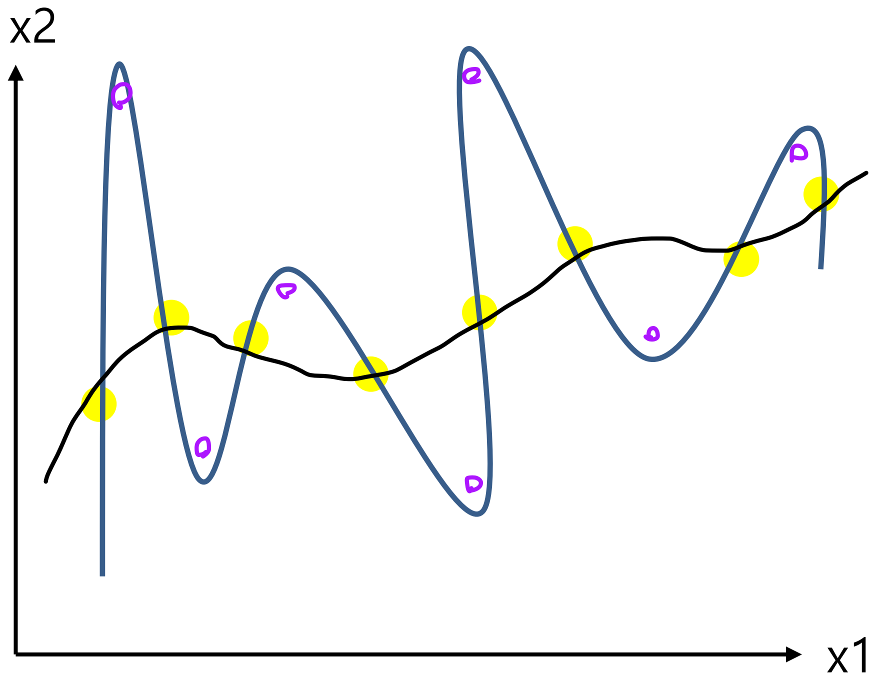
Deep Learning Basic

Jaewon Kim, Dankook Univ.

