Introduction to Data Mining
DATS 6103 - 10, Summer 2018

1 Meeting Time and Location

- Meeting time: Tuesday / Thursday, 5:10 PM - 7:40 PM
- Location: Corcoran Hall 103

2 Instructor

- Name: Yuxiao Huang
- Email: yuxiaohuang@gwu.edu
- Office address: 2100 Pennsylvania Avenue, Suite 200, Room 281
- Office hours: Monday - Thursday, 4:00 PM - 5:00 PM
- Note: If you cannot make my scheduled office hours and need to talk outside of class, please send email to set up an appointment. I will try to respond within 24 hours. Please be aware that I may be unable to answer emails about Homework and Final project before the deadline, if they are received less than 24 hours before they are due.

3 Teaching Assistant

- Name: Deepak Aggarwal
- Email: deepakagarwal@email.gwu.edu
- Office address: Monday 7:40 PM - 8:40 PM, Corcoran Hall 204; other times, 2100 Pennsylvania Avenue, Suite 200
- Office hours: Monday 4:00 PM - 5:00 PM, 7:40 PM - 8:40 PM, Tuesday 4:00 PM - 5:00 PM, Thursday 4:00 PM - 5:00 PM

4 Course Description

- This course is an introduction of data mining using Python
- The course has three major parts: Python language (3.5 classes), data visualization and preprocessing (2 classes), and linear models (4 classes)
- Although lectures will include some theory, the emphasis will be on coding

5 Learning Outcomes

As a result of completing this course, students will be able to:

- use Python to visualize and preprocess data
- use Python to implement linear models and apply the models to solve real-world problems
- write technical report and present the results
- work both individually and as a team

6 Textbook

The following book is recommended but not required:

- Raschka S. and Mirjalili V. (2017). *Python Machine Learning. 2nd Edition.*

7 Average Minimum Amount of Out-Of-Class or Independent Learning Expected Per Week

- Going over key concepts and doing lots of problems, beyond what is assigned in class, is integral for success in this course
- You should spend at least 10 hours of out-of-class or independent learning per week

8 Homework

- There will be 6 Homework assignments, which will be solely based on Python programming
- Homework must be completed individually

9 Final Project

The Final project is a good opportunity for you to apply data mining methods to complex, real-world problems. It will be completed by teams of 1, 2, or 3 students. Each team can choose a problem in the domain of their interest.

9.1 Deliverables

- Project proposal
- Code and a readme file (describing how to run the code)
- Final report

9.2 Proposal

The project proposal is 1-page maximum. It should include:

- The title of the project
- The problem definition and motivation
- The proposed method, language, and package you will need for the implementation
- The link to the data
- The responsibility of each team member

9.3 Data and Code

- Each team must use real-world data. Simulated data are not allowed. Please talk to the instructor if you are not sure about the nature of the data. There are many publicly available datasets. For example, UCI and Kaggle provide repositories that could be useful for you project:
  - UCI: `http://www.ics.uci.edu/~mlearn/MLRepository.html`
  - Kaggle: `https://www.kaggle.com/datasets`
- Each team must submit the code with a readme file describing how to run the code
- For full consideration, experiments must be reproducible given the (link to the) data, code, and the readme file

9.4 Final Report

The Final report is 3-4 pages. It must include:

- Title
- Introduction (including problem definition and motivation)
- Proposed method and the idea behind it (e.g. why should the method work)
- Experimental results and analysis (e.g. why the results look like this)
- Conclusions

9.5 Presentation

- A presentation should be no longer than 10 minutes, and will be followed by a Q & A session (no longer than 2 minutes)
- All team members should present

10 Submission

- Homework must be completed by individual students. Final project will be completed by groups of 1, 2, or 3 students. Both the Homework and Final project should be submitted to Blackboard.
- Homework and Final project will be due for submission through Blackboard by Tuesday or Thursday at 11:59 PM (Eastern time). Submission will no longer be accepted after the deadline, and will receive a grade of 0.

11 Grading Scheme

- 42% Homework (6)
- 28% Final project (1)
  - 5% Proposal
  - 9% Data and Code
  - 9% Final report (3-4 pages)
  - 5% Presentation (10 minutes)
- 30% Exams
  - 10% Midterm Examination
  - 20% Final Examination

12 Grade Appeals

- A grade becomes permanent one week after you receive the grade
- Grade appeals and questions must be raised in writing (email) within one week after the day on which the grade was received

13 Letter Grade Distribution

$[93, 100]$	A
$[90, 93)$	A-
$(87, 90)$	B+
$[83, 87]$	B
$[80, 83)$	B-
$(77, 80)$	C+
$[73, 77]$	C
$[70, 73)$	C-
<70	F

14 University Policies

14.1 University Policy on Observance of Religious Holidays

In accordance with University policy, students should notify faculty during the first week of the semester of their intention to be absent from class on their day(s) of religious observance. For details and policy, see: `https://students.gwu.edu/accommodations-religious-holidays`

14.2 Academic Integrity Code

Academic dishonesty is defined as cheating of any kind, including misrepresenting one's own work, taking credit for the work of others without crediting them and without appropriate authorization, and the fabrication of information. For details and complete code, see: `https://studentconduct.gwu.edu/code-academic-integrity`

14.3 Safety and Security

In the case of an emergency, if at all possible, the class should shelter in place. If the building that the class is in is affected, follow the evacuation procedures for the building. After evacuation, seek shelter at a predetermined rendezvous location.

15 Support for Students Outside the Classroom

15.1 Disability Support Services (DSS)

Any student who may need an accommodation based on the potential impact of a disability should contact the Disability Support Services office at 202-994-8250 in the Rome Hall, Suite 102, to establish eligibility and to coordinate reasonable accommodations. For additional information see: `https://disabilitysupport.gwu.edu/`

15.2 Mental Health Services 202-994-5300

The University's Mental Health Services offers 24/7 assistance and referral to address students' personal, social, career, and study skills problems. Services for students include: crisis and emergency mental health consultations confidential assessment, counseling services (individual and small group), and referrals. For additional information see: `https://counselingcenter.gwu.edu/`

16 Tentative Schedule

Class Date	Topic	Assignment Given	Assignment Due
05/22	Python: jupyter notebook, python syntax, data type, and control flow		
05/24	Python: numpy, scipy, and function	Homework 1 Given	
05/29	Python: object orient programming		Homework 1 Due
05/31	Python: pandas and matplotlib Data visualization	Homework 2 Given	
06/05	Midterm Data preprocessing		Homework 2 Due
06/07	Data preprocessing (continued)	Homework 3 Given	
06/12	Linear model: linear regression	Homework 4 Given	Homework 3 Due
06/14	Linear model: logistic regression	Homework 5 Given	Homework 4 Due
06/19	Linear model: perception & Adaline	Homework 6 Given	Homework 5 Due
06/21	Case study (putting everything together)		Homework 6 Due Final project Code and Report Due
06/26	Presentation		
06/28	Final exam Preparation for Machine Learning I		