DATA 5610: Advanced Machine Learning for Analytics

Spring Semester, 2022

Course Information

- Course Dates: January 10 - May 4, 2022
- Course Time: MW 9:00 - 10:15 AM
- Course Room: EBB 120
- Slack Channel
- Canvas Page

Intructor Information

- Tyler J. Brough
- Office Hours: By Appointment
- Office: BUS 512
- Email: tyler.brough@usu.edu

Syllabus

Course Description

This course provides a foundational understanding of predictive modeling, machine learning, and data analysis for business and economic decision making. The focus of the class is on practical applications rather than deep theoretical development. Theory will be introduced to support successful implementation.

The main purpose of the course is to foster three modes of thought to enable you to become competent scholars of business data analytics:

1. **Computational logic:** Computational thinking helps break a problem down and find a practical path to successful implementation. Important techniques from computational mathematics and statistics will be introduced as they pertain to data analysis.

2. **Statistical logic:** The main focus of the course will be learning to think statistically about problems in business analytics. All meaningful problems in data analysis are necessarily embedded in conditions of uncertainty. There exists a core statistical logic that is distinct from the mathematical details. Learning to develop this mode of thinking is an essential step in the life of any successful business data analyst. We will find that statistical reasoning is essential for proper business decision making.

3. **Economic logic:** Arbitrage is the central underlying concept of all of economics and finance. Developing skills in economic reasoning is an essential element in business data analytics.

These three ways of thinking are not independent. Quite to the contrary, we will find that they are strongly mutually reinforcing.

Method of Teaching and Learning

This course will be taught as a graduate seminar style course. **That means that your preparation and participation is crucial.** We will get out of this course what we individually and collectively put into it.

We will follow the Socratic Method during lectures. A good amount of class time will also be dedicated to guided computational exploration. The idea behind this exploration is to develop a methodology for **computational thinking**. Mathematical theory will serve as an important background to this enterprise but will not itself be the focus.

Your preparation and participation is absolutely essential to the success of this course!

Assessment

The grade that you will earn will be determined by the weighted total points accumulated. The weights given to each part of the class are as follows:

- *Class Preparation & Participation (10%)* - In a Socratic style seminar student preparation and participation is essential!

- *Student Presentations (10%)* - Throughout the course a set of specialized readings will be distributed. Each student will have an opportunity to prepare a presentation to the class on one or more of the assigned readings.

- *Annotated Bibliography & Wiki-style Articles (15%)*

    - Throughout the course, a set of specialized readings will be distributed. Each student will prepare an annotated bibliography of each reading.
    - Students will prepare a Wikipedia-style summary article on one specialized topic from each of the four areas of concentration. The following are examples (not exhaustive) of topics that might be chosen:
        1. Decision theory: sequential analysis, utility and loss functions, risk analysis, etc.
        2. Supervised learning: pattern classification, Lasso regression, Random forests, etc.
        3. Time series analysis: cointegration and error-correction, time series bootstrap methods, hidden markov models, etc.
        4. Reinforcement learning: k-armed bandits, Thompson sampling, AlphaZero, etc.

- *Projects (65%):*

    - Kaggle Titanic classification (16.25%) - Kaggle Titanic - Machine Learning from Disaster

- Stock returns prediction (16.25%)
- Pairs trading prediction (16.25%)
- Auction/Reinforcement learning simulation (16.25%)

- *Extra Credit:* There will several opportunities for extra credit work.

Communications

We will use Canvas only to share documents and for certain communications.

For all other communication purposes we will use a course Slack channel. Students will receive an invitation to the Slack the first week of class. You should check for communications and announcements often.

Readings

Unfortunately, there is no single source that covers all of the topics that we will cover in this course. The readings and a reading schedule will be distributed through the course Canvas page. We will take readings from the following sources.

- Analysis of Integrated and Cointegrated Time Series with R by Bernhard Pfaff

- An Introduction to Statistical Learning: with Applicaitons in R by James, Witten, Hastie & Tibshirani

- Applied Predictive Modeling by Kuhn & Johnson

- Computational Statistics Handbook with MATLAB by Martinez & Martinez

- Economic Forecasting by Elliot & Timmermann

- Introduction to Machine Learning by Ethem Alpaydin

While these are the main sources there will be additional sources distributed in addition to the above.

Topics

We will cover the following main modules:

1. Introduction to statistical decision theory
2. Supervised learning
3. Time series analysis
4. Reinforcement learning

Import dates:

- **First day of classes** - Jan 10
- **Martin Luther King, Jr. Day** - Jan 17

- **Presidents' Day** - Feb 21
- **Spring Break** - Mar 7 - 11
- **Last day of classes** - Apr 26
- **Final exams** - Apr 28 - May 4