

2023년 제 1회 통계세미나

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

일시 : 2023년 3월 10일(금) 오전 11시

장소 : 고려대학교 정경관 508호

연사 : 김동하 교수 (성신여자대학교 통계학과)

Approaches for solving data contamination issues using deep neural networks

<Abstract>

Learning with contaminated data leads to degradation of model performances, which is observed clearly for deeply structured models due to their enough complexity in memorizing even the contamination. This talk covers handling data contamination problems using deep learning. We consider two scenarios: 1) data with noisy labels and 2) data with anomalies, and introduce two novel methods to solve each.

First, we identify clean-annotated samples when data with noisy labels are given. We find out a new observation of over-fitted deep neural networks that the similarity of the two neighborhood distributions derived by the feature space and original input space depends on the cleanness of the target sample. Based on this finding, we develop a new method to accurately differentiate cleanly labeled data from noisy ones.

Second, we also discuss filtering anomalies from given data when any label information of anomalousness is unattainable. When we train a deep generative model with data with outliers, we notice that the model first memorizes inliers before outliers in the early learning stage. We exploit this finding and devise a powerful and efficient method to identify outliers.

**고려대학교 통계연구소
BK21 통계학교육연구팀**



고려대학교부설통계연구소
Korea University Institute of statistics



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



Data Science PLUS
데이터사이언스 플러스