

Part.06

Class Imbalanced Problem

| SMOTE변형 기법들

FASTCAMPUS
ONLINE

머신러닝과 데이터분석 A-Z

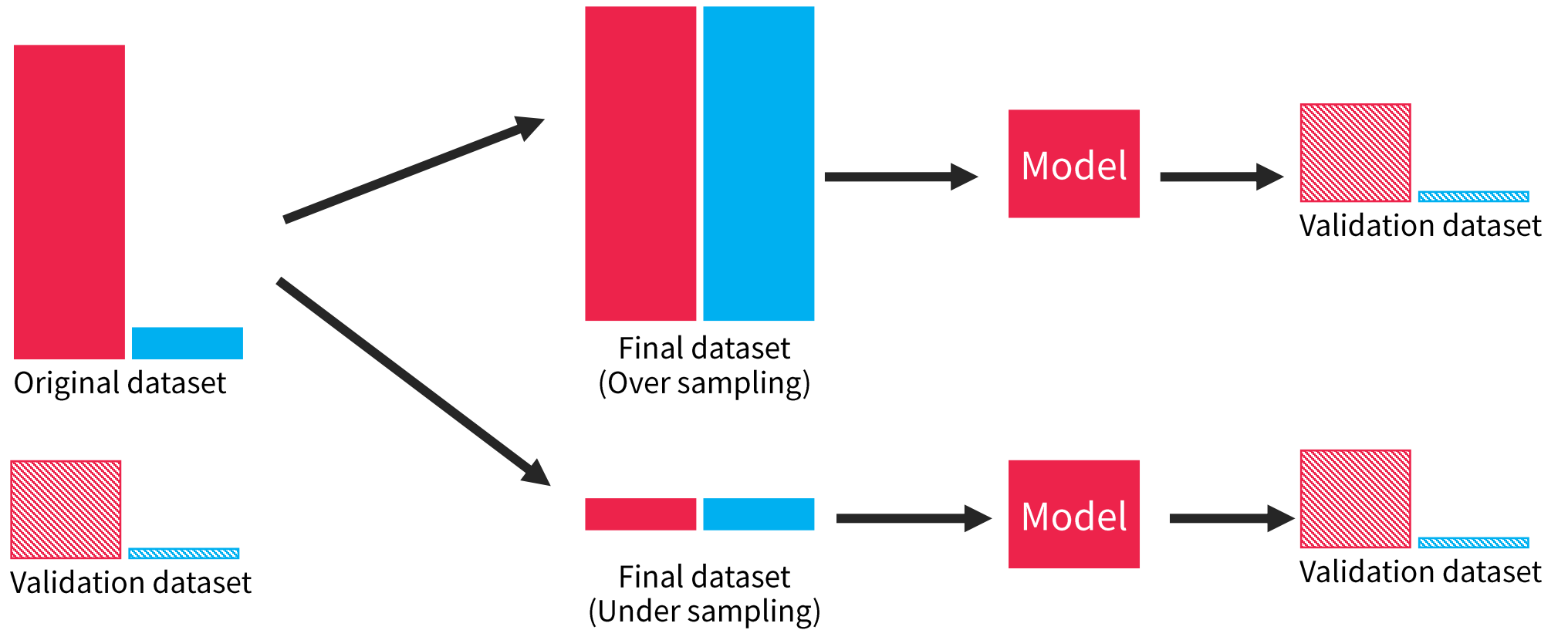
강사. 이경택

ISMOTE변형 기법들

- 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

