

The Analysis of Gene Expression Using Statistical Learning Approaches

Di Gao, Ph.D.
Assistant Professor
Sam Houston State University

The classification problem of gene expression level, specifically gene expression analysis, is a major research area in biostatistics. There are several classical methods to solve the classification problem. For example, hypothesis testing with the Benjamini-Hochberg method, Logistic regression, Hidden Markov Chain (HMM) model, and Support Vector Machine (SVM). These classical methods can be applied and perform well with some strict assumptions. One of such is the observations in the data should be independent. However, in gene analysis, these assumptions are violated. The Conditional Random Field (CRF) was introduced to solve the problem. This presentation will discuss some classical methods and the CRF model. Finally, the Least Absolute Selection and Shrinkage Operator (LASSO) penalty, a dimensional reduction method, is introduced to improve the CRF model.