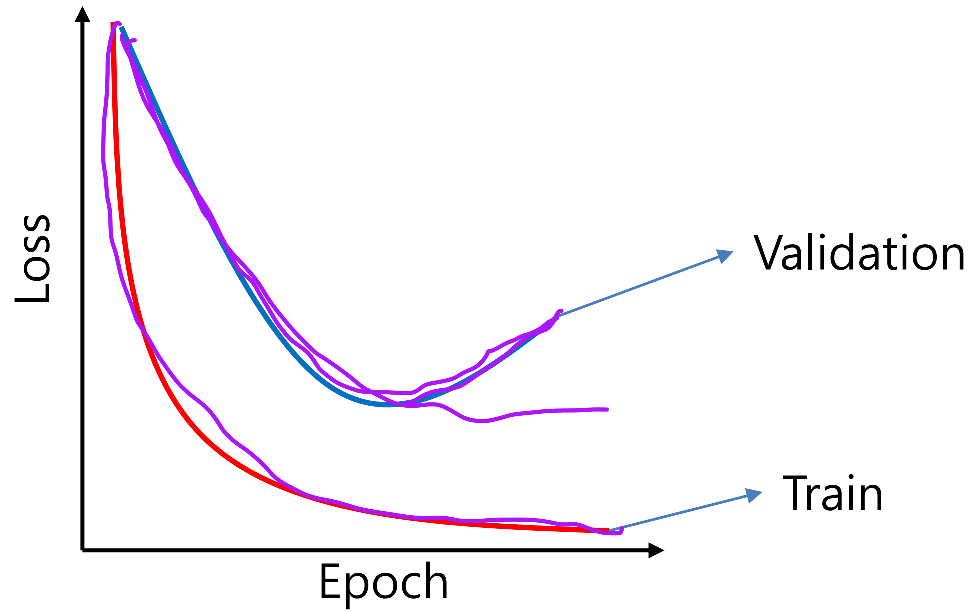
Chapter 3-1

Contents

Overfitting Review



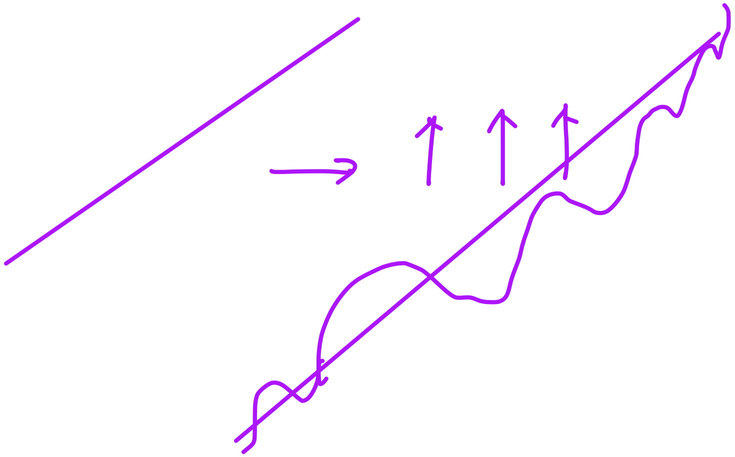
Overfitting Review



How to solve Overfitting ?

