STATISTICS COLLOQUIUM

MONDAY, OCTOBER 7, 2013
TALK: 4:15 PM — SCIENCE CENTER RM. 705
RECEPTION: 3:50 PM — SCIENCE CENTER, 7TH FLOOR

“Leverage-score sampling for linear regression”

Garvesh Raskutti
SAMSI
(Statistical and Applied Mathematical Sciences Institute)
Research Triangle Park, North Carolina

ABSTRACT

Sampling plays an important role in many statistical contexts. Here we are interested in sampling for linear regression where information for the covariates X is easily available but information for the response Y is difficult to attain (e.g., X represents demographic information for a patient and Y requires a blood test). Sampling based on the “leverage scores” of X is a recent approach developed in the computer science community with encouraging results both in terms of theoretical guarantees and for large-scale data applications. In this talk, I provide a statistical perspective for leverage-score sampling and show that leverage-score sampling performs favorably in terms of mean-squared error compared to uniform random sampling approaches.

This talk is based on joint work with Michael Mahoney (Stanford Univ, Math Dept).

.