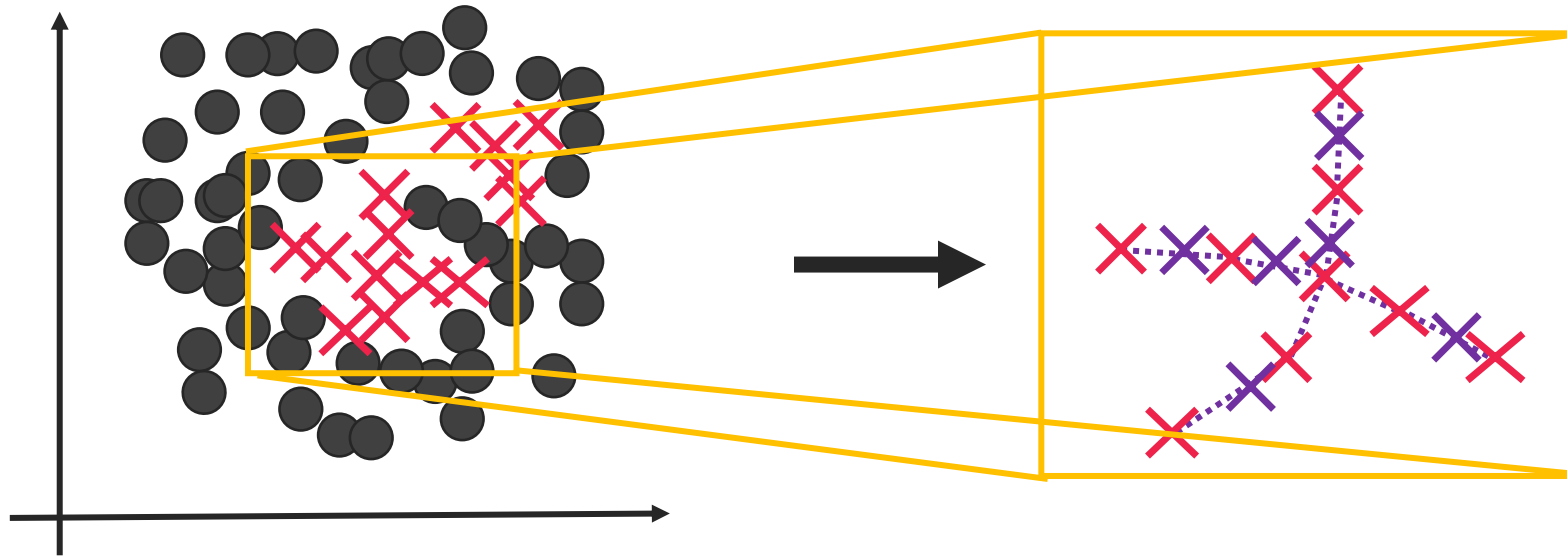
ISMOTE변형 기법들

- Class Imbalanced Problem을 해결하기 위한 방법
 - Resampling method



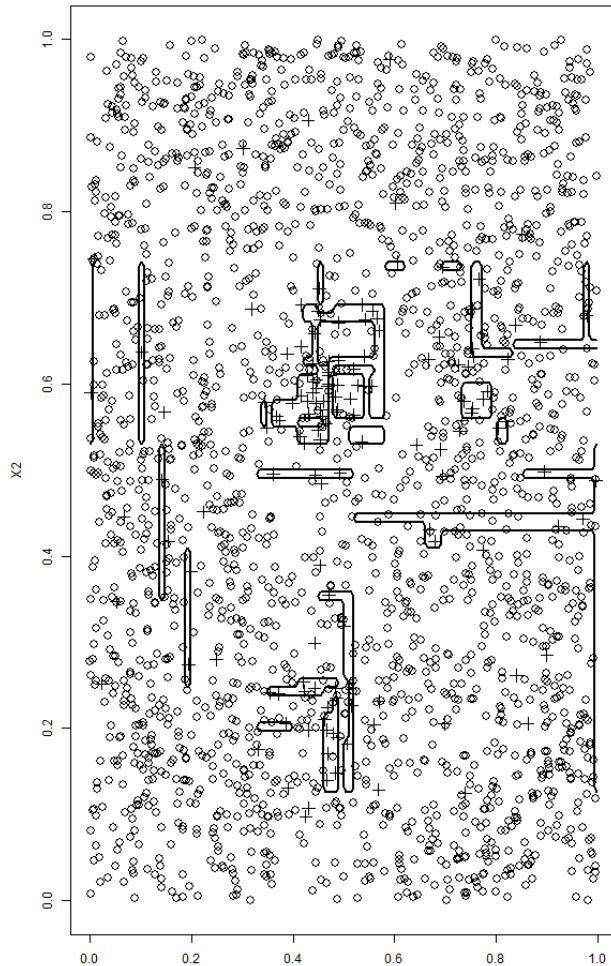
I SMOTE변형 기법들

- SMOTE(Synthetic Minority Over-sampling Technique)
