



THE UNIVERSITY OF
CHICAGO

Department of Statistics

STATISTICS COLLOQUIUM

TONG ZHANG

Department of Statistics
Rutgers University

Statistics in Big Data Optimization

MONDAY, May 2, 2016, at 4:00 PM
Eckhart 133, 5734 S. University Avenue

ABSTRACT

Many modern big-data machine learning problems encountered in the industry involve optimization problems so large that traditional optimization methods are difficult to handle. In this talk, I will present a novel statistical sampling method for multi-class logistic regression that can be used to select a small number of the most effective data points. Asymptotically we show that the proposed method can achieve variance no more than s times that of the full-data MLE with no more than $1/s$ of the full data in the worst case; moreover the required sample size can be significantly smaller than $1/s$ of the full data when the classification accuracy is relatively high. We demonstrate how to use sampling methods in real applications.