

Train, Test, Validation

≡ Type

Preprocessing

Outline

- Train data
 - 모델을 학습하기 위한 데이터
 - 학습 과정
- Test data
 - 학습한 모델을 평가하기 위한 데이터
 - 평가 과정
- Validation
 - 학습한 모델들 중에서 최적의 모델 선정
 - 검증 과정



Train vs Test

Given Data

- If Train 100%
 - 모델이 **overfitting(과적합)**이 될 확률이 높음
 - 학습한 데이터에 대해서만 100%를 보임

- 모델의 과적합을 막기 위해 Test데이터가 필요
 - 8:2의 비율을 가지고 총 데이터 중에서 데이터를 나눔
 - 학습에 쓰이지 않는 데이터

Validation

- Train data하고 Test data하고 비교 했을 때 차이가 발생함
 - 그 차이를 줄이기 위해 Validation 사용
- Train data에서 8:2 비율로 보통 잡는 것이 일반적