 - minority 데이터마다 가장 가까운 다른 minority data point를 찾는다(KNN을 적용)
 - 두 point를 연결하는 segment상에 존재 하는 data를 random sampling

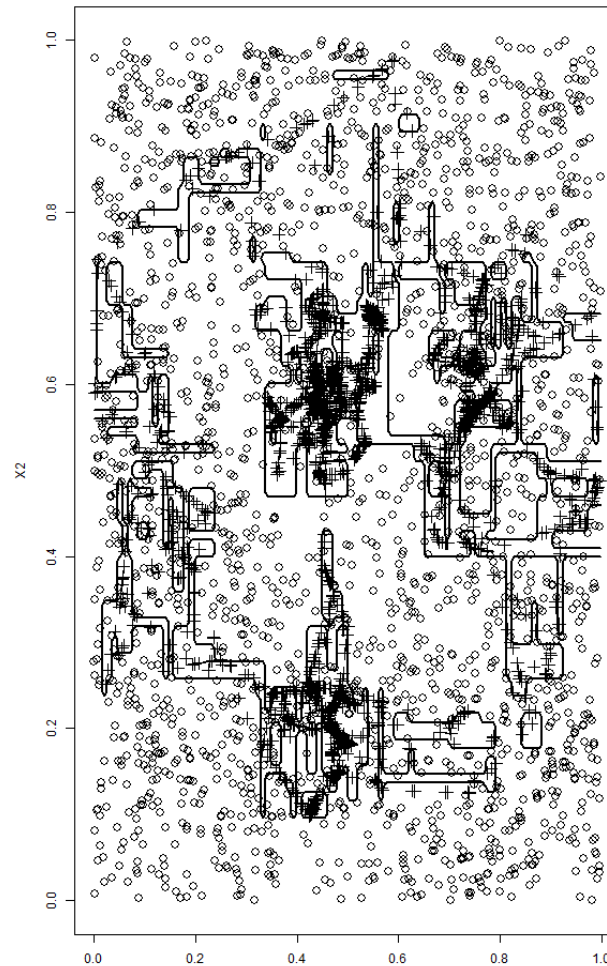


I SMOTE변형 기법들

- SMOTE(Synthetic Minority Over-sampling Technique)



SMOTE적용 전 (0.19)

