Syllabus

229351: Statistical Learning for Data Science 1

Semester 1/2025

1. Course Information

- **Course Code:** 229351
- **Course Title:** Statistical Learning for Data Science 1
- **Credits:** 3 (3-0-6)
- **Prerequisites:** I will assume that you are familiar with basic statistics and programming (preferably Python).

2. Instructor Information

- **Instructor:** Donlapak Ponnoprat (https://donlapark.pages.dev)
- **Lecture Session:** Monday, 9:00 AM - 11:00 AM, in Room SCB4202.
- **Lab Session:** Thursday, 9:00 AM - 11:00 AM, in Room SCB4405-06.
- **Course Website:** https://donlapark.pages.dev/229351/ for course schedule and slides
- **Canvas:** https://mango-cmu.instructure.com/courses/18110 for lab assignments

3. Course Description

This course covers the first topics in statistical learning. Topics include principal component analysis (PCA), linear regression, time series analysis, and logistic regression for classification. We will primarily use Python libraries such as `numpy`, `pandas`, `statsmodels` and `scikit-learn`.

4. Course Schedule

The following is my tentative schedule for the course. I may make adjustments based on our class progress and your interests.

Week	Topics	Lecture Hours	Lab Hours
1	1. Introduction: Syllabus and Course Overview	1	-
2-3	2. Basic Linear Algebra 2.1 Vectors and Matrices 2.2 Transpose, Identity Matrices, Matrix Inverse	2	2
4-5	3. Principal Component Analysis (PCA)	4	4
6-8	5. Linear Regression 5.1 Simple Linear Regression 5.2 Multiple Linear Regression 5.3 Model Diagnostics and Variable Selection	8	8
9	Midterm Exam		
10-13	6. Time Series Analysis 6.1 Data Preparation for Time Series 6.2 Classical Decomposition Techniques 6.3 Exponential Smoothing Models 6.4 ARIMA Models	8	8
14	7. Logistic Regression	2	2
15	8. Evaluation of Classification Models	2	2
16	Final Exam		

5. Assessment and Grading

5.1 Grading Scheme

I will determine your final grade based on the following components:

- **Labs (10-12 assignments):** 30%
- **Midterm Examination:** 35%
  - **Date:** Saturday, August 30, 2025
  - **Time:** 15:30 - 18:30 (3 hours)

  Please fill out the form (link in Canvas) so that we can decide on the best time slot.
- **Final Examination:** 35%
  - **Date:** Sunday, November 2, 2025
  - **Time:** 15:30 - 18:30 (3 hours)

5.2 Grading Policies

- **Mandatory Exams:** You must take both the midterm and final examinations. If you miss either exam, you will receive an automatic 'F' grade for the course.
- **Grading:** Your final grade may be curved at the end of the semester.
- **Lab Assignments:**
  - I require you to submit your lab assignments as Google Colab notebooks, saved to a public GitHub repository. Your submissions on Canvas should be the **public GitHub link** to your '.ipynb' file.
  - I will provide detailed step-by-step instructions on how to sign up for GitHub, create a repository, and save your Google Colab notebooks to GitHub in a separate guide.
  - Each lab is due 23:59 on the next Wednesday.
  - Your two lowest lab scores will be dropped from the total score calculation.

6. Course Policies

- **Academic Honesty:** I expect all work you submit to be your own. Plagiarism, cheating, or any form of academic dishonesty will not be tolerated and will result in severe penalties, including a failing grade for the assignment or the course. Please refer to the university's academic integrity policies.
- **Attendance:** While I may not directly grade attendance, I highly encourage active participation in both lectures and lab sessions as it is crucial for your success in this course. It is your responsibility to catch up on any missed classes.
- **Late Submissions: Late submissions will not be accepted under any circumstances.**
- **Communication:** I will make all important announcements via email. You are responsible for regularly checking your inbox. For individual questions, please send a direct message via Canvas or direct email.
- **Disability Services:** If you have a disability and may need accommodations in this class, I encourage you to contact me as early as possible so that we can arrange reasonable accommodations.

7. Required and Recommended Resources

7.1 Main Course Materials

- My lecture notes and materials will be provided on the course website.

7.2 Recommended Books and Online Resources

- **An Introduction to Statistical Learning: with Applications in R** By Trevor Hastie et al.
  - Book PDF available at: https://www.statlearning.com/
  - Python code examples available at: https://github.com/JWarmenhoven/ISLR-python
- **Forecasting: Principles and Practice** By Rob J Hyndman and George Athanasopoulos.
  - Accessible online at: https://otexts.com/fpp2
- **Linear Algebra and Learning from Data** By Gilbert Strang.

Disclaimer: Please note that this syllabus is my guide for the course and may be subject to change at my discretion. I will announce any changes in class and on the course website.