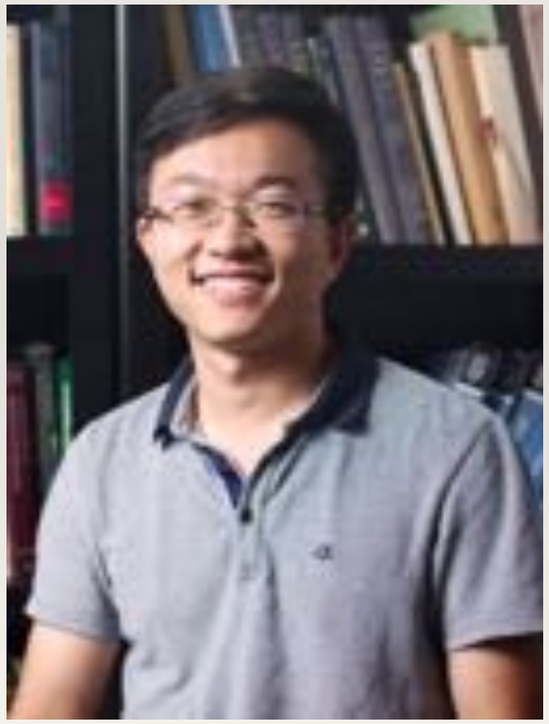
Practical Statistical Learning (F18)

Feng Liang



Changbo Zhu

Yinyin (Elaine) Chen

Joshua Daniel Loyal

Overview

- Types of statistical learning problems
- Why learning is difficult?
- Bias variance tradeoff
- An example: kNN *vs* Linear Regression (in a separate pdf file)
- Not all about prediction

Problems (I)

- Project 1 (Ames Housing Data): Predict the sale price of a house given its features.
- Project 2 (Sales Forecasting): Provide sales forecasting for Walmart for each department in each store based on historical data.

Y: Target
X: Feature / Covariates

Problems (II)

- Project 3 (Lending Club): Determine the chance that a borrower will miss a payment next month given various characteristics of the borrower and the loan.
- Project 4 (Sentiment Analysis): Determine whether a movie review is positive or negative.

Y: Target
X: Feature/Covariates

Problems (III)

- Based on the recent real estate transactions at Ames, Iowa, can we identify any home buying/selling trends? Further, can we identify distinctive groups of buyers?

- Based on the transaction data at Walmart, can we recommend any marketing strategies to Walmart?

Association Rule (chap 14.2 of ESL)
Market Segmentation (cluster customers)

Problems (III)

- Based on the recent real estate transactions at Ames, Iowa, can we identify any home buying/selling trends? Further, can we identify distinctive groups of buyers?
- Based on the transaction data at Walmart, can we recommend any marketing strategies to Walmart?

Y: ~~Target~~

X: Feature/Covariates

Types of Statistical Learning Problems

- Supervised Learning
  - Regression: response is a number
  - Classification: response is a label (binary or multi-class)

Semi-supervised Learning *Recommender System*

- Unsupervised Learning: identify latent structures in the data, e.g., clustering, association rule, HMM, etc.