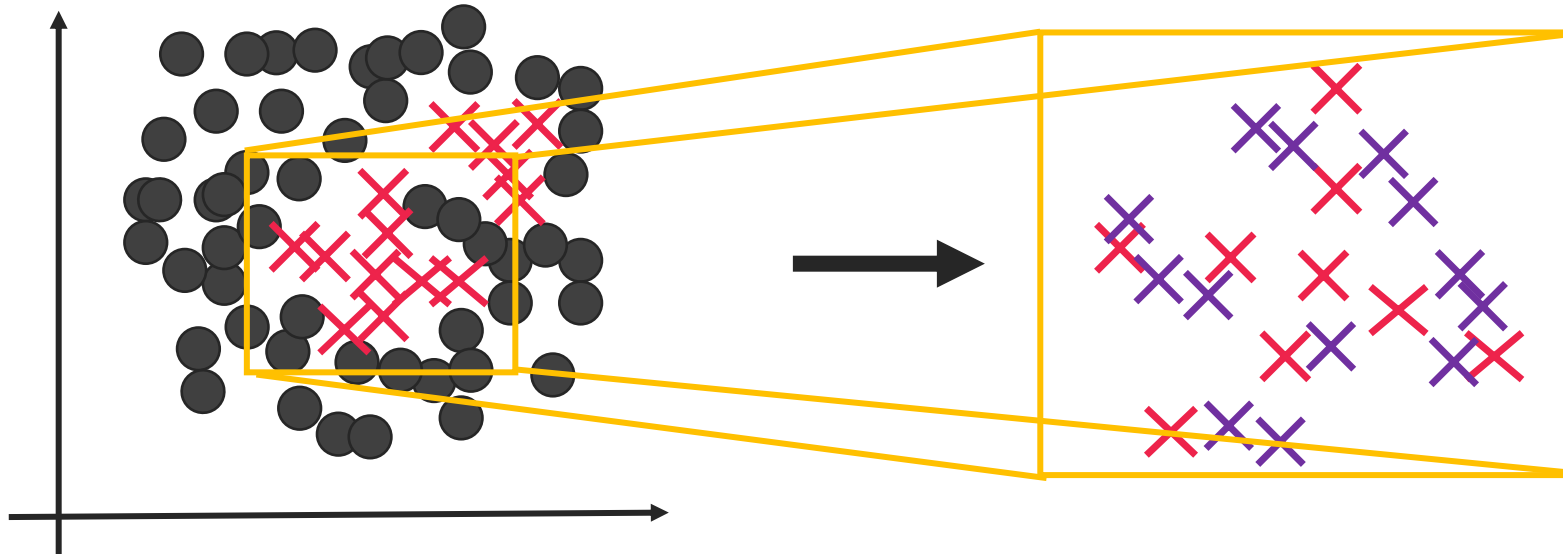


SMOTE적용 후 (0.28)

SMOTE변형 기법들

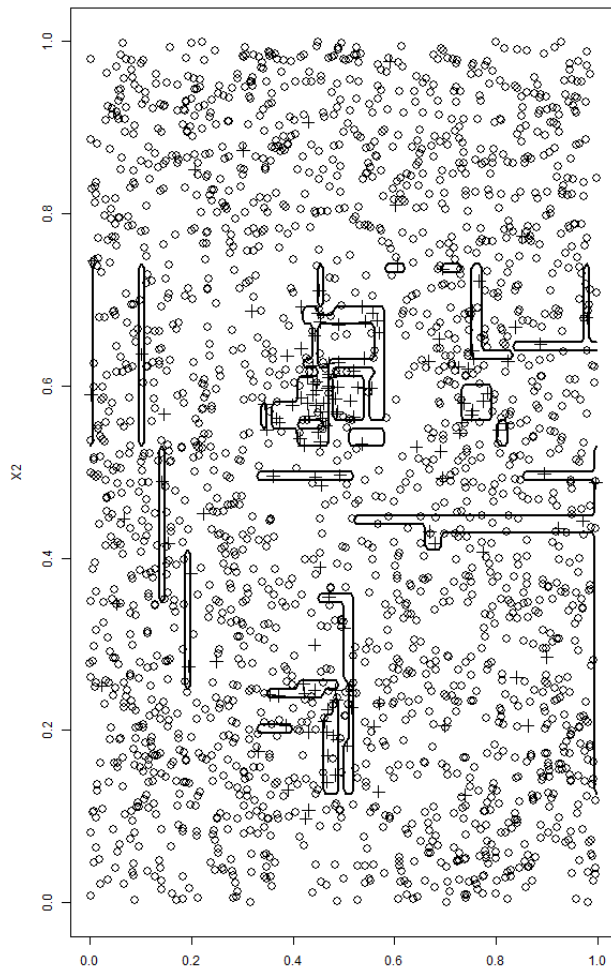
■ BLSMOTE(Borderline SMOTE)

