

2021년 제 9회 통계세미나

고려대학교 통계연구소와 BK21 통계학교육연구팀이 다음과 같이 공동으로 세미나를 개최하오니 많은 참여 바랍니다.

일시 : 2021년 9월 24일 (금) 오전 10시

장소 : 온라인(ZOOM)으로 진행

연사 : 정현우 (김경희 교수 연구실 소속)

The Effect of Rebalancing on LDA in Imbalanced Classification

<Abstract>

In binary classification, class imbalance is undesirable in that it may worsen the performance of a classifier. One of the remedies to handle this problem is rebalancing with the optimal rate. Theoretical derivation of the rate is not usually considered and often empirically detected, because it is complex and depends on the classifier. To simplify this, we used a linear discriminant classifier, deriving the theoretical optimal rate that maximizes the Matthews Correlation Coefficient (MCC) and F1 score assuming normality. We showed that adjusting the size of each class to be equal is not always the best solution. Instead, we found that there exists the optimal rate depending on the level of class imbalance and the Mahalanobis distance between two classes. Conducting extensive simulation studies and real data analyses, we confirmed that rebalancing with the optimal rate improves the test MCC and F1 score. These findings suggest that with a careful consideration on the level of class imbalance and the separability between two classes, we can achieve better classification results in presence of class imbalance.

고려대학교 통계연구소
BK21 통계학교육연구팀