

Graduate Seminar



Efficient sampling for signals on large graphs and large sample linear regression

Aarti Singh

Associate Professor
Machine Learning Department
Carnegie Mellon University

Thursday, February 2nd 4:30 pm Scaife Hall 125

ECE Seminar Committee

Aswin Sankaranarayanan
saswin@ece.cmu.edu

Maysam Chamanzar
mchamanz@andrew.cmu.edu

Swarun Kumar
swarun@cmu.edu

Abstract:

In many applications, we have access to large datasets (such as location of major road intersections in a state, healthcare records, database of building profiles, and visual stimuli), but the corresponding labels (such as traffic or wind speed at the intersections, customer satisfaction, energy usage, and brain response, respectively) are hard to obtain. We investigate the question of how to efficiently sample such large datasets, under label budget constraints. Our solutions apply to the problem of sampling for large-scale linear regression, as well as related problems of sampling signals on large graphs and constrained compressed sensing.

We derive computationally feasible and near minimax optimal sampling strategies for both with and without replacement settings for prediction as well as estimation of the regression coefficients. Experiments on both synthetic and real-world data confirm the effectiveness of our sampling algorithm for small label budgets, in comparison to popular competitors such as uniform sampling, leverage score sampling and greedy methods.

Bio:

Aarti Singh is an Associate Professor in the Machine Learning Department at Carnegie Mellon University. She obtained her PhD in Electrical and Computer Engineering from University of Wisconsin, Madison in 2008 and was a postdoctoral research associate in the Program in Applied and Computational Mathematics at Princeton University before joining CMU in 2009. Her research lies at the intersection of machine learning, statistics and signal processing, and focuses on designing statistically and computationally efficient algorithms that can leverage inherent structure of the data in the form of clusters, graphs, subspaces and manifold using direct, compressive and active queries. Her work is recognized by an NSF Career Award, a United States Air Force Young Investigator Award, A. Nico Habermann Faculty Chair Award, Harold A. Peterson Best Dissertation Award, and a best student paper award at Allerton. She is currently serving as the co-chair for Artificial Intelligence and Statistics (AISTATS) 2017 conference and guest editor for a special issue of Electronic Journal of Statistics.

SEMINAR NOTES: (REFRESHMENTS SERVED AT 4:00 PM)