ML/DL for Everyone Season2

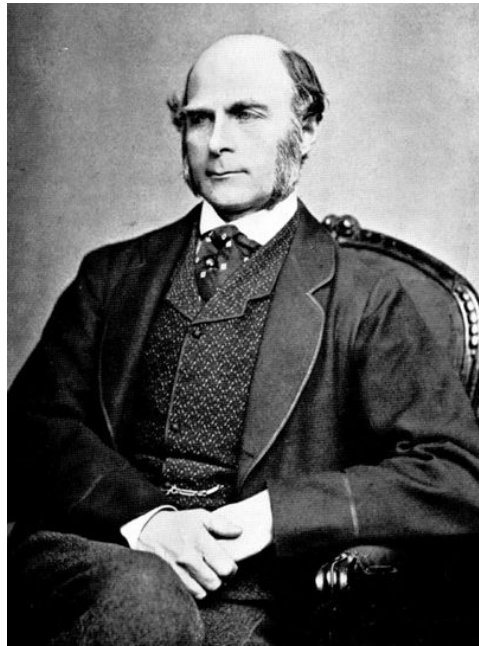with TensorFlow

02 - Simple Linear Regression

다루고자 하는 주제

- Regression
- Linear Regression
- Hypothesis
- Which hypothesis is better?
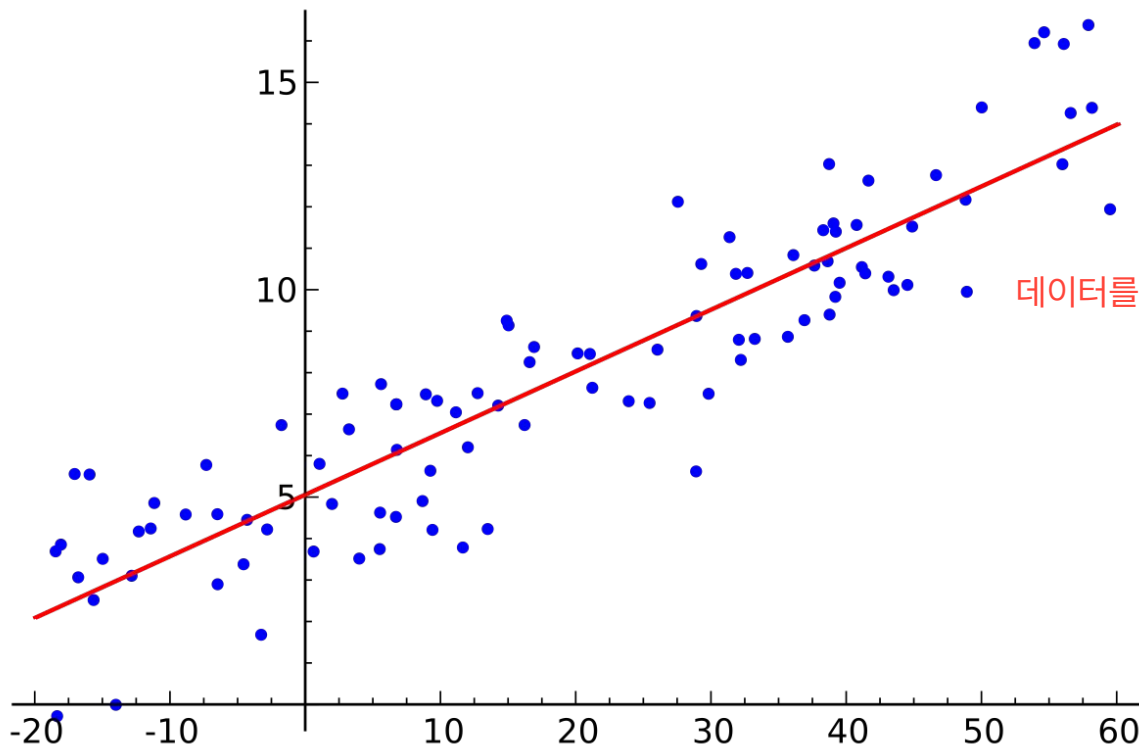- Cost, Cost function
- Goal: Minimize cost

Regression

"Regression toward the mean"

전체의 평균으로 되돌아간다

Sir Francis Galton (1822 ~ 1911)

