

Machine Learning

Chapter 3 지도 학습(Supervised Learning)



START



Smart Media
스마트미디어인재개발원

- 검증 데이터의 필요성을 이해 할 수 있다.
- 교차 검증 기법을 이해 할 수 있다.
- 교차 검증의 장/단점을 알 수 있다.



Cross validation

(교차검증)



Smart Media
스마트미디어인재개발원

Data



학습



Train Data

평가



Test Data

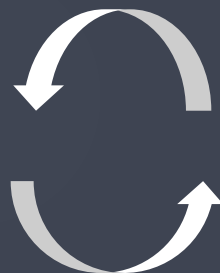
Cross validation(교차검증)

학습



Train Data

ex) 하이퍼파라미터 튜닝



평가



Test Data

테스트 세트에 맞게 학습 될 수 있다.



Cross validation(교차검증)

학습



Train
Data

검증



Validation
Data

평가



Test
Data

Cross validation (교차검증)

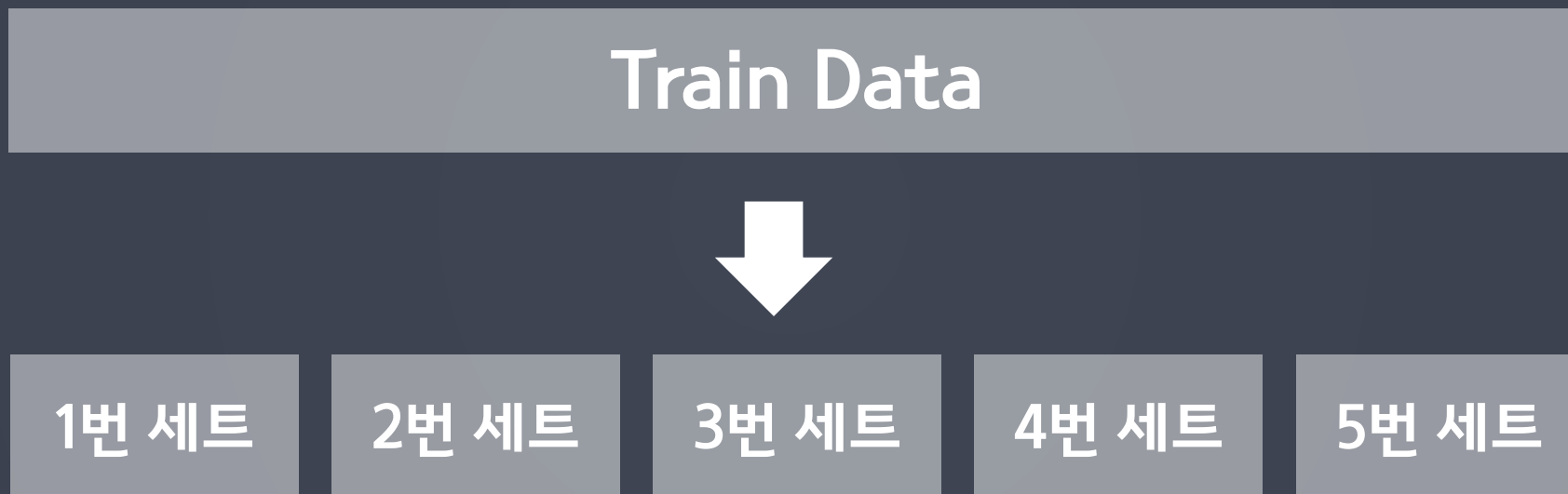
학습-평가 데이터 나누기를 여러 번 반복하여
일반화 에러를 평가하는 방법