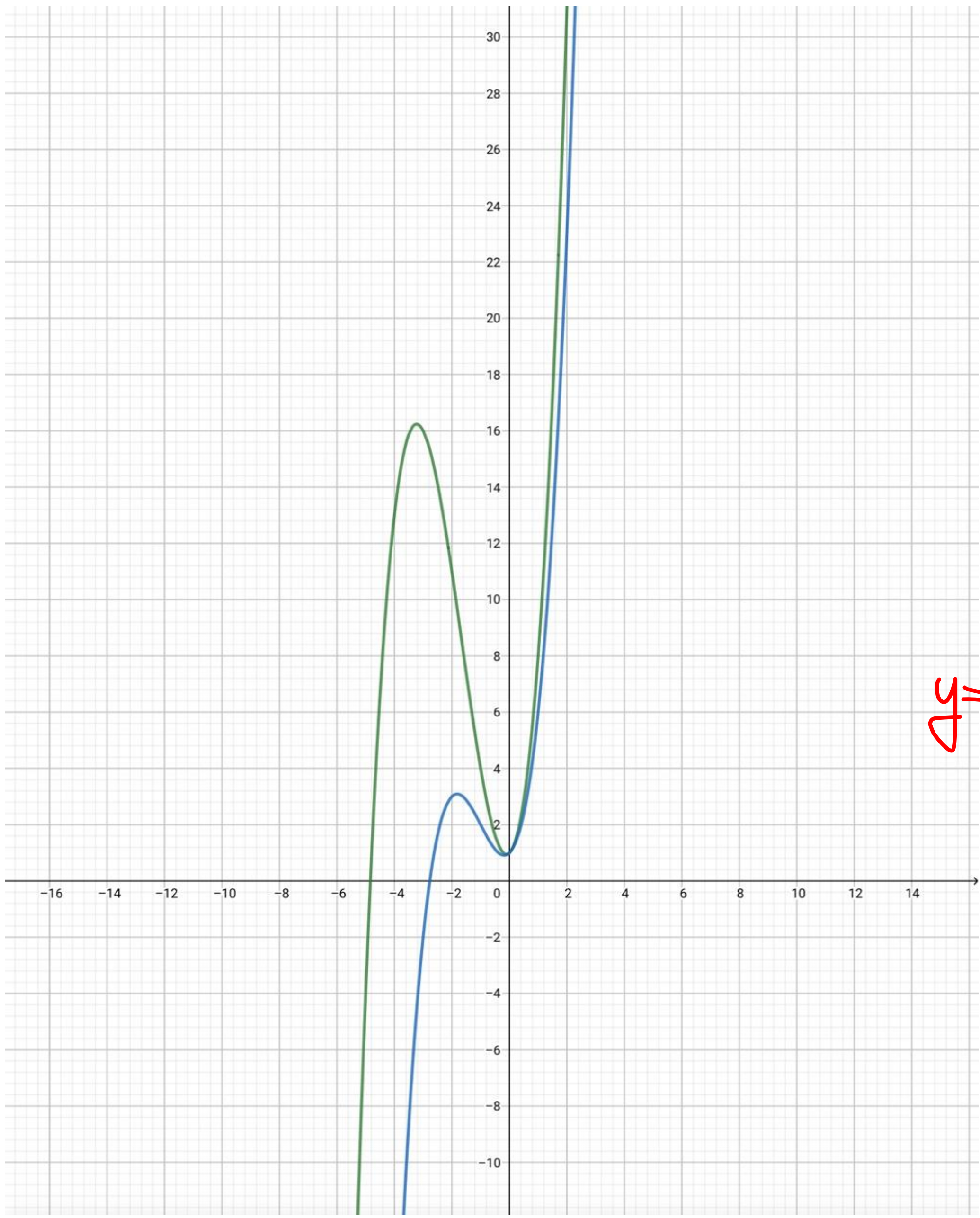
Regularization

정규화, 일반화, 규칙화



- 모델의 복잡도를 줄이고 일반화한다
- 학습 중, Weight가 너무 커지기 않게끔 일종의 penalty를 부과한다.

Regularization



$$f(x) = x^3 + 5x^2 + x + 1$$

$$g(x) = x^3 + 3x^2 + x + 1$$

$$y = w_1 x^3 + w_2 x^2 + w_1 x + 1$$

Norm

두 벡터의 길이 (or 차이, 거리)를 측정.

$$\|\mathbf{x}\|_p := \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$

$p=1 \rightarrow L1$

$p=2 \rightarrow L2$

P = Norm의 차수

L1 : MAE

L2 : MSE

L1-Norm & L2-Norm

L1-Norm

$$d_1(\mathbf{p}, \mathbf{q}) = \|\mathbf{p} - \mathbf{q}\|_1 = \sum_{i=1}^{n} |p_i - q_i|$$

Ex) p=(2, -1, 5) / q = (3, 4, -2)

