

Part.06

Class Imbalanced Problem

| Oversampling기법

FASTCAMPUS
ONLINE

머신러닝과 데이터분석 A-Z

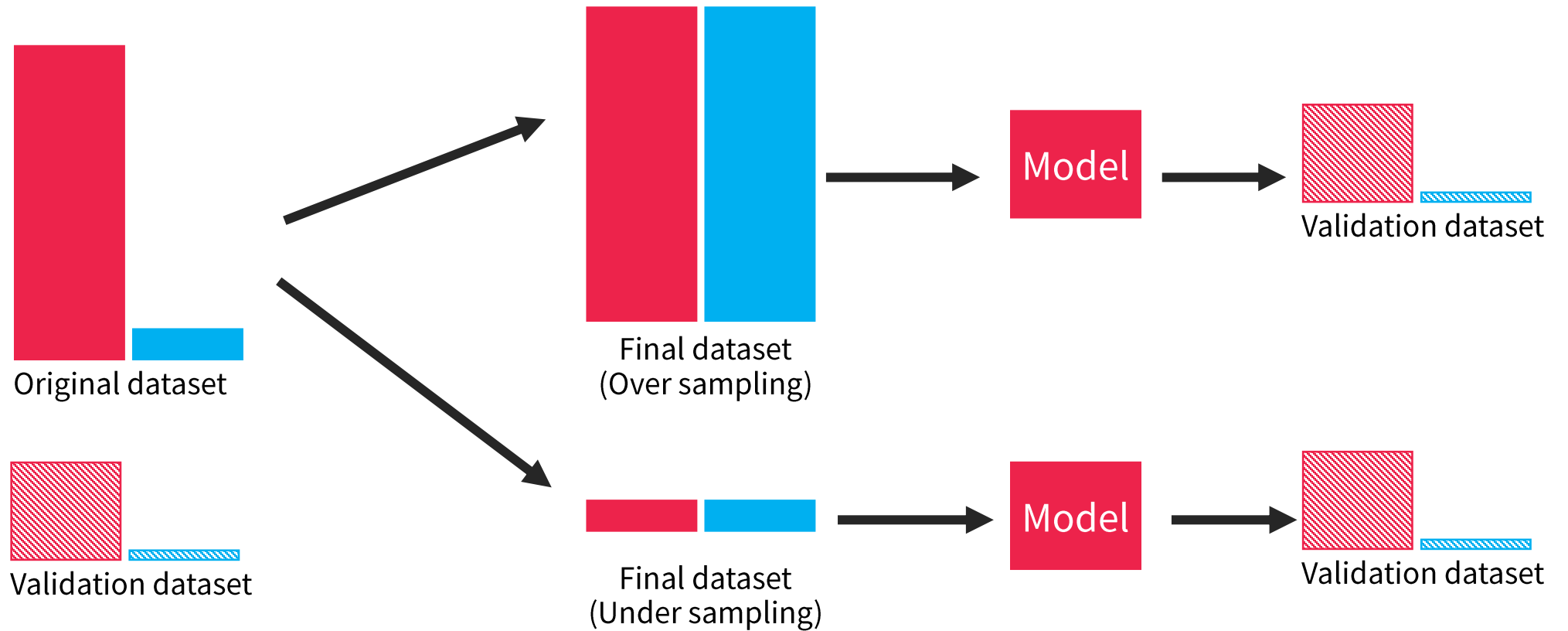
강사. 이경택

I Oversampling기법

- 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 Oversampling기법

- Class Imbalanced Problem을 해결하기 위한 방법
 - Resampling method

