Syllabus for *Applied I*

The principles of estimation for statistical learning and hypothesis testing.

Course Code: DATASCI 520

Instructor: Abhishek Ananth, PAIS 589, abhishek.ananth@emory.edu

Pre-requisite: None

What is this class about?

This course provides a comprehensive understanding of applied data analysis through a structured exploration of various statistical and machine learning techniques. The course is divided into five modules that progressively introduce key concepts and methods used in analyzing data, with a strong emphasis on real-world applications.

Readings:

Hansen, Bruce. Econometrics. Princeton University Press, 2022. **[BH]**

James, Gareth, et al. An Introduction to Statistical Learning: with Applications in Python. New York: springer, 2023. **[ISLP]**

Learning Objectives:

- Develop an intuitive understanding of the core philosophy of data analysis.

- Explore and understand different statistical models and estimators.

- Apply appropriate techniques to analyze data with both continuous and discrete outcomes.

- Tackle challenges related to high-dimensional data.

- Interpret and critically evaluate results from data analysis.

Class Requirements:

1. Problem sets x 5 (14% each)
2. Class participation (5%)
3. Take home final (25%)

You will be given feedback routinely through your problem sets.

Modules:

Week 1: Review of Probability and Statistics

- Essential probability concepts: distributions, expectations, variance - Basic statistical measures: mean, median, mode, standard deviation
- Key concepts in inferential statistics: hypothesis testing, confidence intervals
- Overview of common probability distributions (Normal, Binomial, etc.) - Readings:
- Assignments: Problem set on foundational concepts

Module 0: A Primer in Statistical Decision Theory

- *Week 2-3*
- Introduction to Statistical Decision Theory
- The role of decision theory in data analytics
- Loss functions and risk
- Bayesian vs. Frequentist perspectives
- Readings: [ISLP Ch 2], [BH Ch 2]
- Assignments: Problem set on decision theory concepts

Module 1: Low Dimensional Features with Continuous Outcomes

- *Week 4-6*
- Linear regression models
- Estimation techniques: OLS, MLE
- Model assumptions and diagnostics
- Interpretation of coefficients and model fit
- Case studies with real-world data - Readings: [BH Ch 4], [ISLP Ch 3]
- Assignments: Data analysis project focusing on continuous outcomes Module 2: Low Dimensional Features with Discrete Outcomes
- *Week 7-9*
- Logistic regression and other generalized linear models (GLMs)
- Estimation and inference in discrete outcomes - Model evaluation: ROC curves, AUC, etc.
- Application protocols with categorical data

- Real-world examples and case studies
- Readings: [ISLP Ch 4], [BH Ch 25]
- Assignments: Project on discrete outcome analysis

Module 3: High Dimensional Features with Continuous Outcomes

- *Week 10-12*
- Introduction to high-dimensional data analysis
- Regularization techniques: Lasso, Ridge, Elastic Net
- Model selection and cross-validation
- Applications in genomics, finance, etc.
- Handling multicollinearity and model complexity
- Readings: [ISLP Ch 6-7], [BH Ch 19]
- Assignments: High-dimensional data analysis project

Module 4: High Dimensional Features with Discrete Outcomes

- *Week 13-15*
- Classification techniques in high dimensions
- Support Vector Machines (SVM), Random Forests, etc.
- Feature selection and dimensionality reduction
- Practical challenges and solutions in high-dimensional settings
- Case studies with real-world applications
- Readings: [ISLP Ch 8-9], [BH Ch 28]
- Assignments: Final project on high-dimensional discrete outcome analysis **Final Project:**

TBD

Mental Health:

During your time at Emory, you may experience a range of challenges which could cause barriers to your learning and academic engagement (including, but not limited to: anxiety, relationship issues, elevated levels of stress, substance use problems, grief, sadness, or decreased motivation). Emory's Counseling & Psychological Services (CAPS) is here to help with these and other mental health concerns you might experience. You can access free, confidential mental health services by calling them at: (404) 727-7450.

Honor Code:

The Honor Code is in effect throughout the semester. By taking this course, you affirm that it is a violation of the code to cheat on exams, to plagiarize, to deviate from the teacher's instructions about collaboration on work that is submitted for grades, to give false information to a faculty member, and to undertake any other form of academic misconduct. You agree that the instructor is entitled to move you to another seat during examinations, without explanation. You also affirm that if you witness others violating the code you have a duty to report them to the honor council.

Accessibility Statement:

If you anticipate issues related to the format or requirements of this course, please meet with me. I would like us to discuss ways to ensure your full participation in the course. If you determine that disability-related accommodation is necessary, you may register with Accessibility Services at (404)727-9877 or via e-mail at accessibility@emory.edu. To register with OAS, students must self-identify and initiate contact with the OAS office.

Copyright Statement:

The materials provided by the instructor in this course are for the use of the students enrolled in the course. Materials in this course—unless otherwise indicated—are protected by United States copyright law [Title 17, U.S. Code]. Materials are presented in an educational context for personal use and study and should not be shared, distributed, or sold in print—or digitally—outside the course without permission. Click here for more information.