psd

MASTER'S THESIS PRESENTATION

A Simple Causal Model for Invariant Learning

WHEN
May 10, 2022
1:00 PM

WHERE
Jones Laboratory, Room 304

Tanmay Gupta, MS candidate

Deep neural networks are very good at making predictions using image and text data. In other words, they can detect patterns in the data to predict a label Y from an input X . A significant drawback is that models might learn features of the input X which help it predict Y in the training set which shouldn't matter, i.e. associations which might not hold in test data. For example, a model to predict sentiment in Amazon product reviews might associate product category with sentiment, which will result in biased performance on unseen data. Causality lends itself very well to separate such spurious correlations from genuine, causal, ones. In this paper, we present a simple causal model for data and a method using which we can train a classifier to predict a category Y from an input X, while being invariant to a variable Z which is spuriously associated with Y. We first lay down the theory and then present experimental results on text and image data.

stat.uchicago.edu

THE UNIVERSITY OF CHICAGO | DEPARTMENT OF STATISTICS