

Stat Lab Seminars

Robust Optimal Scoring Discriminant Analysis & Discrete Wavelet Packet Transform Based Discriminant Analysis for Whole Genome Sequences

Prof. Hsin-Hsiung Huang (University of Central Florida)

Abstract: We develop a novel classification model which is robust against outliers by introducing a loss function to optimal scoring discriminant analysis. The discriminant vectors and scoring vectors are solved by an iteratively reweighted least squares algorithm. It inherits good properties from these two ideas for reducing the influence of these outliers on estimating the discriminant directions. In the asymptotic stability analysis, we show that the influence function is bounded and discriminant vectors and scoring vectors are both consistent as the percentage of outliers goes to zero. The experimental results are presented to confirm that the robust optimal scoring discriminant analysis is effective and efficient.

In recent years, alignment-free methods have been widely applied for genome sequences comparisons, since these methods compute efficiently and provide desirable phylogenetic analysis results. These methods have been successfully combined with hierarchical clustering methods for finding phylogenetic trees. However, it may not be suitable to apply these alignment-free methods directly to the existing statistical classification methods, since there still lacks an appropriate statistical classification theory for integrating with the alignment-free representation methods. In this article, we propose a discriminant analysis method which uses discrete wavelet packet transform to represent the whole genome sequences and discriminant analysis to

Monday, April 23 at 4:00 PM in SEO 612
--

classify the genome sequences. We show that the proposed alignment-free representation statistics of features approximately follow normal distributions. The data analysis results indicate that the proposed method provides accurate classification in real time.

Drink and snack will be served for this talk.

Monday, April 23 at 4:00 PM in SEO 612