An L_1-norm procedure for fitting overlapping clustering models to proximity data

Anil Chaturvedi

AT&T Laboratories, USA

J. Douglas Carroll

Rutgers University, New Jersey, USA

Abstract: We describe a general approach to finding Least Absolute Deviation estimates of two-way and three-way overlapping clustering models called ADCLUS and INDCLUS. The suggested approach utilizes a combinatorial optimization approach that takes advantage of a separability property of this loss function for fitting these models. Our approach helps in robustifying the solutions in the presence of extreme outliers in the data.

Key words: Cluster analysis, overlapping clusters, combinatorics, trilinear models, least absolute deviation.

AMS subject classification: Primary 62H30; secondary 92J45.

1 Introduction

Shepard and Arabie (1979) first introduced an overlapping clustering model for classification based on similarity data called ADCLUS (for additive clustering). Subsequently, Arabie and Carroll (1980), provided a mathematical programming approach for fitting this model. Carroll and Arabie (1983), proposed an individual differences generalization of this model, which they called INDCLUS (for individual differences clustering), and also devised a procedure for fitting this three-way generalization, thus providing a methodology for three-way overlapping clustering. These algorithms optimize a least squares, or L_2-norm based, loss function. The theoreti-