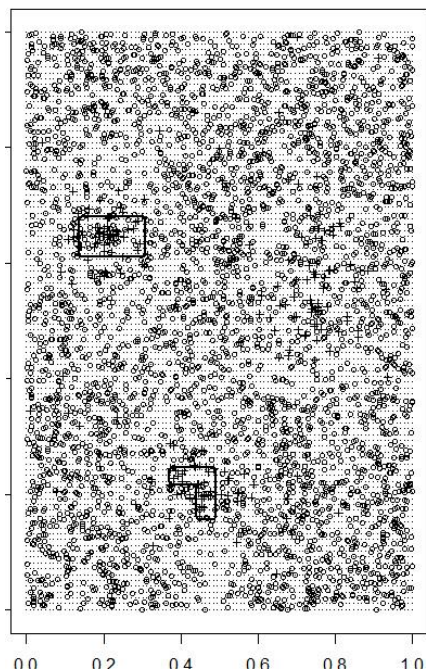


I Oversampling기법

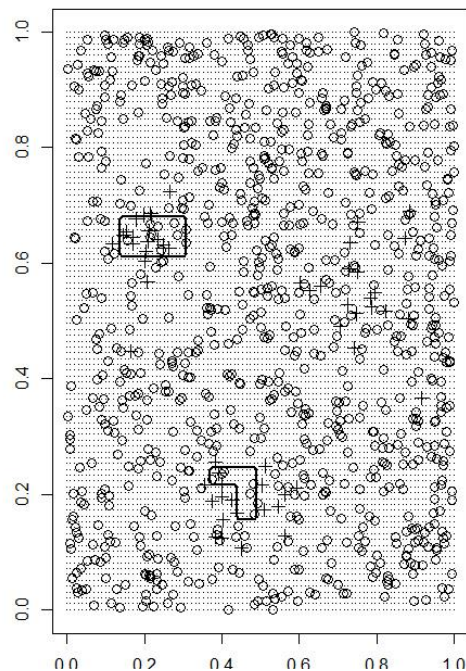
■ Class Imbalanced Problem을 해결하기 위한 방법

- Resampling method

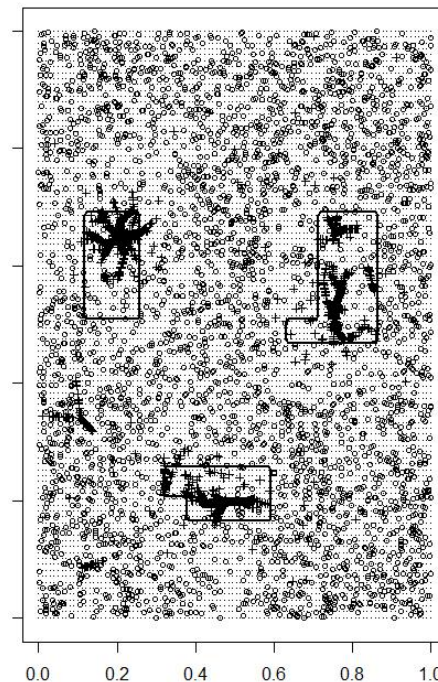
Oversampling를 사용한 예시



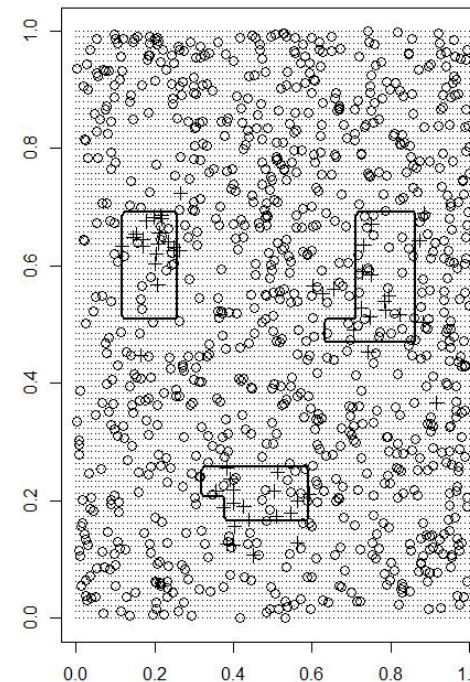
학습데이터



검증데이터
(F-measure : 0.25)



