EC994: Applications of Data Science

Thiemo Fetzer

University of Warwick

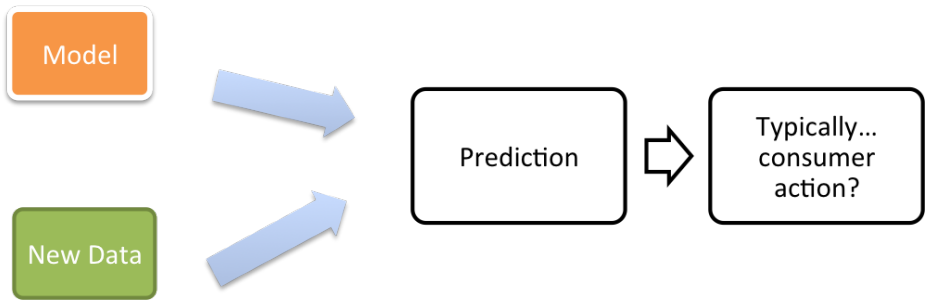
October 4, 2019

What is Data Science?

What is Data Science?

- Lies at the intersection of statistics, computer science, engineering and business
- Data science is "the art of extracting value from (unstructured) data."
- Data insight driven decision making becomes an increasingly important practice in private as well as public sector.

A Data Scienctist's Work Flow



Topics Covered

Part 0: Concepts	Statistical Learning, Bias-Variance Tradeoff
Part 1: (Numeric) Prediction	Linear regression, Model selection, Shrinkage, Regression Trees
Part 2: Classification	Logistic regression, Naive Bayes, Nearest Neighbours, SVM, Trees
Part 3 : Dimensionality Reduction	Principal Component, k-means clustering, mixture models, Topic modelling

Useful Textbook (free): An Introduction to Statistical Learning with Applications in R, James, G., Witten, D., Hastie, T. and Tibshirani, R. (4th edition), Springer Statistics.

What do (Applied) Economists care about...

- Internal validity (or minimizing systematic error or **bias**)
- Focus on causal identification of theoretically founded mechanism
- A wealth of methods

What do Data Sciencists care about...

- Strong focus on (out of sample) prediction

  e.g. online conversion rates, ad placement, price discrimination
- Model optimization to achieve robust inference
- Typically atheoretic models