오버샘플링 기법이란?
: 분류 문제에서 자주 발생하는 Imbalance(클래스 불균형) 문제를 해결하는 방법 중 하나.
오버샘플링의 종류는 주로 SMOTE에서 확장된 방법들이 많고 최근에는 GAN을 활용한 오법샘플링 기법도 많이 연구되고 있음

오버샘플링이 필요한 이유?
: 모델은 적은 수의 클래스(Minority)의 분포를 제대로 학습하지 못하게 돼서 모델은 많은 수의 클래스(Majority)의 분포에 과대적합되어 어떤 데이터가 들어오더라도 Major Class로 분류하게 되는 문제가 발생하기 때문에 모델을 학습할 때, 이러한 클래스 불균형을 조정해야함.

■ SMOTE(Sythetic Minority Over-Sampling Technique)
임의의 Minor 클래스로부터 인근 Minor 클래스 사이에 새로운 데이터를 생성하는 것. 인근 K개의 X와 원래 X 사이에 임의의 새로운 데이터 X'를 생성하는 개념.

■ Borderline-SMOTE
다수와 소수 클래스가 인접해있는 경계선인 Borderline에 있는 소수 클래스의 데이터에 대해 SMOTE 적용하는 방법

이론)
보더라인 스코트는 임의의 소수 클래스 데이터 X에 가장 근접한 K개의 데이터를 찾는다. 인근 K개 클래스 수에 따라서 [Danger, Safe, Noise] 3개로 구분하고 SMOTE방법을 적용.

[Danger] : K/2 < K'< K ----> SMOTE 적용
X 주변 데이터 중 절반 이상이 다수 클래스일 경우 위험하다고 간주하고 SMOTE를 적용
[Noise] : K = K'
X에 가장 근접한 K개의 클래스가 전부 Major일 경우 Noise라 생각하고 SMOTE를 미적용
[Safe] : 0 <= K'<= K/2
절반 이상이 Minor일 경우 Safe라 생각하고 SMOTE 미적용

■ ADASYN(Adaptive Synthetic Sampling)
아다신 방법은 모더라인 스모트에서 응용된 알고리즘이다. 기존의 보더라인 스모트 방법이 경계선 근처에서 [Danger, Safe, Noise] 3가지 경우로 판단하여 SMOTE를 적용했다면, 아다신은 가중치를 사용하여 SMOTE를 적용히는 방법이다.

1) Ratio = Major 수 / K개
2) Scaling
3) 다수 클래수 수 - 소수 클래스 수 = G
4) Scaling*G 두 값을 곱하면 생성할 샘플수가 나온다.

요약하자면, 인접한 다수 클래스의 비율에 따라 SMOTE를 다르게 적용
https://dining-developer.tistory.com/27