

STANFORD UNIVERSITY
DEPARTMENT OF STATISTICS
DEPARTMENTAL SEMINAR

4:15 p.m., Tuesday, June 5, 2007
Sequoia Hall Room 200
(Cookies at 3:45 in 1st Floor Lounge)

Hege Marie Bøvelstad
Department of Mathematics
University of Oslo, Norway

Predicting survival from microarray data - a comparative study

Survival prediction from gene expression data and other high-dimensional genomic data has been subject to much research during the last years. These kinds of data are associated with the methodological problem of having many more gene expression values than individuals. In addition, the responses are censored survival times. Most of the proposed methods handle this by using Cox's proportional hazards model and obtain parameter estimates by some dimension reduction or parameter shrinkage estimation technique. In the talk I will present a comparative study of the prediction performance seven such survival prediction methods: univariate selection, forward stepwise selection, principal components regression, supervised principal components regression, partial least squares regression, ridge regression, and the lasso. The comparison is done by application of the methods to three well-known microarray gene expression data sets.

From our study, we find that statistical learning from subsets should be repeated several times in order to get a fair comparison between methods. Based on such a comparison, methods using coefficient shrinkage or linear combinations of the gene expression values have much better performance than the simpler variable selection methods. In particular, ridge regression has the overall best performance on the studied data sets.

The talk is based on joint work with Ståle Nygård, Hege L. Størvold, Magne Aldrin, Ørnulf Borgan, Ole Christian Lingjærde, and Arnaldo Frigessi.