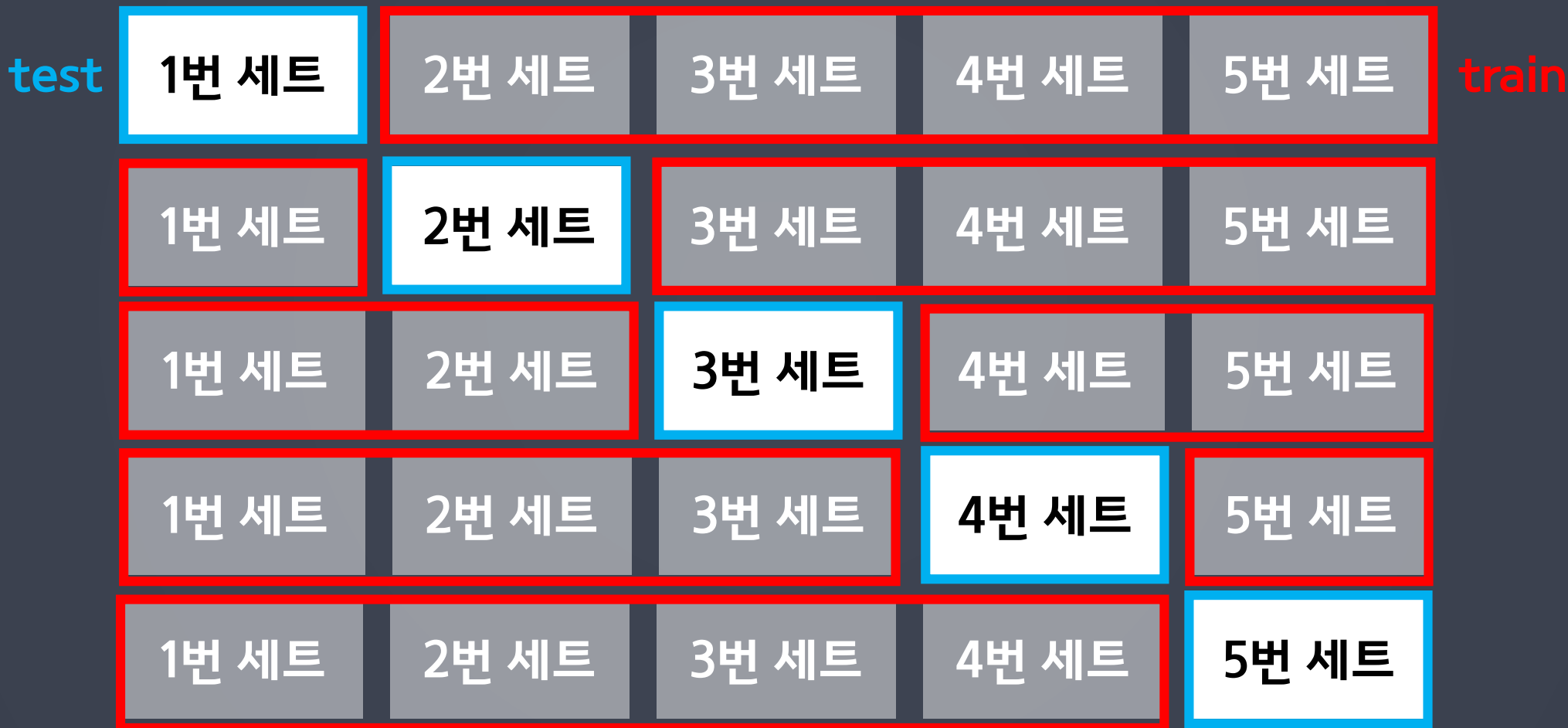
K-fold cross-validation 동작 방법

1. 데이터 셋을 k 개로 나눈다.
2. 첫 번째 세트를 제외하고 나머지에 대해 모델을 학습한다.
그리고 첫 번째 세트를 이용해서 평가를 수행한다.
3. 2번 과정을 마지막 세트까지 반복한다.
4. 각 세트에 대해 구했던 평가 결과의 평균을 구한다.

K-fold cross-validation 동작 방법



K-fold cross-validation 동작 방법



cross-validation 장/단점

- 데이터의 여러 부분을 학습하고 평가해서 일반화 성능을 측정하기 때문에 안정적이고 정확하다. (샘플링 차이 최소화)
- 모델이 훈련 데이터에 대해 얼마나 민감한지 파악가능 (점수 대역 폭이 넓으면 민감)
- 데이터 세트 크기가 충분하지 않은 경우에도 유용하게 사용 가능하다.
- 여러 번 학습하고 평가하는 과정을 거치기 때문에 계산량이 많아진다