psd

MASTER'S THESIS PRESENTATION

Adaptive Data Analysis: A Theoretical and Empirical Overview

WHEN
May, 6, 2021
3:30 PM, CDT

WHERE
Via ZOOM
ZOOM information will be provided in the email announcement for this seminar.

Irina Cristali, MS candidate

In order to ensure valid statistical conclusions, one would ideally select an analysis to be performed on a given dataset independently of the data themselves. However, this independence does not hold any longer when data are re-used across analyses and the analysis to be performed at a given step depends on the results of previous analyses. A typical example of such a situation occurs in Kaggle competitions, where a contestant may be repeatedly tuning their model until it achieves a desired predictive accuracy, all the while using the same training data which has a limited size. In this work, we aim to provide a deep insight into the state-of-the-art methods for preserving the validity of statistical conclusions when re-using data, focusing on the core theoretical frameworks, the experimental performance of these methods, as well as their limitations, and further research directions.

stat.uchicago.edu

THE UNIVERSITY OF CHICAGO | DEPARTMENT OF STATISTICS