

STANFORD UNIVERSITY
DEPARTMENT OF STATISTICS
DEPARTMENTAL SEMINAR

4:15 p.m., Tuesday, October 2, 2001
Sequoia Hall Rm. 200
(Cookies at 3:45 in 1st Floor Lounge)

Jerome H. Friedman
Department of Statistics
Stanford University
Stanford, CA 94305

Weighted Harmonic Distance Clustering

A dissimilarity measure for value-attribute data is proposed for use in cluster analysis. It assigns small dissimilarities to observation pairs that have close values on any subset of the attribute variables regardless of their values on the complement set of variables. Using this measure in conjunction with dissimilarity based clustering algorithms encourages the detection of subgroups of observations that preferentially cluster on subsets of the variables. The relevant variable subsets for each individual cluster can be different and partially (or completely) overlap with those of other clusters. Enhancements for increasing sensitivity for detecting especially low cardinality groups clustering on a small subset of variables are discussed. Applications in several different domains, including gene expression arrays, are presented.

Joint work with Jacqueline Meulman, Leiden University.