COMP 4432: Machine Learning

Course Overview

This course explores machine learning techniques and theory. The course covers how to use popular machine learning libraries to develop, train, evaluate, and deploy predictive models on prepared data. Both design principles (machine learning types and tasks) and technical tools/languages will be covered.

Prerequisites: COMP3006, COMP4442

Learning Outcomes

After taking this course, students will be able to:

- Distinguish between supervised and unsupervised machine learning algorithms.
- Pre-process data and feature engineer to meet modeling requirements.
- Select and employ appropriate Python libraries for training, testing, and validation of parametric and nonparametric predictive models.
- Employ appropriate strategies to mitigate model overfitting and underfitting.
- Optimize models via. hyperparameter selection and tuning.
- Articulate the differences between ensemble and boosting algorithms.
- Evaluate model results using appropriate objective functions.
- Describe the strengths, limitations and tradeoffs of a given model.
- Present model results in relation a problem and/or research question in an effective and persuasive manner.

Textbooks and Materials

Required textbooks for COMP 4432 Machine Learning:

Géron, A. (2019). *Hands-on machine learning with scikit-learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems* (2nd ed.). O'Reilly Media.

Recommended reading:

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013.An introduction to statistical learning. New York: Springer. Free PDF version (2nd ed): https://www.statlearning.com/.

Bishop, C.M. (2006). Pattern Recognition and Machine Learning. Springer. Free PDF version. https://tinyurl.com/y6b8z5qv.

Alpaydin, E. (2020). *Introduction to Machine Learning* (4th ed.). The MIT Press.

Assignment datasets:

Assignment datasets are located at: https://github.com/arjayit/cs4432_data. Instructors will provide additional guidance when they assign each deliverable. Please reach out to your instructor if you have any questions or trouble accessing the data sets.

Grading

Assignment/Assessment	Points	Weight on Final Grade
Assignment 1	100	20%
Assignment 2	100	20%
Assignment 3	100	20%
Assignment 4	100	20%
Assignment 5	100	20%

Grading Scale

A 93-100   A- 90-92.99

B+ 86-89.99 B 83-85.99 B- 80-82.99

C+ 76-79.99 C 73-75.99 C- 70-72.99

D+ 66-69.99 D 63 - 65.99 D- 60-62.99 F < 60

Weekly Schedule

There will be a graded assignment assigned each odd week and due the following week by midnight the day prior to the live session. The schedule also includes many asynchronous exercises in addition to the assignments. Please complete each week's asynchronous exercises 24 hours before each live session.

Week 1. Machine Learning Basics
Readings:
- Reading 1: Géron, Chapter 1

Week 2. Data Analytics Project and Process Management
Readings:
- Reading 1: Géron, Chapter 2

Complete Assignment 1

Week 3. Classification
Readings:
- Reading 1: Géron, Chapter 3

Week 4. Model Training
Readings:
- Reading 1: Géron, Chapter 4

Complete Assignment 2

Week 5. Support Vector Machines
Readings:
- Reading 1: Géron, Chapter 5

Week 6. Decision Trees
Readings:
- Reading 1: Géron, Chapter 6

Complete Assignment 3

Week 7. Ensemble Trees
Readings:
- Reading 1: Géron, Chapter 7

Week 8. Clustering – K Means

Readings:

- Reading 1: Géron, Chapter 9.
- Reading 2: Luxburg, U.V., Williamson, R.C., and I. Guyon. 2012. *Clustering: Science or Art*? JMLR: Workshop and Conference Proceedings 27:65–79. http://proceedings.mlr.press/v27/luxburg12a/luxburg12a.pdf

Complete Assignment 4

Week 9. Clustering Gaussian Mixtures

Readings:

- Reading 2. Saxena et al. 2017. *A review of clustering techniques and developments*. Neurocomputing 267: 664-681.

Week 10. Hierarchical and Density-Based

Readings:

- Reading 1: Wegmann et. al. *A review of systematic selection of clustering algorithms and their evaluation*. arXiv:2106.12792.

Complete Assignment 5

Attendance Policy

Attendance at all live session meetings is expected.

Program Mission

Our MS in Data Science provides students with a broad course of study in programming, algorithms, statistics, and data management, as well as a depth of understanding in specific fields such as data mining, machine learning, and parallel systems. Graduates of the data science program go on to work in a wide variety of careers, including business, government, education, and the natural sciences.

Honor Code and Academic Integrity

All students are expected to abide by the University of Denver Honor Code. These expectations include the application of academic integrity and honesty in your class participation and assignments. Violations of these policies include but are not limited to

- Plagiarism, including any representation of another's work or ideas as one's own in academic and educational submissions
- Cheating, including any actual or attempted use of resources not authorized by the instructor(s) for academic submissions
- Fabrication, including any falsification or creation of data, research, or resources to support academic submissions

Violations of the Honor Code may have serious consequences including, but not limited to, a zero for an assignment or exam, a failing grade in the course, and reporting of violations to the Office of Student Conduct.

Diversity, Inclusiveness, Respect

DU has a core commitment to fostering a diverse learning community that is inclusive and respectful. Our diversity is reflected by differences in race, culture, age, religion, sexual orientation, socioeconomic background, and myriad other social identities and life experiences. The goal of inclusiveness, in a diverse community, encourages and appreciates expressions of different ideas, opinions, and beliefs, so that conversations and interactions that could potentially be divisive turn instead into opportunities for intellectual and personal enrichment.

A dedication to inclusiveness requires respecting what others say, their right to say it, and the thoughtful consideration of others' communication. Both speaking up AND listening are valuable tools for furthering thoughtful, enlightening dialogue. Respecting one another's individual differences is critical in transforming a collection of diverse individuals into an inclusive, collaborative, and excellent learning community. Our core commitment shapes our core expectation for behavior inside and outside of the classroom.