Calculating Robust Depth Measures for Large Data Sets

I. Al-Furaih, Syracuse University
T. Johnson, AT&T Labs
S. Ranka, University of Florida

Abstract

Data analysis often requires the finding of the median of the data set. However multidimensional median finding is a difficult problem, and proposed algorithms for multidimensional median finding are expensive to compute and do not scale well. We present approximate algorithms for computing two measures of a median on a 2-dimensional data set, the regression depth and the halfspace depth. Our algorithms require a few passes over the entire data set and provide tight error bounds.