$$|p-q| = |2-3| + |-1-4| + |5--2|$$

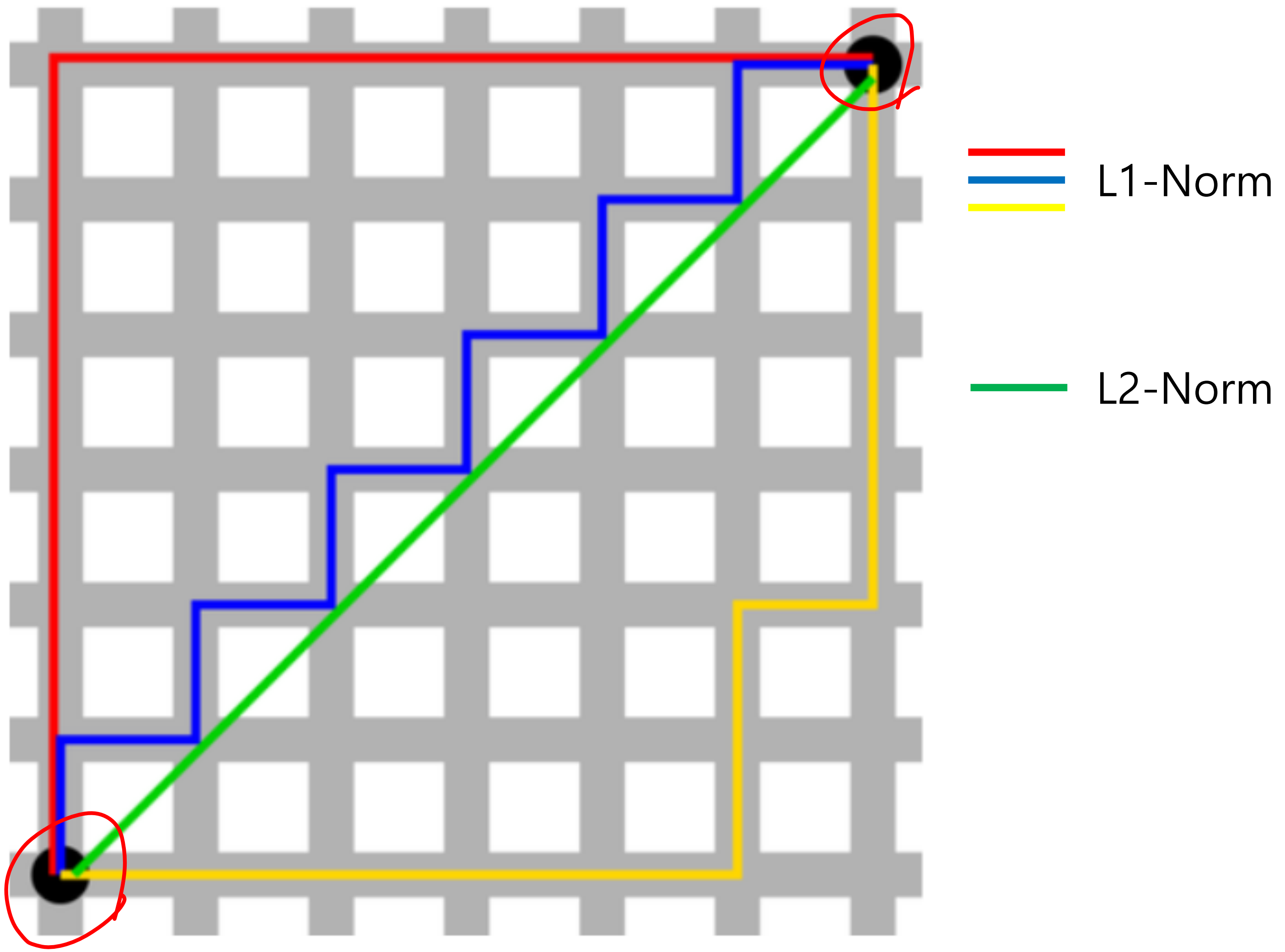
$$= 1 + 5 + 7 = 12$$

L2-Norm

P와 q의 직선거리

$$\|\boldsymbol{x}\|_2 := \sqrt{x_1^2 + \cdots + x_n^2}.$$

Difference with L1-Norm and L2-Norm



About Loss

L1-Loss

MAE

$$L = \sum_{i=1}^{n} |y_i - f(x_i)|$$

L2-Loss

MSE

$$L = \sum_{i=1}^{n} \left(y_i - f(x_i)\right)^2$$

Regularization

Loss

L1-Regularization

learning rate.

$$Cost = \underline{\frac{1}{n}\sum_{i=1}^{n}\left\{L\left(y_i, \hat{y}_i\right)\right.} + \frac{\lambda}{2}|w| \Big\}$$

↳ weight가 너무 커지지 않게끔 일정의 패널티 부과.

