

## 2023년 제 7회 통계세미나

고려대학교 통계연구소와 BK21 통계학교육연구팀, 그리고 DS+가 다음과 같이 공동으로 세미나를 개최하오니 많은 참여 바랍니다.

일시 : 2023년 5월 12일(금) 오전 11시

장소 : 고려대학교 정경관 206호

연사 : 안정연 교수 (KAIST 산업및시스템공학과)

### Double data piling and negatively ridged classifiers in high dimensions

#### <Abstract>

Data piling refers to the phenomenon that training data vectors from each class project to a single point for classification. While this interesting phenomenon has been a key to understanding many distinctive properties of high-dimensional discrimination, the theoretical underpinning of data piling is far from properly established. In this work, high-dimensional asymptotics of data piling is investigated under a spiked covariance model, which reveals its close connection to the well-known ridged linear classifier. In particular, by projecting the ridge discriminant vector onto the subspace spanned by the leading principal component directions and the maximal data piling vector, we show that a negatively ridged discriminant vector can asymptotically achieve data piling of independent test data, essentially yielding a perfect classification. The second data piling direction is obtained purely from training data and shown to have a maximal property. Furthermore, asymptotic perfect classification occurs only along the second data piling direction. This interesting phenomenon is shown to also occur in multi-category classification problems, in which the second data piling subspaces are estimated by negatively ridged discriminant subspaces. We demonstrate that negative ridge parameters can be optimal in classification of well-known image and microarray datasets.

고려대학교 통계연구소

BK21 통계학교육연구팀



고려대학교 통계연구소  
Institute of Statistics



4단계 BK21 통계학교육연구팀  
BK21 FOUR R&E Team for Statistics



Data Science PLUS  
데이터사이언스 플러스