A proposal for cost-considerate variable selection in statistical models

Michael Lerch

In statistical modeling, an important step is often variable selection: what variables are to be included in the final model. Naturally, there is no single best strategy to perform variable selection and the number of strategies abound. However, one of the most practically important aspects often goes unconsidered: cost of data collection. Researchers may ask the intuitive question: "How *good* must this variable be to justify the cost of measurement?" Regularization is becoming more and more important for model fitting (especially when prediction is the goal) as datasets become larger and wider. Regularization includes pruning variables and also coefficient shrinkage. Without regularization, models can easily become overfit. Variable selection can help prevent overfitting but it is an all-or-nothing approach to regularization. Techniques that perform coefficient shrinkage may find a balance between variable selection and no regularization. In this talk, I plan to review what literature exists on cost-considerate variable selection and to propose a new approach to cost-considerate variable selection for linear models that also performs coefficient shrinkage. The inspiration is the Lasso, the Least Angle Shrinkage and Selection Operator. With the new strategy, researchers may be able to perform cost-considerate variable selection and coefficient shrinkage in a computationally efficient way.

This talk is the public component of Michael Lerch's dissertation proposal and oral comprehensive exam.

Location: AJM 233
Time: Thursday, Oct. 29, 9:00AM