Speaker: Daniela Witten, Ph.D. Candidate, Dept of Statistics, Stanford

Title: A Penalized Matrix Decomposition, with Application to Sparse Clustering

Abstract

We present a penalized matrix decomposition, a new framework for computing a rank K approximation for a matrix. We approximate a $n \times p$ matrix $\mathbf{X}$ as $\sum_k d_k \mathbf{u}_k \mathbf{v}_k^T$, where $\mathbf{u}_k$, d_k, and $\mathbf{v}_k$ minimize the squared Frobenius norm of $\mathbf{X} - \hat{\mathbf{X}}$, subject to constraints on $\mathbf{u}_k$ and $\mathbf{v}_k$. Here, $\mathbf{u}_k$ and $\mathbf{v}_k$ are vectors of lengths n and p, and d_k is a scalar. We show that 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.