학습데이터

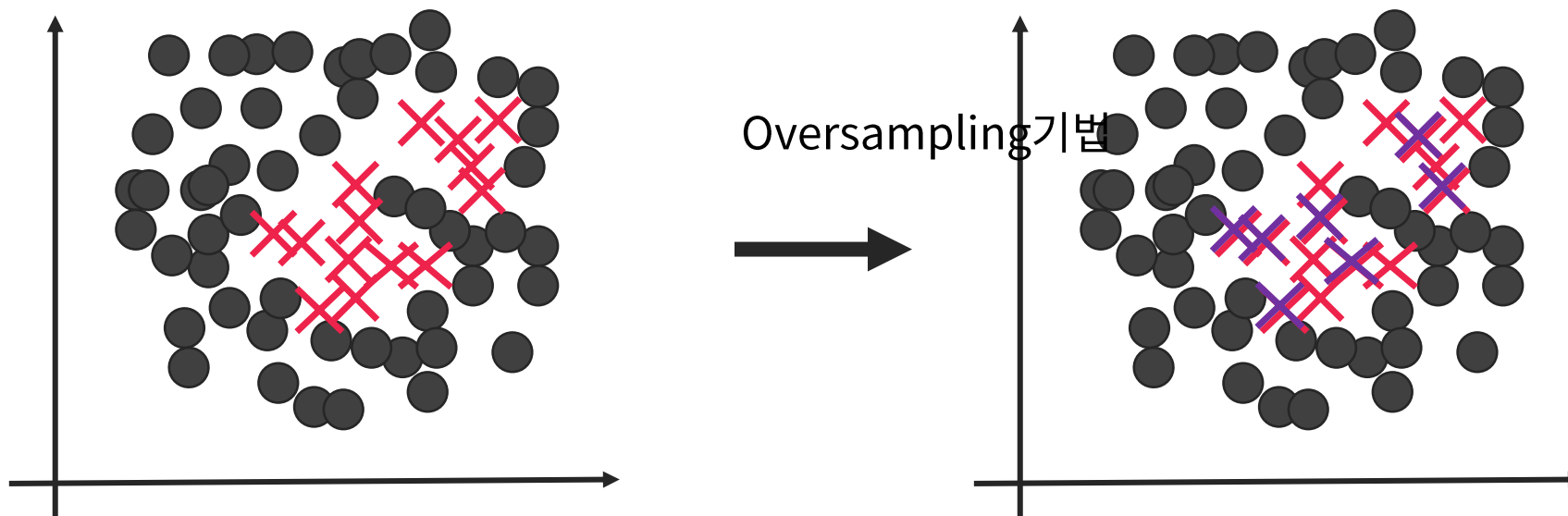


검증데이터
(F-measure : 0.37)

