



THE UNIVERSITY OF
CHICAGO

DEPARTMENT OF STATISTICS

PHD DISSERTATION PRESENTATION

YONGHOON LEE

Department of Statistics
The University of Chicago

Robust Estimation and Distribution-Free Inference for Supervised Learning

TUESDAY, April 18, 2023, at 11:00 AM

Jones 303, 5747 S. Ellis Avenue

ABSTRACT

We discuss problems where we have limited access to the information of underlying distribution of training data, which can be caused by imperfect data or insufficient prior knowledge.

We first look into the binary classification problem, in the setting where the label observations are corrupted by noise. We establish that corruption acts as a form of regularization, and we compute precise upper bounds on estimation error in the presence of corruption. Our results suggest that the presence of corrupted data points is beneficial up to a small fraction of the total sample, scaling with the square root of the sample size.

Next, we study the regression problem in the distribution-free setting. We show that there are three regimes in terms of the possibility of a meaningful inference, which are characterized by 'effective support size' of the feature distribution. Our result implies that there exists an counterintuitive in-between regime where we can still expect to obtain a meaningful inference for a future input which is unlikely to have the value we have observed before.

We also develop distribution-free methods for predictive inference with hierarchically structured datasets. For the special case where we have i.i.d repeated measurements, we propose to bound the expected squared conditional miscoverage rate in order to have a better control of the conditional coverage, and extend existing methods to construct distribution-free prediction sets that achieve the bound.