Linear Regression

$$y = ax + b$$

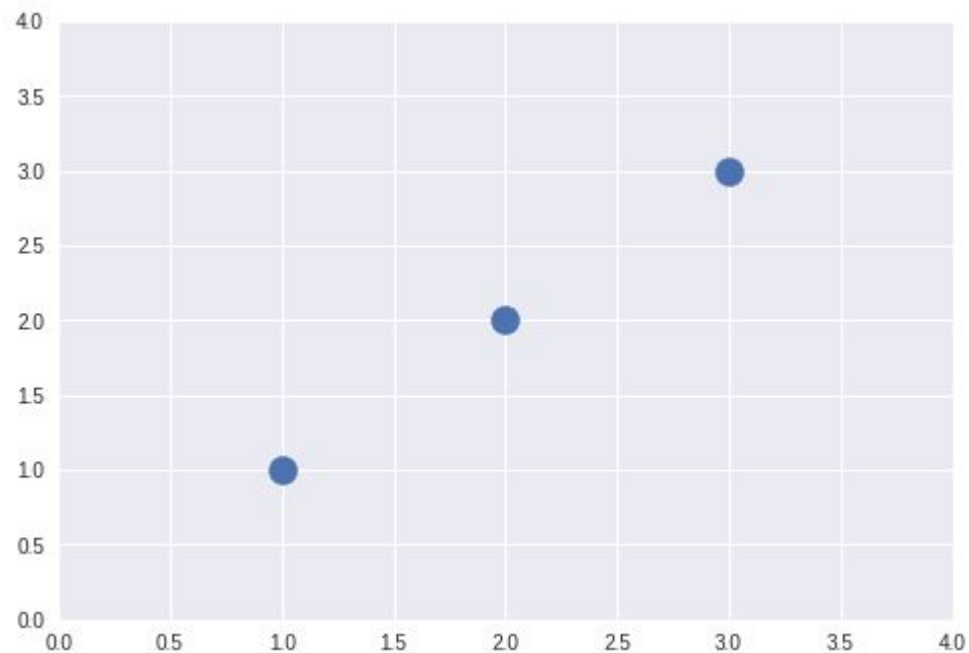데이터를 가장 잘 대변하는 직선의 방정식을 찾는 것

Predicting exam score: regression

x (hours)	y (score)
10	90
9	80
3	50
4	30

Regression

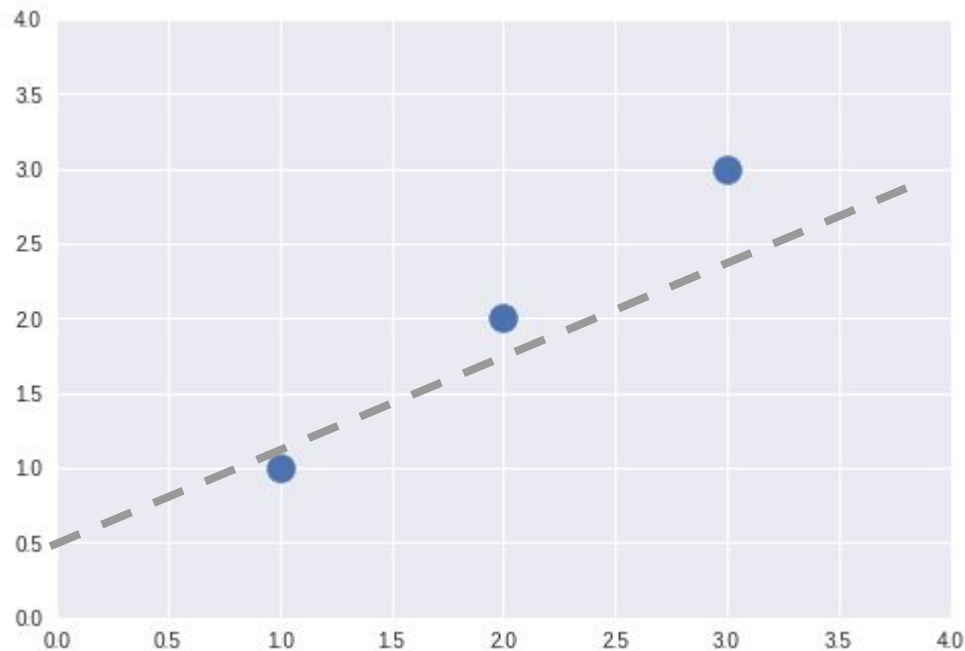| x	y