I Oversampling기법

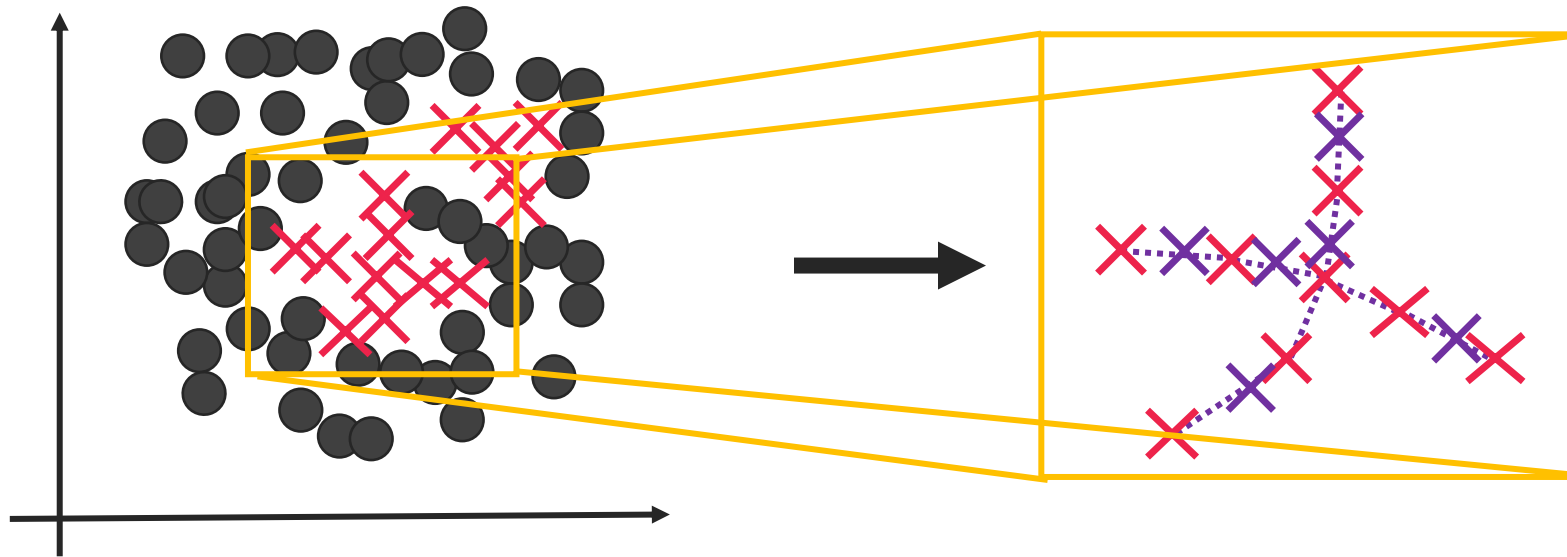
■ Random Over Sampling

- 소수의 데이터를 random하게 sampling시켜 기존데이터와 합치는 방법



I Oversampling기법

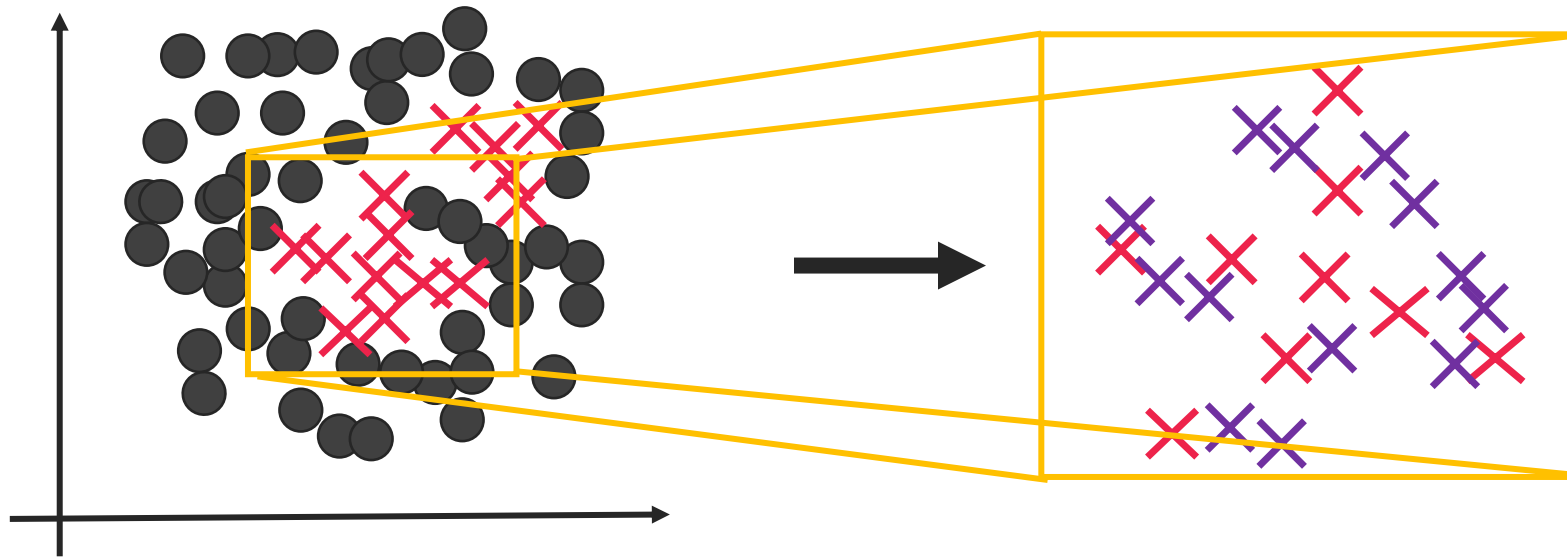
- SMOTE(Synthetic Minority Over-sampling Technique)
 - 소수의 데이터의 sample에 KNN을 적용 후, 샘플과 이웃간의 사이에 random하게 데이터 생성



