(PRACTICAL) STATISTICAL LEARNING

STAT542/CS598 **SYLLABUS AND COURSE POLICIES**[1] Fall, 2018

Instructor: Feng Liang
113D Illini Hall
liangf AT illinois DOT edu
(217) 333-6017

Teaching Assistants: Yinyin (Elaine) Chen, ychen409 AT illinois DOT edu
Joshua Daniel Loyal, jloyal2 AT illinois DOT edu
Changbo Zhu, changbo2 AT illinois DOT edu

Course Web Page: https://piazza.com/illinois/fall2018/stat542cs598/home

Lectures: MWF 1:00-1:50 pm, Noyes Laboratory 161

Text

- Trevor Hastie, Robert Tibshirani, and Jerome Friedman (2nd edition). *The Elements of Statistical Learning: Data Mining, Inference and Prediction.*
- Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani (4th edition). *An Introduction to Statistical Learning with Applications in R.*

Prerequisites

Knowledge of basic multivariate calculus, statistical inference, and linear algebra. You should be comfortable with the following concepts: probability distribution functions, expectations, conditional distributions, likelihood functions, random samples, estimators and linear regression models.

Course Description

This course provides an introduction to modern techniques for statistical analysis of complex and massive data. Examples of these are model selection for regression/classification, nonparametric models including splines and kernel models, regularization, model ensemble, recommender system, and clustering analysis. Applications are discussed as well as computation and theoretical foundations.

[1] This syllabus is subject to change/revision at the instructor's discretion.

Grading

40%: 10 quizzes;
12%: 4 coding assignments;
48%: 4 mini projects.

Assignments are due by 11:30 pm on the due day. Check the due days on the google calendar for this course [Link].

Quizzes will be posted on Compass or Coursera. For each quiz, students are given up to three attempts and the highest counts. The score for each quiz is different and the final score for quiz is cumulative over all quizzes.

You can discuss quiz problems, assignments and projects with other students but should finish your quizzes and assignments independently. Copying solutions from another student or webpages is cheating and plagiarism.

Computing

The assignments and projects will involve some computing. You are expected to have some prior programming experience with R or Python.

Class Outline

Here is a list of tentative topics.

1. **Introduction**: 1 week
2. **Linear regression**: 1 week
3. **Regularization**: 1 week
4. **Regression tree models**: 1 week
5. **Nonlinear regression**: 1 week
6. **Clustering**: 2 weeks
7. **Classification**: 1 week
8. **Linear classifier**: 1 week
9. **SVM**: 1 week
10. **Classification tree models**: 1 week
11. **Recommender system**: 1 week
12. **Large scale machine learning**: 1 week