- Minority 데이터에 대하여 세개의 category로 분류(safe, borderline, noise)
- Minority 데이터에 대하여 이웃 데이터를 찾아 그 중 절반이상이 Majority 데이터라면 borderline에 있다고 판단
- Borderline data에 대하여 SMOTE적용

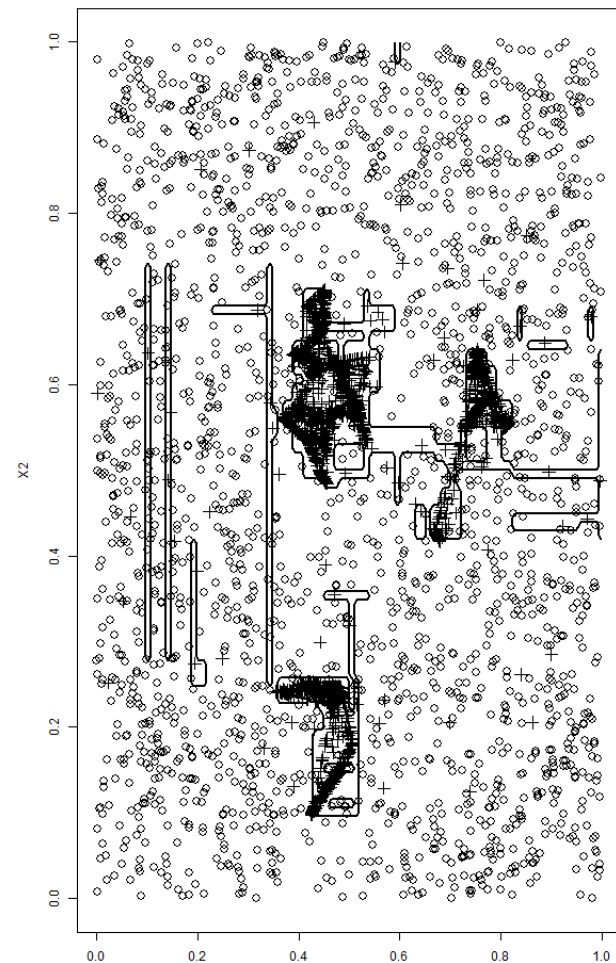


ISMOTE변형 기법들

■ BLSMOTE(Borderline SMOTE)



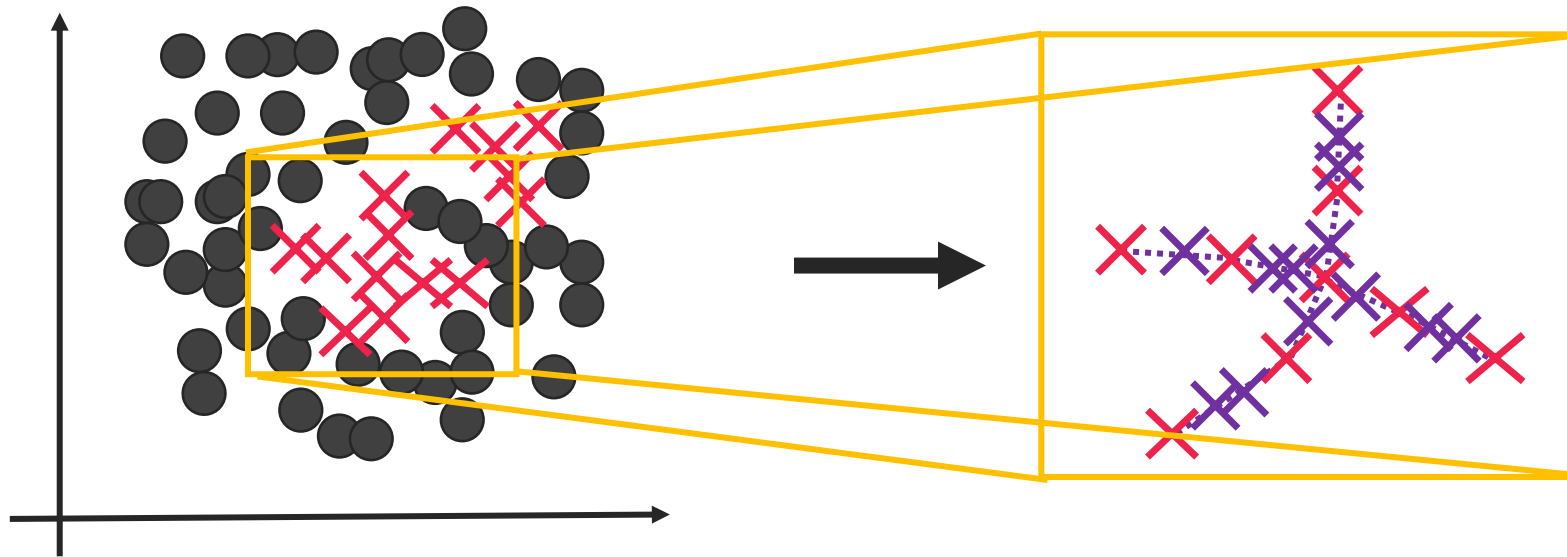
BLSMOTE적용 전 (0.19)