I Oversampling기법

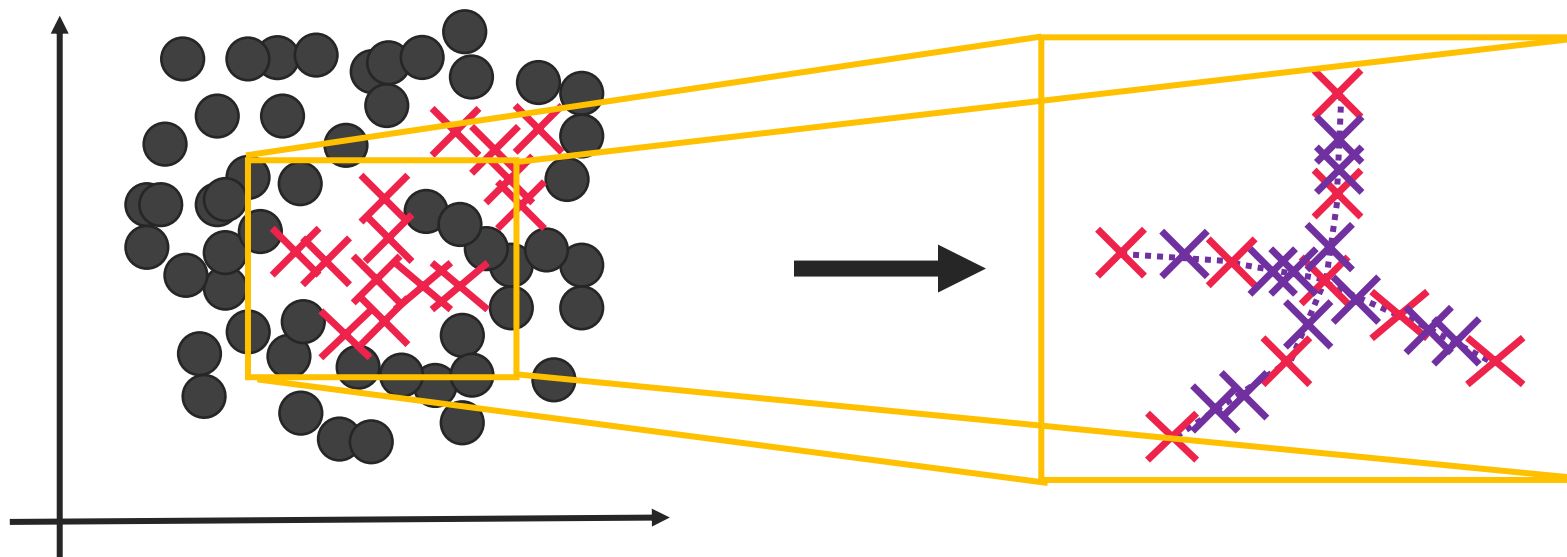
- BLSMOTE(Borderline SMOTE)

- Borderline에 있는 data는 class imbalance problem에 큰 영향을 미친다고 판단하여 해당 dataset에만 SMOTE를 적용



I Oversampling기법

- DBSMOTE(DBSCAN SMOTE)
 - DBSCAN cluster 생성 후 , cluster내에서 SMOTE 적용



Part.06

Class Imbalanced Problem

| SMOTE변형 기법들

FASTCAMPUS
ONLINE

머신러닝과 데이터분석 A-Z

강사. 이경택