

NN adv 6주차 (Training Error와 Generalization Error 사이 간극을 줄이기) 1-2

1. 교차 검증(Cross-validation)

- 모델 학습 시, 주어진 데이터셋을 여러 개의 부분집합으로 나누어 각각 **훈련용과 검증용으로 교차**하여 선택하여 모델을 여러 번 평가하는 방법이다.
- 가장 대표적인 방법론으로는 **K-Fold Cross Validation**, Hold-out Cross Validation 이 있다.

• 장단점

- 특정 데이터셋에 대한 과적합을 방지할 수 있고, 데이터셋 규모가 적을 경우에는 과소적합을 방지할 수 있다.
- 더욱 일반화된 모델을 생성할 수 있게 된다.
- 단점으로 모델 훈련 및 평가 소요 시간이 더욱 증가할 수 있다. (반복 학습 횟수 증가)

• K-Fold Cross Validation 설명

1. 전체 데이터를 K개 fold로 분할한다.(K는 하이퍼파라미터 값으로 보통 5~10 fold를 사용한다.)
2. 분할된 fold 중 test data로 할당된 적이 없는 fold 1개를 test data로 할당시킨다.
3. 위 2.의 과정을 K번 반복시킨다.
4. K개의 모델 성능 평가 결과값을 평균을 냄으로써 최종 결과값에 활용한다.

→ 이를 통해 모델의 일반화 능력을 더 정확하게 평가할 수 있다.

▼ 참고자료

- 2022 ADsP 한 권으로 끝내기
- <https://heytech.tistory.com/113>