MGT-469 Machine Learning With Python

Professor Yinghui (Catherine) Yang
yiyang@ucdavis.edu

Wednesday, 1:10-5:00pm 9/25, 10/2, 10/9, 10/16, 10/23, 10/30, 11/6, 11/13, 11/20, 12/4

Office Hours will be announced at the start of classes.
Zoom link for Office Hours: https://ucdavis.zoom.us/j/3085360932

Course Description

This course introduces popular machine learning methods to address big data problems businesses face. It covers topics in clustering, association rules for market basket analysis, classification, numeric prediction, model evaluation, etc. This course provides an entry point for students to use Python to apply machine learning models to analyze various types of data.

This is a very hands-on class. You will be learning Python to process data and perform data mining and machine learning tasks on different types of data. You do not need to have a programming background, but need to be comfortable with learning to program in Python.

The course is recommended for students interested in understanding the techniques and applications of using data mining and machine learning methods to process data, and also for students who are interested in learning Python from scratch and using it to program and analyze data.

After taking this class, the students should be able to understand the data mining process, comprehend several popular machine learning and data mining methods, use Python to apply data mining and machines learning methods on different types of data, and also formulate problems for a given data set.

Prerequisite: There are no prerequisites for this course. You do not need coding experience to take this class. You are recommended to obtain some basic statistics knowledge before taking this class (e.g. conditional probability and linear regression).

Course Materials

Required Book 1:
Shmueli, G., Bruce, P. C., Gedeck, P., & Patel, N. R. (2019). *Data mining for business analytics: Concepts, techniques, and applications in Python*. Wiley. ISBN-13: 978-1119549840

Required Book 2:

Matthes, E. (2023). *Python crash course: A hands-on, project-based introduction to programming* (3rd ed.). No Starch Press. ISBN-13: 978-1718502703

Assignment and Assessment Information

There will be nine assignments total. Late homework submissions will not be accepted because the solution will be discussed right after the due time. There will be one take-home final exam.

Course Assessments

Assignment/Assessment	Points	Weight on Final Grade
Assignment 1	5	5%
Assignment 2	5	5%
Assignment 3	5	5%
Assignment 4	6	6%
Assignment 5	7	7%
Assignment 6	9	9%
Assignment 7	4	4%
Assignment 8	11	11%
Assignment 9	8	8%
Final Exam (Take Home)	30	30%
Class Participation	10	10%

Course Schedule

Note: The assigned readings may help you with your assignments and deepen your understanding of the class materials. If you understand perfectly the materials covered in class, feel free to skip the readings.

Week 1 – Introduction and Clustering 1

Main Topics:

- Four Data Mining Problems
- Data Mining Process
- Clustering

Other items:

- Get to know each other
- Discuss syllabus

- Install Python
- Run some basic Python
- More exercises if time allows

Reading: Book 1, Chapters 1.1–1.6, 2.1–2.3, 15.1–15.3, 15.5

Week 2 – Python 1 and Clustering 2

Main Topics:

- Introduction to Python
- Use Python for Clustering

Other items:

- Discuss Homework 1 solution
- Install Association Rule package
- Clustering in Python: Cluster profiles, weights
- Cluster for another data set
- Exercises: Python and clustering
- Expand on functions and classes, if time allows

Readings: No reading

Week 3 – Python 2

Main Topics:

- Python Data Types
- Python: Conditions, Loops, Files

Other Items:

- Discuss Homework 2 solution
- More Python exercises

Readings: Book 2, Chapters 2, 3, 4, 5, 6, 7 (You can gradually go through these chapters over several weeks.)

Week 4 – Association Rule 1

Main Topics:

- Association Rule: Definition and Metrics
- Association Rule: Discovery
- Python: Association Rules for Transaction Data

Other Items:

- Discuss Homework 3 solution
- More Python exercises

- Exercises: Rules for transactions
- Exercises: Python—Rules for transactions

Readings: Book 1, Chapter 14.1

Week 5 – Association Rule 2

Main Topics:

- Association Rule: Non-Transaction Data
- Python for Finding Association Rules

Other Items:

- Discuss Homework 4 solution
- Explore Python association rule results more, more functions to filter results
- More Python exercises
- Exercises: Rules for non-transactional data
- Discover rules on a new data set

Readings: No readings

Week 6 – Classification 1

Main Topics:

- Classification Methods
- Classification: Overfitting, Stratified Sampling, Confusion Matrix

Other Items:

- Discuss Homework 5 solution
- Lift curve in theory

Readings: Book 1, Chapters 9.1–9.6, 9.8, 5.1-5.4, 10

Week 7 – Classification 2

Main Topics:

- Classification with Python
- Cross-Validation
- Models

Other Items:

- Discuss Homework 6 solution
- Lift curve on real results
- Classification using another dataset

Readings: No readings

Week 8 – Numeric Prediction

Main Topics:

- Numeric Prediction
- Numeric Prediction: Python

Other Items:

- Discuss Homework 7 solution
- Additional topics on numeric prediction
- Numeric prediction on another dataset

Readings: Book 1, Chapters 9.7, 4.5, 5.2, 6

Week 9 – Problem Structure and Miscellaneous Topics

Main Topics:

- Python Functions
- Classification: Class Weight
- Construct Other Problems on IBM Data
- Missing Data

Other Items:

- Discuss Homework 8 solution
- More exercises on Python functions
- Look into the six datasets, explore missing data

Readings: Book 2, Chapter 8; Book 1, Chapter 5.5

Week 10 – Data Sets: Four Problems

Main Topics:

- One data set, Four problems
- Classification and Numeric Prediction
- Clustering and Association Rules

Other Items:

- Discuss Homework 9 solution
- Go through the six datasets to structure problems

Readings: No readings

Code of Academic Conduct: http://sja.ucdavis.edu/files/cac.pdf