

STATISTICS COLLOQUIUM

TUESDAY, FEBRUARY 02, 2010

TALK: 4:00 PM — SCIENCE CENTER ROOM 705

RECEPTION: 5:15 PM — SCIENCE CENTER, 7TH FLOOR

“A Penalized Matrix Decomposition, with Application to Sparse Clustering”

Daniela M. Witten

Ph.D. Student
Stanford University

ABSTRACT

We present penalized matrix decomposition, a new framework for computing a low-rank approximation for a matrix. This low-rank approximation is a generalization of the singular value decomposition. While the singular value decomposition usually yields singular vectors that have no elements that are exactly equal to zero, our new decomposition results in sparse singular vectors. When this decomposition is applied to a data matrix, it can yield interpretable results. Moreover, when applied to a dissimilarity matrix, this leads to a method for sparse hierarchical clustering, which allows for the clustering of a set of observations using an adaptively-chosen subset of the features. These methods are demonstrated on the Netflix data and on a genomic data set.

This is joint work with Robert Tibshirani and Trevor Hastie.