

Title:

Penalized Logistic Regression for Detecting Gene Interactions

Author(s):

Mee Young Park and Trevor Hastie

Technical Report number (Dept. of Statistics, Stanford Univ.):

2006-15

Date:

September 2006

Abstract:

We propose using a variant of logistic regression with L_2 regularization to fit gene-gene and gene-environment interaction models. Studies have shown that many common diseases are influenced by interaction of certain genes. Logistic regression models with quadratic penalization not only correctly characterizes the influential genes along with their interaction structures but also yields additional benefits in handling high-dimensional, discrete factors with a binary response. We illustrate the advantages of using an L_2 regularization scheme, and compare its performance with that of *Multifactor Dimensionality Reduction* and *FlexTree*, two recent tools for identifying gene-gene interactions. Through simulated and real datasets, we demonstrate that our method outperforms other methods in identification of the interaction structures as well as prediction accuracy. In addition, we validate the significance of the factors selected through bootstrap analyses.