

Overfitting solution

1. Data engineering

학습에 사용될 데이터셋을 다시 분석하여 불필요한 feature가 있는지, outlier가 존재하는지 파악한 후 적절히 처리한다.

2. Cross validation

모델을 학습시킬 때 모든 데이터가 최소 한 번은 validation set으로 쓰이도록 하는 기법이다. 이를 통해 우연의 일치로 noise가 다수 포함된 training set만을 학습하게 되는 경우를 방지한다.

3. Regularization

Linear regression, Logistic regression, Support vector machine 등 머신러닝 알고리즘에 활용되는 기술로 cost function (C_0)에 weight regularization term을 더해주는 것이다. 이를 통해 cost 뿐만 아니라 가중치 또한 작아지는 방향으로 학습하므로 가중치가 비 이상적으로 커지는 것을 막을 수 있다. (weight decay). 분모의 2는 미분과 관련이 있다.

3. Early stopping

모델을 학습하는 과정에서 validation set의 정확도가 최대에 이르렀을 때, 학습을 중단한다. 혹은, training cost와 validation cost에 큰 차이가 발생하는 순간 학습을 중단한다.

4. Dropout

Deeplearning에서 대표적으로 사용되는 방법으로 training 과정에서 일부 neuron은 학습을 생략하는 것이다. (validation 과정에서는 모든 neuron을 이용한다.) 이를 통해 weight들이 동화 (co-adaptation) 되는 것을 막을 수 있다.

5. Dropconnect

학습하는데 Neuron을 생략하는 것이 아니라 weight를 생략한다. Dropout과의 차이점은 dropout은 neuron과 연결된 weight들이 함께 학습을 생략하는 반면에 dropconnect는 해당 weight만이 학습을 생략한다.

6. Batch normalization

각 layer마다 neuron 값의 distribution 차이가 커지는 것을 막기 위해 각 layer의 neuron 값을 평균을 0으로 하도록 normalization한다. Gradient Vanishing / Gradient Exploding을 막는 방법으로 더 널리 알려져 있다.