Syllabus

229352: Course Name: Statistical Learning for Data Science 2

Semester 1/2025

1. Course Information

- **Course Code:** 229352
- **Course Title:** Statistical Learning for Data Science 2
- **Credits:** 3 (3-0-6)
- **Prerequisites:** I will assume that you are comfortable with linear algebra (specifically matrix-vector and matrix-matrix operations), probability and programming (preferably Python). I also recommend knowing some calculus as a tool for optimization, but it is not required.

2. Instructor Information

- **Instructor:** Donlapak Ponnoprat (https://donlapark.pages.dev)
- **Lecture Session:** Friday, 9:00 - 11:00 AM, in Room STB205.
- **Lab Session:** Tuesday, 9:00 - 11:00 AM, in Room SCB4405-06.
- **Course Website:** https://donlapark.pages.dev/229352/ for course schedule and slides.
- **Canvas:** https://mango-cmu.instructure.com/courses/18109 for lab assignments.

3. Course Description

We will cover classification models, clustering methods, deep learning models and their applications. We will use `scikit-learn` for machine learning models and `Pytorch` for deep learning models.

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: What is Statistical Learning?	1	0
2-3	2. Data Preprocessing 2.1 Tabular and Time Series Data Preprocessing 2.2 Image and Text Data Preprocessing	2	2
4-6	3. Classification 3.1 K-Nearest Neighbors (KNN) 3.2 Naive Bayes 3.3 Decision Trees 3.4 Random Forests 3.5 Support Vector Machines (SVM) 3.6 Gradient Boosting	10	10
7-8	4. Clustering 4.1 K-mean and Hierarchical Clustering 4.2 Gaussian Mixture Models (GMM)	2	2
9	Midterm Exam		
10-15	5. Deep Learning Models 5.1 Neural Networks & Backpropagation 5.2 Convolutional Neural Networks (CNNs) 5.3 Transformers 5.4 Generative Models	14	14
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% 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:** 8:00 - 11:00 (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 the course website (https://mango-cmu.instructure.com/courses/18109) 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 Monday.
- Your lowest lab score 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

- **Hands-on Machine Learning with Scikit-Learn, Keras and TensorFlow:** By Aurélien Geron. Notebooks online at https://github.com/ageron/handson-ml3
- **Dive into Deep Learning (D2L):** By Aston Zhang, Zack C. Lipton, Mu Li and Alex J. Smola. Accessible online at https://d2l.ai/
- **Deep Learning Tutorials:** The University of Amsterdam. Accessible online at https://uvadlc-notebooks.readthedocs.io/en/latest/index.html
- **Artificial Intelligence Course:** Alfredo Canziani, Ernest Davis. Accessible online at https://atcold.github.io/NYU-AISP24/

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.