
PROJECT PROPOSAL

Wentao Wu

Paper

Title: A framework for feature selection in clustering

Author: Witten DM and R Tibshirani

Journal: Journal of the American Statistical Association 105(490): 713-726.

Abstract

This paper addresses the problem of clustering with large set of features. It is expected that the true underlying clusters present in the data differ only with respect to a small fraction of the features, and will be missed if one clusters the observations using the full set of features. They proposed a novel framework for sparse clustering, in which one clusters the observations using an adaptively chosen subset of the features. The method uses a lasso-type penalty to select the features. They used their framework to develop simple methods for sparse K-means and sparse hierarchical clustering. A single criterion governs both the selection of the features and the resulting clusters.
