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
講　題：Are significant variables good predictors? Prediction with I-score
時　間：2016/1/18（星期一）11：10～12：00
地　點：理學院四樓理 SC 4009-1 室
茶　會：10:30 於理 SC 4010 室（系辦公室）

摘要

Thus far, genome-wide association studies (GWAS) have been disappointing in the inability of investigators to use the results of identified, statistically significant variants in complex diseases to make predictions useful for personalized medicine. Why are significant variables not leading to good prediction of outcomes? We point out that this problem is prevalent in simple as well as complex data, in the sciences as well as the social sciences. We offer statistical insights on why higher significance cannot automatically imply stronger predictivity. We also demonstrate that highly predictive variables do not necessarily appear as highly significant, thus evading the researcher using significance-based methods. We point out that what makes vari- ables good for prediction versus significance depends on different properties of the underlying distributions. If prediction is the goal, we must lay aside significance as the only selection standard. We suggest that progress in prediction requires efforts toward a new research agenda of searching for a novel criterion to retrieve highly predictive variables rather than highly significant variables. We offer an alternative approach that was not designed for significance, the partition retention method, which was very effective predicting on a long-studied breast cancer data set, by reducing predicting error rate from 30% to 8%.
(If time permit)

The new alternative allows us to approaching prediction from a theoretical framework grounded in maximizing correct prediction rates. We lay out an objective function for correct prediction rates for which we need to maximize, then consider and ultimately reject an estimator based on the sample analog of the solution to the maximization problem. We demonstrate that the Partition Retention method's I-score is a measure that asymptotically approaches this alternative solution. These show that the I-score can capture highly predictive variable sets and serves as a lower bound for the out sample correct prediction rate. Future research in the avenue of alternative solutions as sample-based measures of predictivity is much desired.

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://math.nsysu.edu.tw/ezfiles/87/1087/img/779/NSYSUMAPmath990705.jpg

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