BLSMOTE적용 후 (0.22)

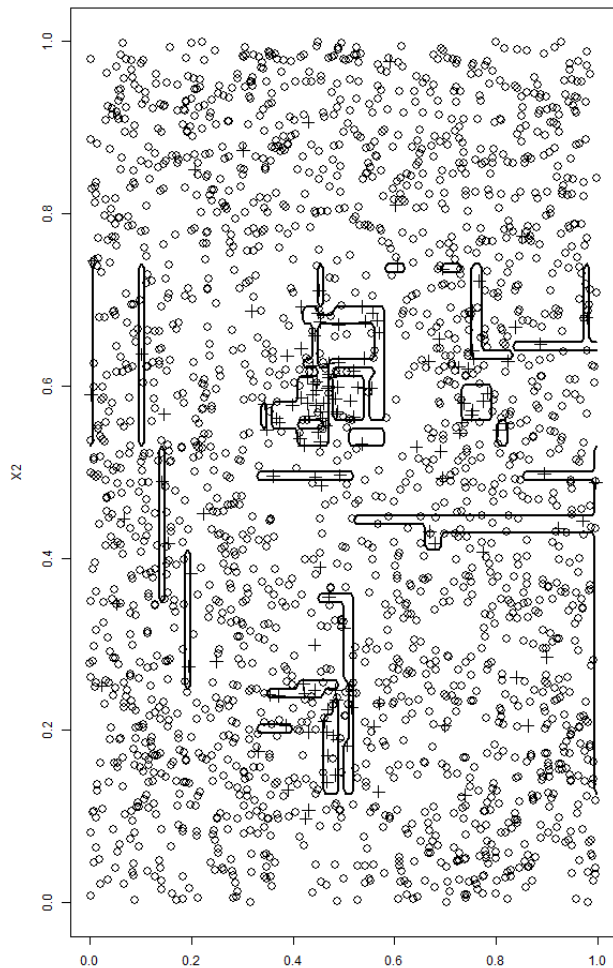
ISMOTE변형 기법들

- DBSMOTE(DBSCAN SMOTE)
- Minority data에 대하여 DBSCAN cluster 생성
- 각 cluster의 centroid와 data 사이의 segment 상에서 random sampling

