MGV-269V Machine Learning With Python

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.

There is no official prerequisite for this course, but you should have some basic knowledge of statistics (e.g., probability theory and linear regression) before taking this class.

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. (2019). *Python crash course: A hands-on, project-based introduction to programming* (2nd ed.). No Starch Press. ISBN-13: 978-1593279288

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	30	30%
Class Participation	10	10%

Grading Scale

Letter Grade	Percent Grade
A+	97–100
A	93–97
A-	90–93
B+	87–90
B	83–87
B-	80–83
C+	77–80
C	73–77
C-	70–73
D+	67–70
D	65–67
E/F	65 and below

Assignment and Assessment Information

There will be nine assignments total, one for each week from Week 1 to Week 9. Each assignment will be due one hour before the Live Session. Late homework submissions will not be accepted because the solution will be discussed right after the deadline.

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 the asynchronous work and Live Sessions, feel free to skip the readings.

Make sure you work on the exercises included in the asynchronous work, since we will discuss them in the Live Session.

Week 1 – Introduction and Clustering 1

Asynchronous materials:

- Four Data Mining Problems
- Data Mining Process
- Clustering

Live Session plan:

- Get to know each other
- Discuss syllabus
- Q&A for asynchronous materials
- 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

Asynchronous materials:

- Introduction to Python
- Use Python for Clustering

Live Session plan:

- Discuss Homework 1 solution
- Q&A for asynchronous materials
- 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

Asynchronous materials:

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

Live Session plan:

- Discuss Homework 2 solution
- Q&A for asynchronous materials
- 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

Asynchronous materials:

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

Live Session plan:

- Discuss Homework 3 solution
- Q&A for asynchronous materials
- More Python exercises on asynchronous materials
- Exercises: Rules for transactions
- Exercises: Python—Rules for transactions

Readings: Book 1, Chapter 14.1

Week 5 – Association Rule 2

Asynchronous materials:

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

Live Session plan:

- Discuss Homework 4 solution
- Q&A for asynchronous materials
- 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

Asynchronous materials:

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

Live Session plan:

- Discuss Homework 5 solution
- Q&A for asynchronous materials
- Lift curve in theory

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

Week 7 – Classification 2

Asynchronous materials:

- Classification with Python
- Cross-Validation
- Models

Live Session plan:

- Discuss Homework 6 solution
- Q&A for asynchronous materials
- Lift curve on real results
- Classification using another dataset

Readings: No readings

Week 8 – Numeric Prediction

Asynchronous materials:

- Numeric Prediction
- Numeric Prediction: Python

Live Session plan:

- Discuss Homework 7 solution
- Q&A for asynchronous materials
- 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

Asynchronous materials:

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

Live Session plan:

- Discuss Homework 8 solution
- Q&A for asynchronous materials
- 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

Asynchronous materials:

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

Live Session plan:

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

Readings: No readings

Attendance Policy

You can get up to one point for each week. You will be assessed based on the exercise solutions for the asynchronous materials and your participation in the Live Session.

Program Mission

The mission of the UC Davis Graduate School of Management is to be a global leader in management research and education. As part of the world's premier public university system, we pursue significance, excellence, and scholarly rigor in our research, teaching, and service to the people of California. We emphasize curiosity, creativity, and high standards in the generation and transmission of theoretical and practical knowledge relevant for business.

Honor Code and Academic Integrity

Academic integrity exists when students and faculty seek knowledge honestly, fairly, with mutual respect and trust, and accept responsibility for their actions and the consequences of those actions. Without academic integrity, there can be no trust or reliance on the effectiveness, accuracy, or value of a

University's teaching, learning, research, or public service activities. It is therefore key that we understand what academic integrity is, why it is important, and how to help it flourish on college campuses.

1. It is expected that all class members will treat each other with respect and dignity.

2. It is not acceptable behavior to insult, harass, or demean any member of the class.

3. Professional business behavior should be modeled in the classroom, including the use of appropriate language, jokes, or stories.

In general, students should adhere to the UC Davis Principles of Community, copied below.

The University of California, Davis, is first and foremost an institution of learning and teaching, committed to serving the needs of society. Our campus community reflects and is a part of a society comprising all races, creeds, and social circumstances. The successful conduct of the university's affairs requires that every member of the university community acknowledge and practice the following basic principles:

We affirm the inherent dignity in all of us, and we strive to maintain a climate of justice marked by respect for each other. We acknowledge that our society carries within it historical and deep-rooted misunderstandings and biases, and therefore we will endeavor to foster mutual understanding among the many parts of our whole.

We affirm the right of freedom of expression within our community and affirm our commitment to the highest standards of civility and decency towards all. We recognize the right of every individual to think and speak as dictated by personal belief, to express any idea, and to disagree with or counter another's point of view, limited only by university regulations governing time, place, and manner. We promote open expression of our individuality and our diversity within the bounds of courtesy, sensitivity, and respect.

We confront and reject all manifestations of discrimination, including those based on race, ethnicity, gender, age, disability, sexual orientation, religious or political beliefs, status within or outside the university, or any of the other differences among people which have been excuses for misunderstanding, dissension, or hatred. We recognize and cherish the richness contributed to our lives by our diversity. We take pride in our various achievements, and we celebrate our differences.

We recognize that each of us has an obligation to the community of which we have chosen to be a part. We will strive to build a true community of spirit and purpose based on mutual respect and caring.

For more information, please review the Academic Conduct Booklet:
https://gsm.ucdavis.edu/sites/default/files/2020-10/code_of_conduct_booklet_2020.pdf