

FCS - NRC Joint 2005-2006 Seminar Series

Primal-Dual methods for data depth By David Bremner

Wednesday, February 23rd, 2005 3:30 – 4:30pm ITC317

Given a set of vectors (variously, feature vectors, experimental results, and measurement vectors) a common task is to compute a "representation" for the set that either makes it easier for humans to understand (in e.g. computational statistics or data mining) or perhaps provides a concise data structure for "likeness queries" (in e.g. machine learning). Arguably the simplest such representation is a single prototype vector. In the computational statistics literature various such protype vectors are referred to as "centres". The naïve centre of a set of vectors is the coordinate-wise average or barycenter. While trivial to compute, it is extremely non-robust, since one erroneous vector can destroy any information content. The natural generalization of the median (called the Tukey Median) is much more robust, but is extremely difficult to compute both in theory and in practice. The Tukey median is the point (or points) of maximum "halfspace depth", a measure of how close to the surface of the convex hull of set of vectors a candidate centre is. In this talk I will describe some work in progress on algorithms and software for computing halfspace depth. These "primal-dual" techniques have the useful property that at every stage they maintain an upper and lower bound on the final answer, and thus provide the user with a partial solution even when terminated early. This talk describes joint work with Komei Fukuda (ETH Zuerich) and Vera Rosta (McGill University).

STUDENTS ARE ENCOURAGED TO ATTEND