

國立中山大學應用數學系

學術演講

講者：羅小華教授

美國哥倫比亞大學統計系

時間：2015/06/03（星期三）15：30～17：30

地點：理學院四樓理 SC 4009-1 室

Interaction-Based Learning for High-Dimensional Data with Continuous Predictors

High-dimensional data, when dealing with continuous variables such as gene expressions in microarray experiments may contain substantial amount of useful but hidden information. This is because, the information, relevant variables and their joint interactions are usually diluted by noises due to a large number of non-informative variables. Consequently, variable selection should play a pivotal role for learning in high dimensional data problems. Most of the traditional feature selection methods, such as Pearson's correlation between response and predictors, stepwise linear regressions and LASSO are among the popular linear methods. These methods are effective in identifying linear marginal effects but are limited in detecting non-linear, non-additive nor higher order interaction effects. It is wellknown that epistasis (gene - gene interactions) may play an important role in gene expression where unknown functional forms are difficult to identify. Motivated by these needs, we introduce a novel nonparametric measure to screen and do feature selection based on information from nearest neighborhoods. We propose a backward elimination algorithm based on this measure which leads to identification of many influential clusters of variables. We also study procedures to combine these groups of individual classifiers to form a final predictor. The nonparametric measure introduced here is flexible, easy to understand and can be easily extended and applied to other areas of Big data and studies.

Why Aren't Significant Variables Automatically Good Predictors

It is noticed recently that newly discovered significant variables from GWAS studies are not very useful in improving prediction rate. Why so? We offer statistical insights by clarifying what makes variables good for significance and what makes variables good for prediction depend on very different distributional properties. We conclude that progress in prediction requires efforts toward a new research agenda of searching for highly predictive variables sets rather than highly significant variables. We offer an alternative predictive approach that is not limited by statistical significance.

中山大學應用數學系

敬請公告！歡迎參加！

應用數學系：<http://math.nsysu.edu.tw>

校園地圖：<http://web.nsysu.edu.tw/files/11-1000-1503.php?Lang=zh-tw>

交通資訊：<http://www.nsysu.edu.tw/files/90-1000-7.php?Lang=zh-tw>



應用數學系



校園地圖



交通資訊