

Stanford University

Department of Statistics

DEPARTMENTAL SEMINAR

4:30pm, Tuesday, January 21, 2020
Sloan Mathematics Center Room 380C

Refreshments served at 4pm in Sequoia Lounge.

Speaker: Jerome H. Friedman, *Stanford Statistics*

Title: Contrast Trees and Distribution Boosting

Abstract:

Often machine learning methods are applied and results reported in cases where there is little-to-no information concerning accuracy of the output. Simply because a computer program returns a result does not ensure its validity. If decisions are to be made based on such results it is important to have some notion of their veracity. Contrast trees represent a new approach for assessing the accuracy of many types of machine learning estimates that are not amenable to standard (cross-) validation methods. They are easily interpreted and can be used as diagnostic tools to reveal and then understand the inaccuracies of models produced by any learning method. In situations where inaccuracies are detected, boosted contrast trees can often improve performance. A special case, distribution boosting, provides an assumption-free method for directly estimating the full conditional distribution of an outcome variable y for any given set of joint predictor variable values x .