

Department of Applied and Computational Mathematics and Statistics Colloquium



Yuhong Yang

School of Statistics
University of Minnesota

Cross-Validation for Optimal and Reproducible Statistical Learning

In data mining and statistical learning, we frequently encounter the task of comparing different methods/algorithms to reach a final choice for pure prediction or a scientific understanding/interpretation of a regression relationship. Cross-validation provides a powerful tool to address the matter. Unfortunately, there are seemingly widespread misconceptions on its use, which can lead to unreliable conclusions. In this talk, we will address the subtle issues involved and present results of minimax optimal regression learning and consistent selection of the best method for the data. In addition, we will propose proper cross-validation tools for model selection diagnostics that will cry foul at an impressive-looking but not really reproducible outcome from a sparse-pattern-hunting method in the wild west of learning with a huge number of covariates.

Monday, November 5, 2018

4:15 PM – 5:15 PM

127 Hayes-Healy Center

Colloquium Tea 3:45 PM to 4:15 PM 101A Crowley Commons Room