Spring 2019

Colloquium Series

February 1, 2019 at 3pm in Bell Hall 143

Dr. Sangjin Kim

The University of Texas at El Paso

High Dimensional Variable Selection with Error Control

The iterative sure independence screening (ISIS) is a popular method in selecting important variables while maintaining most of the informative variables relevant to the outcome in high throughput data. However, it not only is computationally intensive but also may cause high false discovery rate (FDR). We propose to use the FDR as a screening method to reduce the high dimension to a lower dimension as well as controlling the FDR with three popular variable selection methods: LASSO, SCAD, and MCP. Method. The three methods with the proposed screenings were applied to prostate cancer data with presence of metastasis as the outcome. Results. Simulations showed that the three variable selection methods with the proposed screenings controlled the predefined FDR and produced high area under the receiver operating characteristic curve (AUROC) scores. In applying these methods to the prostate cancer example, LASSO and MCP selected 12 and 8 genes and produced AUROC scores of 0.746 and 0.764, respectively. Conclusions. We demonstrated that the variable selection methods with the sequential use of FDR and ISIS not only controlled the predefined FDR in the final models but also had relatively high AUROC scores.