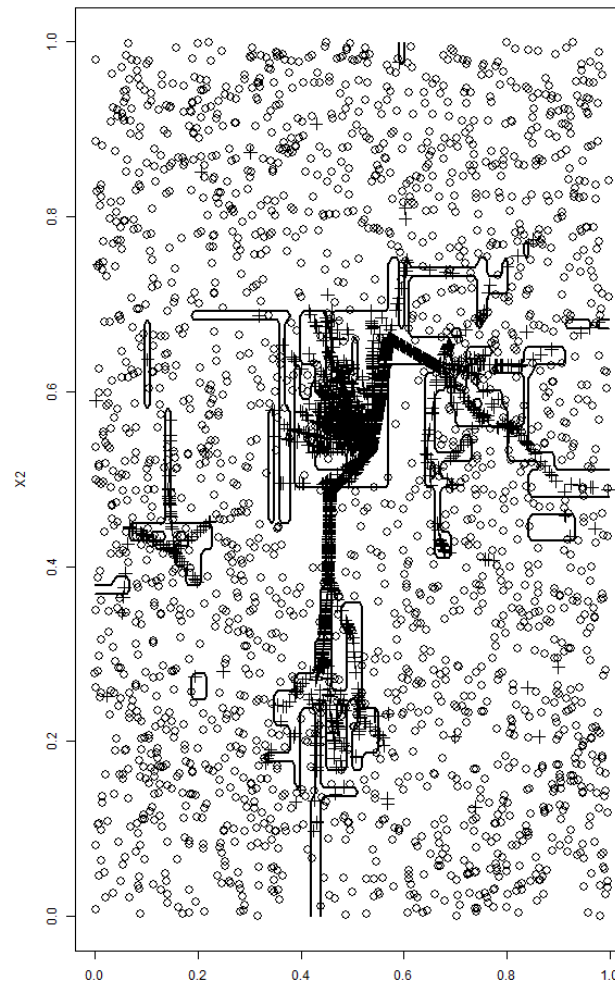


ISMOTE변형 기법들

■ DBSMOTE(DBSCAN SMOTE)



DBSMOTE적용 전 (0.19)

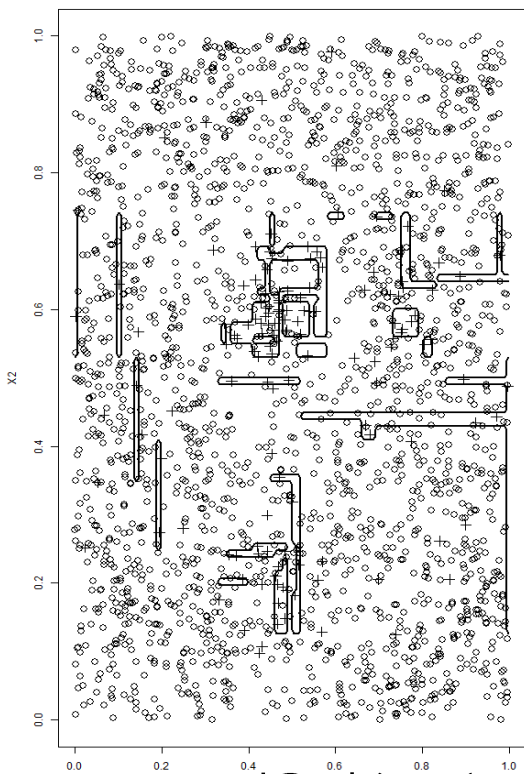


DBSMOTE적용 후 (0.35)

