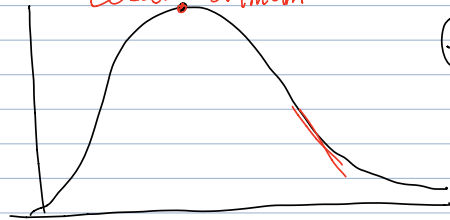


Lab-07-1 Tips

1. Maximum Likelihood Estimation (MLE)

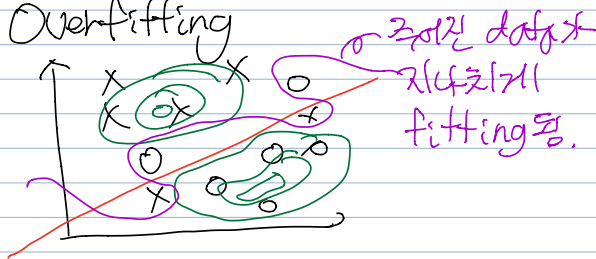
베르누이 분포를 n 번 binomial이 됨.

$$K \sim B(n, \theta) \quad P(K=k) = \binom{n}{k} \theta^k (1-\theta)^{n-k}$$



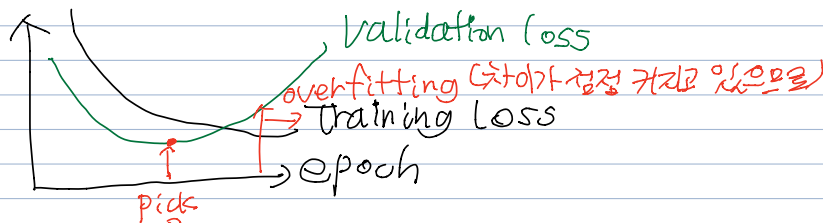
$$\theta \leftarrow \theta - 2 \nabla_{\theta} L(x; \theta)$$

2. Overfitting



overfitting을 방지하기 위하여
로 observations을 나눔.

Training set	validation set develop set	Test set
0.8	0.1	0.1



많은 data를 모으거나 features의 적의면 overfitting을 막을 수 있음.

regularization = overfitting을 막는 방법 중 하나,

early stopping (validation loss가 더 이상

낮아지지 않을 때), network size를 줄인다,

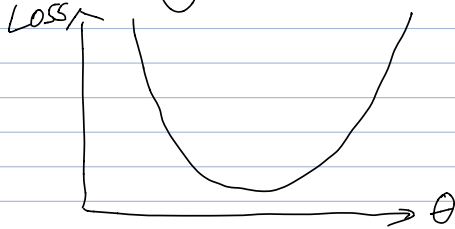
weight의 크기를 제한한다,

dropout, batch normalization

3. DNN 훈련과정

- ① neural network architecture를 만든다.
- ② 훈련시키고 model이 over-fitting이 되는지 확인한다.
(over-fitted되지 않았으면 model size를 증가시키고 over-fitted되었으면 regularization 한다.)
- ③ 2단계를 반복한다.

4. Learning rate



$$\theta \leftarrow \theta - \alpha \nabla_{\theta} L(x; \theta)$$

learning rate가 너무 크면 cost가 발산함.

learning rate가 너무 작으면 cost 값이 거의 변하지 않음.

처음에 0.1로 설정하고 조절한다.

5. Data Preprocessing

$$x'_j = \frac{x_j - \mu_j}{\sigma_j}$$

$$\mu = x_{\text{train}}.mean(dim=0)$$

$$\sigma = x_{\text{train}}.std(dim=0)$$

$$\text{norm_x_train} = (x_{\text{train}} - \mu) / \sigma$$

$$\text{model} = \text{MultivariateLinearRegression}(\text{Model})$$

$$\text{optimizer} = \text{optim.SGD}(\text{model.parameters}(), lr=1e-1)$$

Lab-07-2 MNIST Introduction

for X, Y in data_loader:

$X = X.view(-1, 28 * 28)$

1 epoch: training set 전체 한번 돌리기

batch size: 한 번의 epoch에 사용될 train data 수

iteration: $\frac{\text{train data 수}}{\text{batch size}}$