이번주는 모델 훈련시 train셋과 test셋, 그리고 validation셋을 나누는 이유에 대해 알아보고 워드 및 pdf로 정리해서 토요일에 이야기 나눠봅시다

기본적으로는 overfitting 방지용



Training set과 Test set 두가지로 나누면 편향현상이 생기므로 이를 보완하기 위하여 Training set에서 validation set을 또 나눠낸다. 보통 7 : 2 : 1 로 나눈다.

Training set은 말 그대로 학습을 시키고 validation set은 일반화가 잘 되었나 확인, Test set은 말그대로 정확도 확인만 한다. Test set도 overfitting되지 않게 유의한다.

그렇게 나눠진 training set과 validation set에 교차검증을 사용한다. 보통 8 : 2로 나눠서 한다.

검증(Training, Validation set 이용한) 종류

1.K-겹 교차검증



훈련set를 k개로 나눠서 validation을 1개지정해서 검증 총 k번 검증하게 된다.

2. Hold-out 교차검증



그냥 무작위로 1번 특정한 비율로 나눠서 검증

3.Leave-P-out 교차검증



K겹에서 test set갯수도 지정