1	1
2	2
3	3

Regression

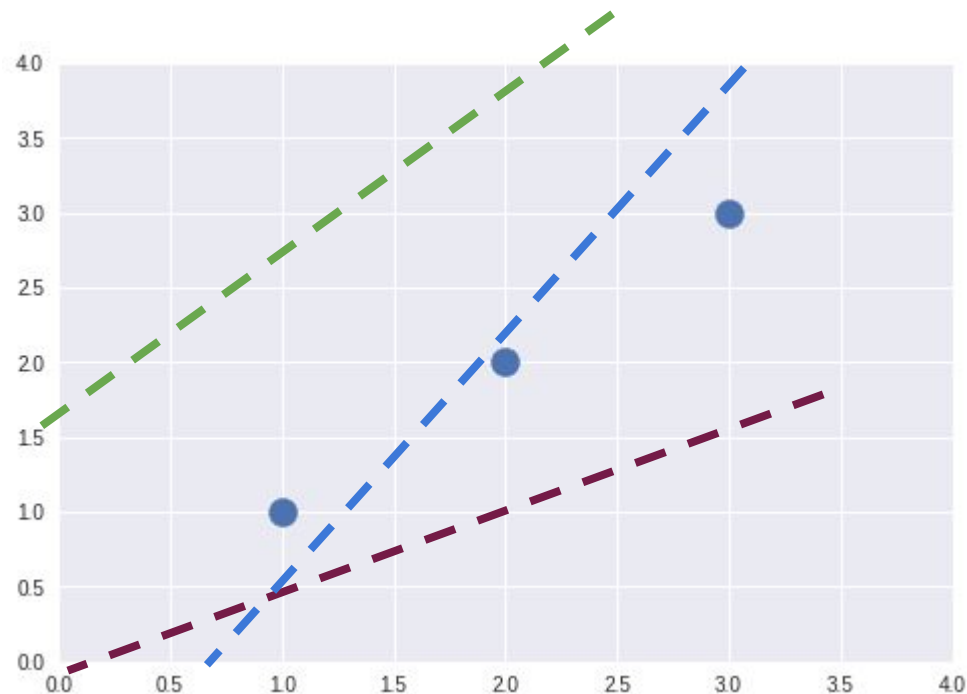| x	y
1	1
2	2
3	3

Hypothesis
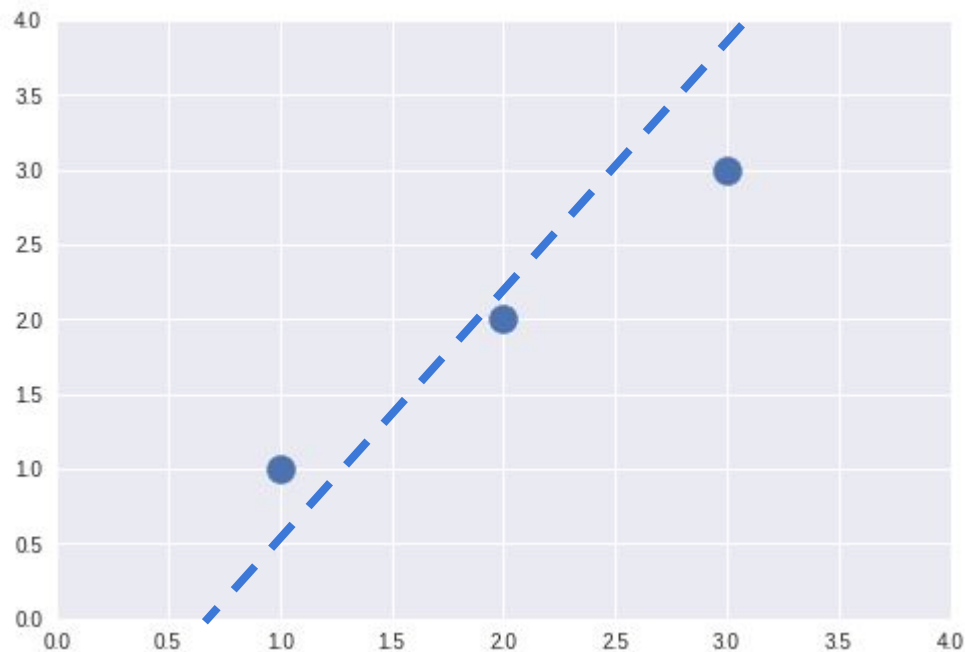## (Linear)

$$H(x) = Wx + b$$

Which hypothesis is better?

$$H(x) = Wx + b$$

Which hypothesis is better?

