

Part.06

Class Imbalanced Problem

# | Undersampling기법

FASTCAMPUS  
ONLINE

머신러닝과 데이터분석 A-Z

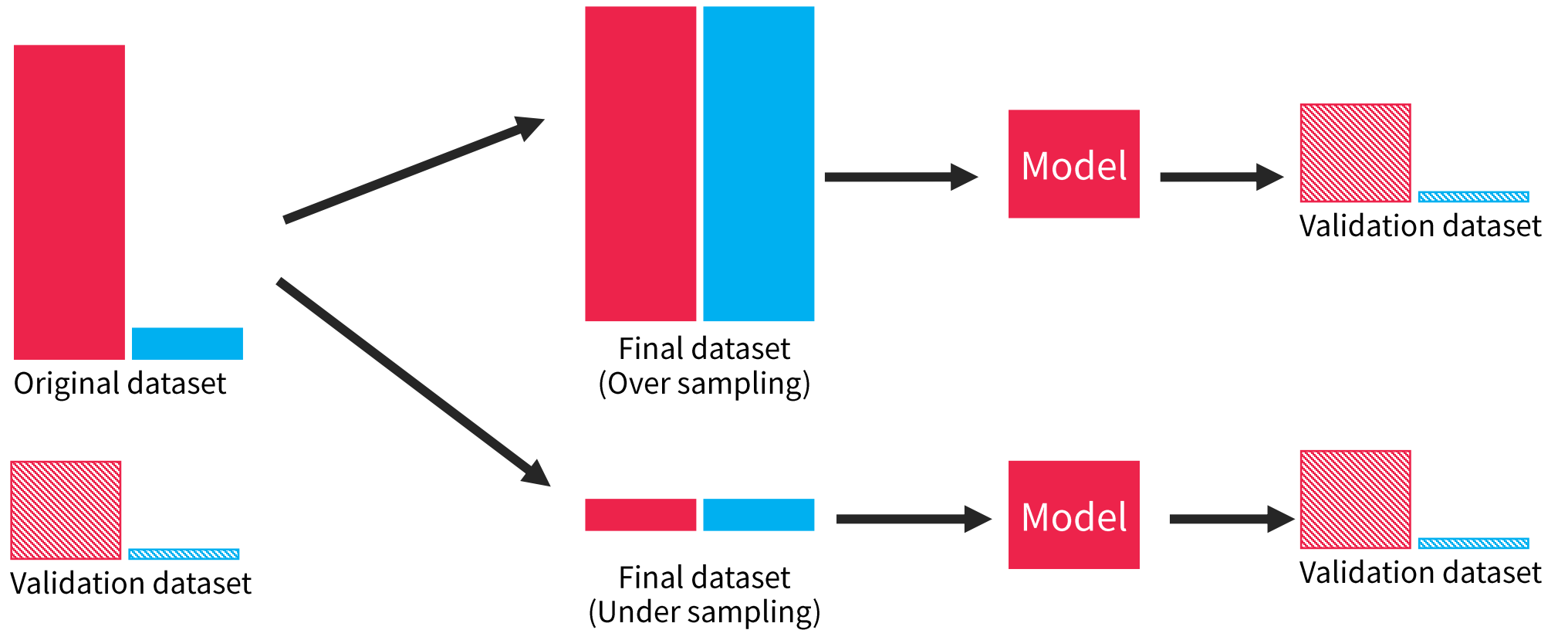
강사. 이경택

# I Undersampling기법

- Class Imbalanced Problem을 해결하기 위한 방법
  - Resampling method
    - Over sampling : 소수의 데이터를 부풀리는 방법
    - Under sampling : 다수의 데이터를 줄이는 방법
    - Hybrid resampling : Over & Under sampling을 결합해서 사용하는 방법
  - Cost-sensitive learning
    - Class의 오 분류에 대한 cost의 가중치를 조절하여 학습하는 방법

