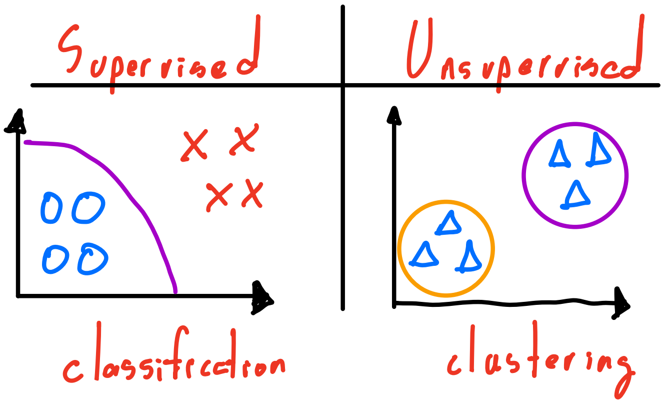
Machine-Learning: teaching a computer certain behavior for a task.

Two types of ML:

- **Supervised:** output/answer is known, learns from correct answers.
- **Unsupervised:** output not given, finds patterns in data.



Regression: predict numbers

$\boxed{\text{training set}}$

↓

$\boxed{\text{learning alg}}$

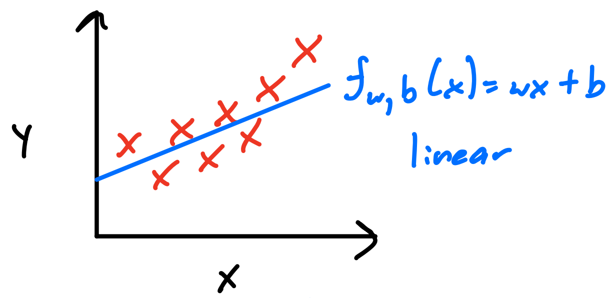
↓

x → f → y

hypothesis, model, prediction

How to represent model?

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



univariate linear regression

Cost Function: tells how well the model is doing.

Training Set

features size (x)	targets price (y)
2104	460
1416	232
1534	315
852	178

note: $x^{(i)}$ is i^{th} example

ex: $x^{(2)} = 1416$

$y^{(3)} = 315$

univariate linear-regression model:

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

$\hat{y}$ is prediction

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

Use squared error cost function:

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

m: number of training examples

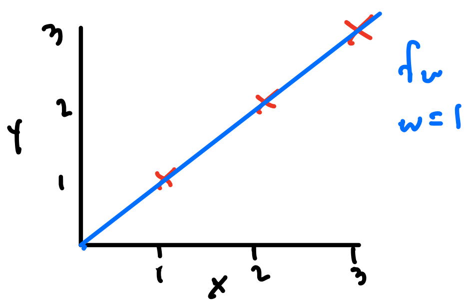
goal: want to find w, b such that $J(w,b)$ is small as possible

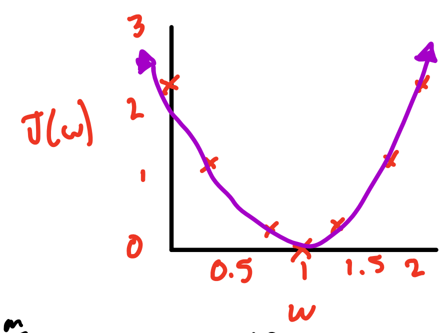
$f_w(x)$

(for fixed w, function of x)

$J(w)$

(function of w)





$w=1$

$$J(w) = \frac{1}{2m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)}\right)^2 = \frac{1}{2m} \sum_{i=1}^{m} \left(wx^{(i)} - y^{(i)}\right)^2 = 0$$

Gradient Descent: Systematic approach of minimizing $J(w,b)$

Outline:

- start w/ some w, b
- keep changing w, b to reduce $J(w,b)$ til you settle near a minimum.

Algorithm:

simultaneous update

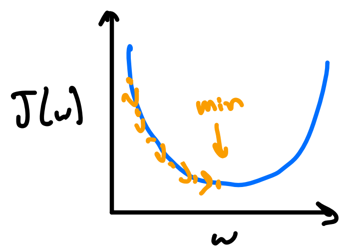
$$\begin{cases} w = w - \alpha \frac{\partial}{\partial w} J(w,b) \\ b = b - \alpha \frac{\partial}{\partial b} J(w,b) \end{cases}$$

α: learning rate

Learning:

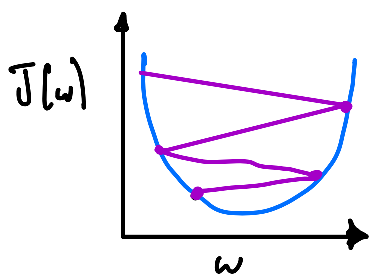
if α is too small...

grad desc can be slow.



if α is too large...

grad desc may overshoot



Gradient Descent for Linear Regression:

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

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

Gradient Descent:

repeat til convergence

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