FALL 2019 - STAT 440 E100

#ÌÁÓÓ . Õ Í ÂÅÒ¿ψωύφ $ÅÌËÖÅÒÙ - ÅÔÈÏÄ¿)Î 0ÅÒÓÏÎ

O e e

Mo 4:30 PM – 5:20 PM
RCB 6125, Burnaby

We 4:30 PM – 6:20 PM
RCB 6125, Burnaby

Elliott, Lloyd
lloyde@sfu.ca
1 778 782-3376
Office: SC-K10550

90 units including STAT 350 and one of STAT 341 or CMPT 225, or instructor approval. STAT 240 is also recommended.

De c

A data-first discovery of advanced statistical methods. Focus will be on a series of forecasting and prediction competitions, each based on a large real-world dataset. Additionally, practical tools for statistical modeling in real-world environments will be explored.

STAT 440 is suitable for senior students who have a minimum of 90 units.

Course Outline

The course will be split into several modules. Each module will focus on a particular dataset. At the start of each module, students will be randomly divided into teams. A subset of the dataset will be given to all teams (the training data) and the rest of the dataset will be withheld (the test data). Students will learn modern machine learning methods for predicting aspects of the test data based on the training data. This test/train paradigm is often encountered in both academic and industrial settings. The methods will include bagging, boosting, deep learning, model blending and cross-validation. The students will learn how to implement these methods using standard software packages such as scikit-learn and tensorflow. They will use these methods (and any other techniques they wish) to predict aspects of the test data. During each module, teams will submit their predictions and see the results of those predictions on the withheld test data. Marks for competition results will be awarded based on the accuracy of their predictions.

G ad

Competition Results

Competition Writeups

Homework

Assignments and Grading Procedures

- Competition Results (50%): The course's modules will be held as competitions. Students will be randomly divided into teams at the start of each module. Half of the marks for competition results will be awarded based on a team's performance relative to other teams, and the other half will be awarded based on a team's performance relative to objective baselines.

- Competition Writeups (30%): At the end of each module, each team will provide a short report describing their code, methods and thought processes.

- Homework (20%): Problem sets will be assigned (to be done individually) following the methods taught in the lectures.

Above Grading is subject to change.

REQUIRED READING:

RECOMMENDED READING:

Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems (A. Géron, 2017, O'Reilly)

Deep Learning (I. Goodfellow and Y. Bengio and A. Courville, 2016, MIT Press)

Machine Learning: A Probabilistic Perspective (K.P. Murphy, 2012, MIT Press); Elements of Statistical Learning (T. Hastie, R. Tibshirani and J. Friedman, 2009, Springer)

Introduction to Machine Learning with Python (A. Müller and S. Guido, 2016, O'Rielly)

Linear Algebra, 5th Edition (S. Friedberg, A. Insel and L. Spence, 2018, Pearson)

Learning R (R. Cotton, 2013, O'Reilly)

Information Theory, Inference, and Learning Algorithms (D. MacKay, 2003, Cambridge University Press)

DEPARTMENT UNDERGRADUATE NOTES:

Students with Disabilites:
Students requiring accommodations as a result of disability must contact the Centre for Accessible Learning 778-782-3112 or csdo@sfu.ca

Tutor Requests:
Students looking for a Tutor should visit http://www.stat.sfu.ca/teaching/need-a-tutor-.html. We accept no responsibility for the consequences of any actions taken related to tutors.

REGISTRAR NOTES:

SFU's Academic Integrity web site http://www.sfu.ca/students/academicintegrity.html is filled with information on what is meant by academic dishonesty, where you can find

c