

國立中央大學

統計研究所

學術演講

主講人：李育杰 教授(中央研究院資訊科技創新研究中心／國立陽明交通大學應用數學系)

講題：Consensus Learning for Sparse Principal Component Analysis

時間：110年12月28日(星期二) 上午11:00 ~ 12:00

地點：中央大學鴻經館 M-429 室

茶會：上午10:30 ~ 11:00 地點：鴻經館 M510 室

ABSTRACT

Learning from data has become the mainstream in modern machine learning applications. The more data we have, our machine learning methods show better results if the data quality is good enough. However, in many real applications, the data owners may not want to share or are not allowed to share the data they have because of privacy issues or legal concerns. To solve this problem, we proposed a consensus learning framework also known as the federated learning named by Google in 2016. This framework has been applied to linear and nonlinear SVM, LASSO as well as PCA. In this work, we will apply this framework to sparse principal analysis (SPCA). The sparse loadings are composed of as few non-zero elements as possible and keep the data variation as much as possible when we project the data onto these sparse loadings. The SPCA is not only an unsupervised dimension reduction algorithm but also helps identify the key explainable variables. Our proposed method, consensus learning SPCA (CLSPCA) will be solved by a distributed optimization algorithm, ADMM. It basically let each data owners train their models separately. There is a central master will collect the models generated by the workers. The central master fuses these models and then broadcasts to each worker let them improve their model. After certain iterations, we will have the resulting model that will be approached to the model generated by pooling all data together. The data owners never share their data with anyone in the whole training process. We will demonstrate our CLSPCA results on the synthetics dataset as well as a public dataset.