

ML lec 02

Simple Linear Regression

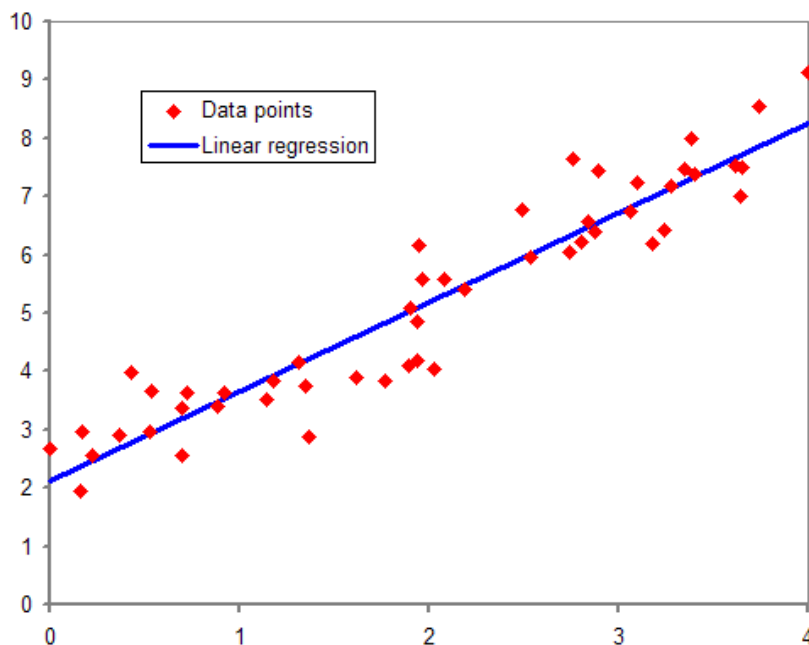
2021년 3월 12일 금요일 오후 10:55

■ Regression(되돌아가다, 후퇴, 퇴보)

"Regerssion toward the mean" (전체 평균으로 회귀하려 한다)

■ Linear Regression(선형회귀)

데이터를 잘 대변하는 직선의 방정식을 찾는 것



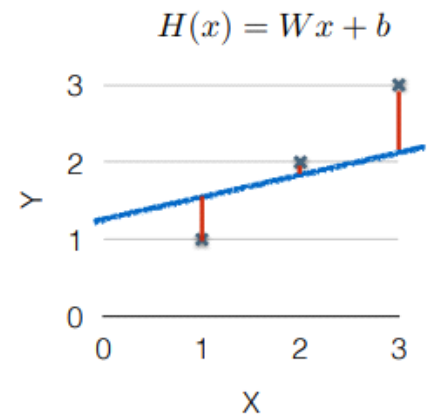
■ Cost(비용)

가설과 실제 데이터와의 차이. 빨간 선들의 총합이 작으면 작을 수록 파란 실선이 데이터를 잘 표현하고 있다고 말할 수 있다.

Cost function

- How fit the line to our (training) data

$$H(x) - y$$



우리의 가설(Hypothesis)과 실제 데이터와의 차이를 Cost, Loss, Error라고 표현한다.

우리는 이 cost를 최소화 해야한다. 근데 cost의 합이 최소가 되게만 하려면 문제가 발생 할 수 있다.

어떤 부분은 비용이 음수가 나타나지만 어떤 부분은 양수로 나타나므로 합을 더하는 것이 무의미해질 수 있다.

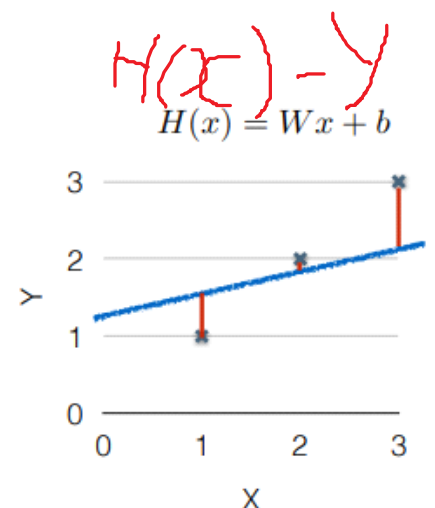
따라서 우리는 가설과 실제값의 차이를 제곱해서 평균을 낸다. 이를 비용함수로 많이 사용한다.


Cost function

- How fit the line to our (training) data

$$\frac{(H(x^{(1)}) - y^{(1)})^2 + (H(x^{(2)}) - y^{(2)})^2 + (H(x^{(3)}) - y^{(3)})^2}{3}$$

$$cost = \frac{1}{m} \sum_{i=1}^m (H(x^{(i)}) - y^{(i)})^2$$





$$cost = \frac{1}{m} \sum_{i=1}^m (H(x^{(i)}) - y^{(i)})^2$$

\sim

0	1	2	3
		x	

아래 가설 식에서 W 를 Weight(가중치), b 를 Bias(바이어스, 편향)라고 한다.

비용함수 $Cost(W, b)$ 는 오차 제곱의 평균으로 정의되며, 우리의 목적은

이 비용함수를 Minimize하는 w 와 b 를 찾아야 한다.

Cost function

$$cost = \frac{1}{m} \sum_{i=1}^m (H(x^{(i)}) - y^{(i)})^2$$

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

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