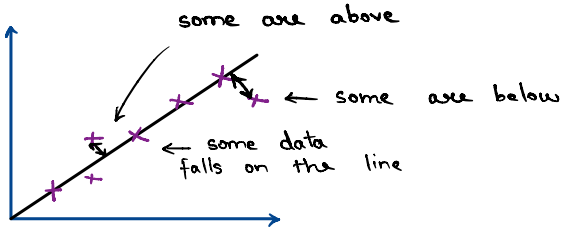
A. For that we're going to take a cost function.

$$(\hat{y} - y) \rightarrow \text{error}$$

$\hat{y}$: prediction value

y : target value

The less the error, the better the function.



Since in data below, the difference can be less than 0, we square the error.

cost function :-

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

error

m = no. of training examples

→ We also need to make sure that the error doesn't get bigger when the dataset gets bigger.

we compute the average squared error

$$= \frac{\text{total squared error}}{m}$$

Conventionally, cost function $= \frac{1}{2m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)} \right)^2$

There are different functions for different machine learning problems.

Squared error function is the most used function for linear regression.