Course Syllabus

DS 637: Python and Mathematics for Machine Learning

Spring 2024

Instructor: Daming(David) Li, email: dli@njit.edu

Course Description and format:

This course aims to equip students with foundational knowledge and practical skills in Python programming and mathematics as they relate to machine learning. By combining theory and practice, students will gain a solid understanding of the basics of Python programming, data analysis, probability, statistics, linear algebra and optimization, and be prepared for more advanced studies in machine learning. This course could be a valuable first step for anyone interested in pursuing a career or further studies in machine learning.

This course will be divided into two sections. The first part will be focused on data analysis. It will start with fundamental Python programming concepts, followed by the use of Numpy and Pandas dataframes, and then how to create plots and visualizations of data. This section will provide a strong foundation in data analysis with Python, which can be valuable for various applications beyond machine learning.

The second part of this course will focus on mathematics as it relates to machine learning. This section will bring together the mathematical concepts and their Python implementations. This can be valuable for those interested in a deeper understanding of machine learning, as mathematical concepts are an essential part of many machine learning algorithms. By gaining experience in their Python implementations, students will be able to apply these concepts practically and see how they can be used in real-world applications.

Textbooks (helpful but not required):

- Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter 3rd Edition by Wes McKinney
- Mathematics for Machine Learning 1st Edition by Marc Peter Deisenroth

Collaboration and Honor Code: Students may discuss problems together but must write up their own solutions. When writing up the solutions, students should write the names of people, if any, with whom they discussed the assignment. Note in particular that copying homework or programming assignments, in full or in part is forbidden. Students found cheating or plagiarizing will be immediately referred to the Dean of Students and the NJIT Committee on Professional Conduct and subject to Disciplinary Probation, a permanent marking on the record, possible dismissal, and an “F” grade in the course. All submitted assignments will be checked for similarities, and plagiarism and guilty students identified.

Grading:

The requirements of this course will consist of participating in lectures, homework, in class computing lab assignments, two exams and a project. The grading breakdown is the following:

- Homework, computing lab exercise (10%)
- Quiz (20%)
- Term Project (20%)
- Midterm (20%)
- Exam (30%)

Homework (10 %)

- Only use Python in homework
- Try to do it independently, discussions allowed, but copying is forbidden.
- 25% penalization per late day;
- Not accepted more than 3 days late

Lab exercise

- Have a lab session every week
- Focus on Python computing exercises
- We will solve some simple problems
- Post your answers by replying on canvas
- Some answers may be selected for discussion by the end of lab session.
- Some problems may become part of homework

Quiz (20%)

- Focus on course materials.
- 4 Quizzes
- Every other week
- Only Python is allowed

Two Term Projects (20%)

- Use Python for your projects
- Use Jupyter
- Submit code and report to summarize what you have done and results you obtained.
- Prepare for presentation and demo.
- 1~4 students a group.
- More details to be announced on canvas
- Cheating/Copying is strictly prohibited.

Two Exams (50%)

- One midterm and one Final (20%+30%)
- In-class
- Final is cumulative
- Only Python

Tentative course topics (Subject to changes according to progress)

1. Class overview and Python basics I
2. Python basics II
3. Array Oriented Programming with NumPy
4. Python pandas Series and Dataframe
5. Data Cleaning and Preparation
6. Visualization and Plotting
7. Data Wrangling - Join, Combine and Reshape
8. Midterm Exam
9. Probability and Distribution
10. Statistical tools
11. Linear Algebra and Analytic Geometry
12. Matrix Decompositions and Solving Systems of Linear Equations
13. Vector Calculus for Machine Learning
14. Root Finding and Continuous Optimization
15. Final Exam