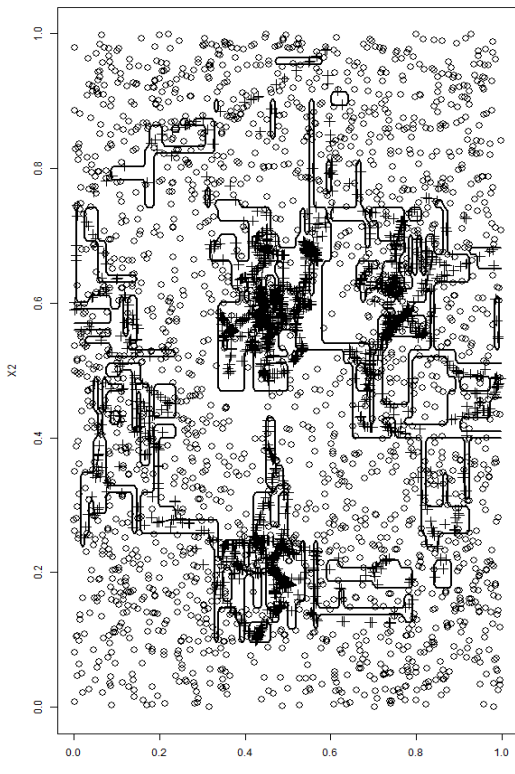
I SMOTE변형 기법들

■ Oversampling의 단점

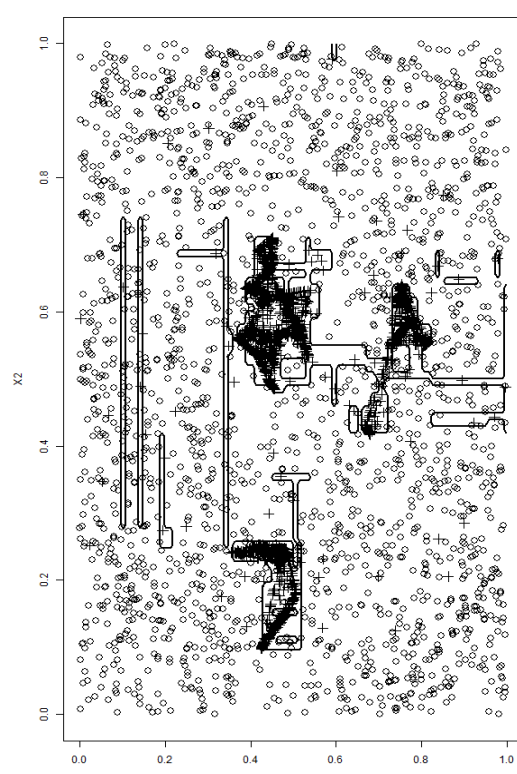
- 데이터마다 성능 편차가 심하며, minority 데이터에 대하여 over-generalization되는 경향성이 있음



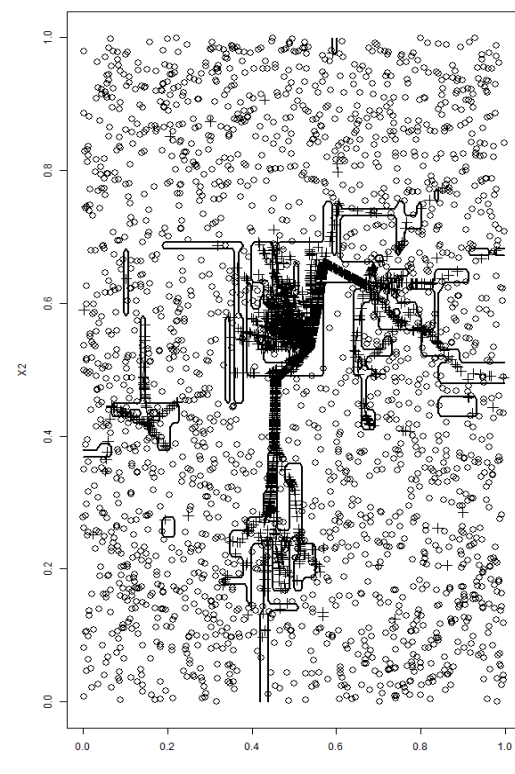
SMOTE적용 전 (0.19)



SMOTE적용 후 (0.28)



BLSMOTE적용 후 (0.22)



DBSMOTE적용 후 (0.35)

Part.06

Class Imbalanced Problem

| Undersampling기법

FASTCAMPUS
ONLINE

머신러닝과 데이터분석 A-Z

강사. 이경택