L2-Regularization

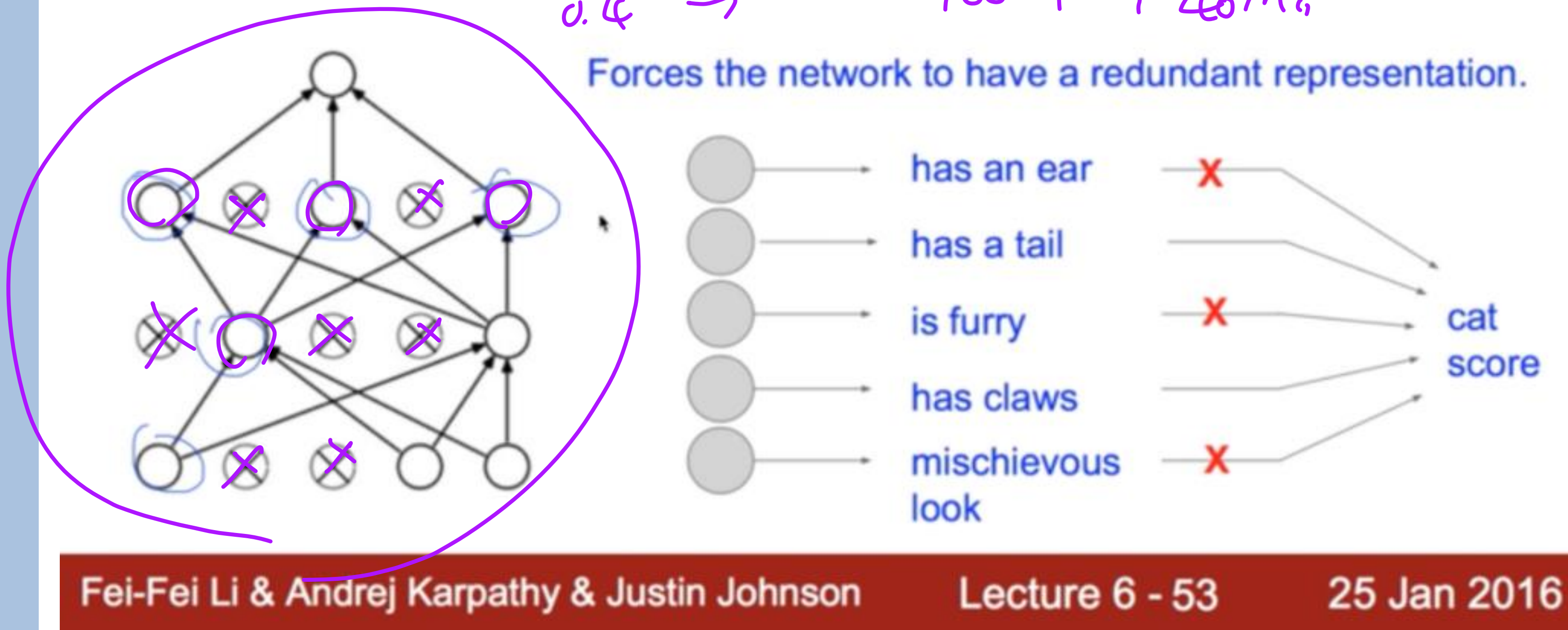
Loss

$$\text{Cost} = \underbrace{\underline{\sum_{i=0}^{N}(y_i - \sum_{j=0}^{M} x_{ij}W_j)^2}}_{\text{Loss function}} + \underbrace{\lambda\sum_{j=0}^{M} W_j^2}_{\text{Regularization Term}}$$

Dropout

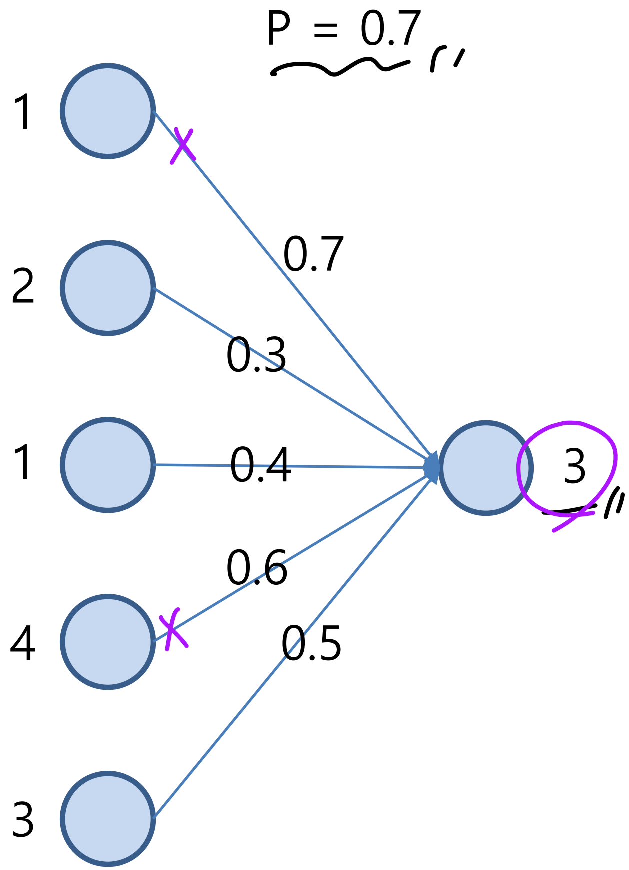
0.5 → 노드 100개 → 50개

0.4 → 100개 → 60개.



- Dropout은 training set에만 적용한다.

Dropout



① Dropout ×

$0.7 + 0.6 + 0.4 + 2.4 + 1.5 = 5.6$

② Dropout ($P = 0.4$)

$0.6 + 0.4 + 1.5 = 2.5$

Thank you..!!!

Reference: https://wooono.tistory.com/221