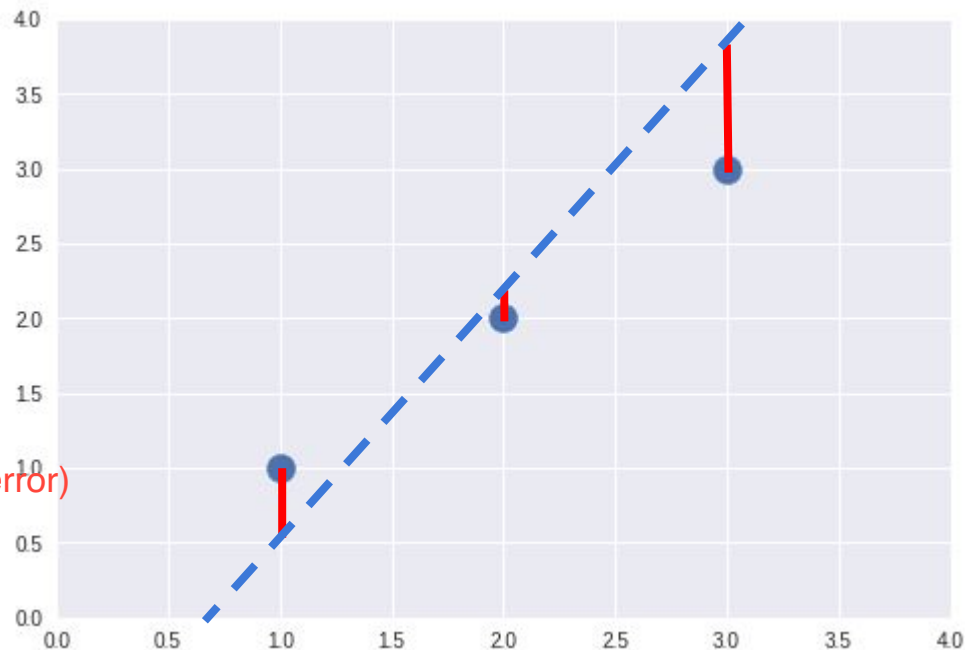$$H(x) = Wx + b$$

Cost

loss, error

$$H(x) = Wx + b$$

$$H(x) - y$$

예측   실제데이터

우리의 가설과 실제 데이터의 차이 (cost, loss, error)
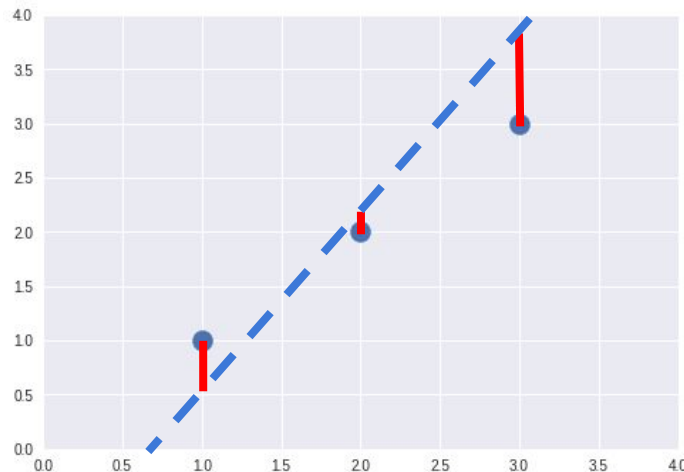
Cost

How **fit** the line to our (training) data

$$H(x) - y$$

$$\frac{(H(x_1)-y_1)^2+(H(x_2)-y_2)^2+(H(x_3)-y_3)^2}{3}$$

$$cost(W) = \frac{1}{m} \sum_{i=1}^{m} (Wx_i - y_i)^2$$

차이를 그냥 더해서 사용하면 그 차이가 음수도 있고, 양수도 있으므로 정확하지 않아 제곱해서 사용

Cost function

오차 제곱의 합의 평균

$$cost(W) = \frac{1}{m} \sum_{i=1}^{m} (Wx_i - y_i)^2$$

$$H(x) = Wx + b$$

Weight, bias

예측(H(x))과 실제(y)의 차이(에러)의 제곱의 합의 평균

$$cost(W, b) = \frac{1}{m} \sum_{i=1}^{m} (H(x_i) - y_i)^2$$

Goal: Minimize cost

$$\underset{W,b}{minimize}\ cost(W, b)$$

목표는 직선의 방정식이 데이터를 가장 잘 대변하도록, 비용이 최소가 되도록 하는 것

즉, Weight와 bias의 함수인 cost를 최소화하는 W와 b를 찾는 것

Summary

- Regression

- Linear Regression

- Hypothesis $H(x) = Wx + b$

- Cost function $cost(W, b) = \frac{1}{m} \sum_{i=1}^{m} (H(x_i) - y_i)^2$

- Goal: Minimize cost

What's Next?

- How to minimize cost