

POMONA COLLEGE

COMPUTER SCIENCE COLLOQUIUM

Pulling Rank: Inference from Incomplete Data

Ben Recht

Cal Tech – Pasadena, CA

Abstract:

How can we infer answers from a survey that is only partially filled out? Suppose we ask a large collection of individuals a series of questions. We collect some data but, unfortunately, many questions are left unanswered. Is it possible for us to make an educated guess about what the missing answers should be? How can we make such a guess? In general, with no additional assumptions, this is impossible. However, if we assume that we can arrange all of the answers into a low rank matrix, there is often a unique assignment for the missing entries.

The rank of a matrix is equal to the dimension of the span of its columns (or rows). Matrix rank is often an efficient way to describe system order, complexity, or dimensionality. Moreover, matrices of very low rank can be uniquely determined from very few measurements.

Consequently, rank minimization---finding the minimum rank matrix that agrees with some partial measurements---is a recurring problem in engineering and computer science. It arises in a diverse set of fields including collaborative filtering, Euclidean embedding, multi-task learning, system identification, and controller design. Unfortunately, although specific instances can often be solved with specialized algorithms, the general rank minimization problem is NP-hard.

This talk will present a comprehensive approach to this class of important problems. I will discuss our theoretical understanding of what kinds of data matrices can be recovered from a small number of measurements and how they can be recovered, focusing on some of my recent results in this area. I will subsequently outline practical algorithms for data recovery that can be efficiently applied to large scale problems with guaranteed success.

Thursday, April 3rd, 4:15 pm

Rose Hills Theater – Smith Campus Center

Pomona College

Refreshments available at 4:00