NYSPI Biostatistics Seminar Series[1]

A road for high-dimensional classification with applications to identify cancer patient subgroups

Yang Feng, Ph.D.
Assistant Professor, Department of Statistics
Columbia University

Tuesday, January 29, 2013

3:30-4:30pm, New PI 6th Floor Multipurpose Room (6602)
Coffee & Cookies provided

Abstract

For high-dimensional classification, it is well known that naively performing the Fisher discriminant rule leads to poor results due to diverging spectra and noise accumulation. Therefore, researchers proposed independence rules to circumvent the diverging spectra, and sparse independence rules to mitigate the issue of noise accumulation. However, in biological applications, there are often a group of correlated genes responsible for clinical outcomes, and the use of the covariance information can significantly reduce is classification rates. In theory the extent of such error rate reductions is unveiled by comparing the misclassification rates of the Fisher discriminant rule and the independence rule. To materialize the gain based on finite samples, a Regularized Optimal Affine Discriminant (ROAD) is proposed. ROAD selects an increasing number of features as the regularization relaxes. Further benefits can be achieved when a screening method is employed to narrow the feature pool before hitting the ROAD.
An efficient Constrained Coordinate Descent algorithm (CCD) is also developed to solve the associated optimization problems. Simulation studies and real data analysis support our theoretical results and demonstrate the advantages of the new classification procedure under a variety of correlation structures.

[1] The PI Biostatistics Seminar Series is held on Tuesdays at New York State Psychiatric Institute. If you are interested in receiving regular announcements for our seminars in the future, or if you need further information, please contact Jina James (jamesji@nyspi.columbia.edu, (212) 543-5589).

Biographical Note

Yang Feng is an assistant professor of Statistics at Columbia University. He got his Ph.D. from the Department of Operations Research & Financial Engineering at Princeton University under the supervision of Professor Jianqing Fan in 2010. His research interest includes High-dimensional statistical learning, Nonparametric and semiparametric methods, Bioinformatics, and Network models.