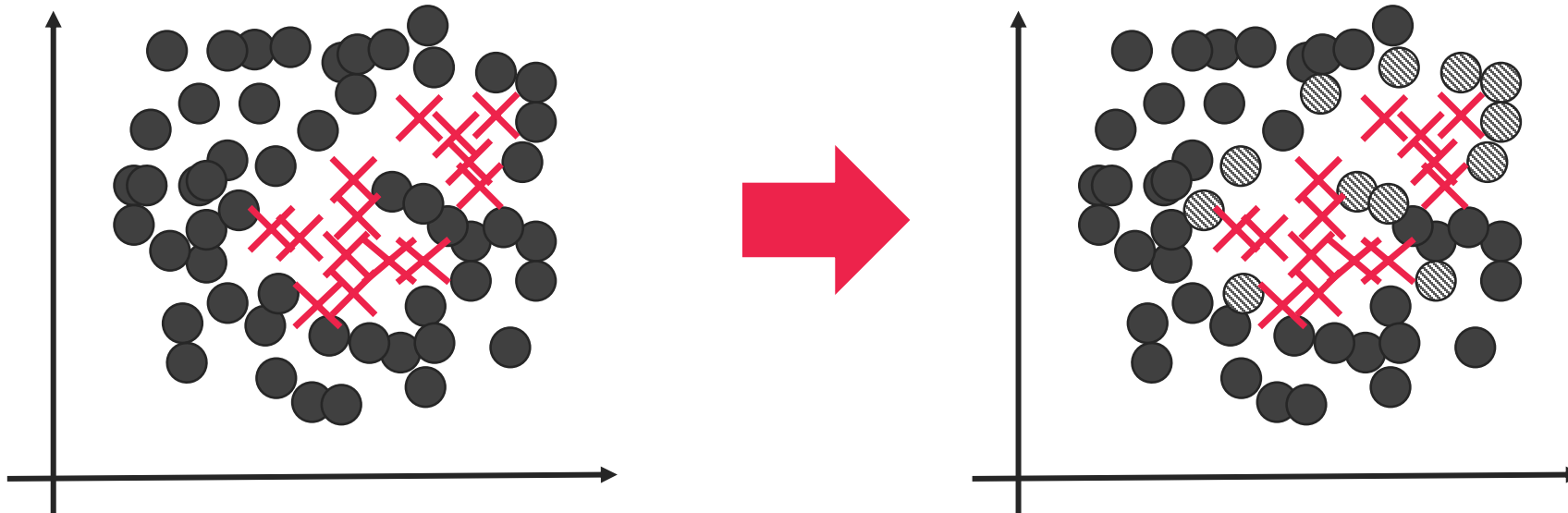
# I Undersampling기법

- Class Imbalanced Problem을 해결하기 위한 방법
  - Resampling method



# I Undersampling기법

- Random Under Sampling
  - Random하게 데이터를 under sampling하는것
- Tomek Links
  - Majority data와 minority data의 거리가 근접해 있을 때 majority data를 제거하는 방법



# I Undersampling기법

## ■ EasyEnsemble

- Majority data와 minority data를 동일 수 만큼 random sampling
- 추출된 데이터를 classifier를 통해 학습
- 위 과정을 K번 반복

## ■ BalanceCascade

- EasyEnsemble과 같이 데이터를 N번 random sampling후 학습
- N-1번째 시행 결과에서 올바르게 분류되는 majority data를 undersampling
- Undersampling 된 수 만큼 negative data를 random sampling하여 추가시켜 N번째 학습 진행

# I Undersampling기법

- Undersampling 장단점
  - 학습에 필요하지 않은 데이터를 제거 줄임으로써 학습 속도 향상을 가져올 수 있음
  - Decision boundary 부근에 있는 데이터를 제거했을 때 학습에 악영향을 끼칠 수 있음

Part.06

Class Imbalanced Problem

# I Hybrid resampling 기법

FASTCAMPUS  
ONLINE

머신러닝과 데이터분석 A-Z

강사. 이경택