DATASCI 207: Applied Machine Learning

Master of Information and Data Science (MIDS)

Course Designer: Dan Gillick

Course Description

This course provides a practical introduction to the rapidly growing field of machine learning—training predictive models to generalize to new data. We start with linear and logistic regression and implement gradient descent for these algorithms, the core engine for training. With these key building blocks, we work our way to understanding widely used neural network architectures, focusing on intuition and implementation with TensorFlow/Keras. While the course centers on neural networks, we will make sure to cover key ideas in unsupervised learning and nonparametric modeling. Along the way, weekly short coding assignments and a midterm exam will connect lectures with data and real applications. A more open-ended final project will tie together concepts in experimental design and analysis with models and training.

Prerequisites

Required

- Core data science courses: W201 and W203
- Intermediate knowledge of Python (we will use NumPy and matplotlib libraries extensively)
- Basics of linear algebra (operations on vectors and matrices)

Useful but optional

- Basics of multivariate calculus
- Experience with Jupyter Notebook or Colab
- Experience with Git and GitHub

Learning Objectives

By the completion of this course, students will be able to:

1. demonstrate a familiarity with a wide range of concepts in the field of machine learning and neural networks in particular.

2. understand and apply the concepts of machine learning using techniques and tools common in industry.
3. demonstrate proficiency with existing coding languages (e.g., Python), packages related to machine learning (numpy, matplotlib, sklearn, and tensorflow), and the application of appropriate machine learning approaches for data science problems and questions.
4. direct their own learning in new and emerging machine learning tools and approaches by navigating API documentation and engaging in experimentation.
5. maintain familiarity with up to date knowledge in the machine learning domain through reading current research papers, and extracting applicable, practical aspects of emerging research and work.

References

No textbook is required, but we suggest some readings from the following sources:

- Chollet, F. *Deep learning with Python*. (Available for free for Berkeley students)
- *Dive into deep learning*. (Free and open source)
- Raschka & Mirjalili (RM). *Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2*. (Available for free for Berkeley students)

Weekly Schedule

Topics and suggested readings for each week of the course.

1. Introduction and Framing
    1. Chollet (1.1, 1.2, 1.3)
    2. RM (1)
2. Linear Regression and Gradient Descent
    1. Chollet (2.2)
    2. RM (10)
3. Feature Engineering
    1. Chollet (5)
    2. RM (4)
4. Logistic Regression
    1. Chollet (3.5)
    2. RM (3)
5. Multiclass Classification and Metrics
    1. Chollet (6)
    2. RM (3)
6. Feedforward Neural Networks
    1. Chollet (2.3, 2.4, 2.5)
    2. RM (12, 13)
    3. Neural network explainer videos (Grant Sanderson)
    4. Backpropagation derivation (Jeremy Jordan)

    5. TensorFlow Playground (Daniel Smilkov and Shan Carter)
7. KNN, Decision Trees, and Ensembles
    1. RM (3, 7)
8. Unsupervised Learning: k-Means and PCA
    1. RM (11)
9. Embeddings for Text
    1. Chollet (4.1, pp. 524–534)
    2. RM (8)
10. Convolutional Neural Networks
    1. Chollet (8.1, 8.2)
    2. RM (15)
11. Network Architecture Design
    1. Chollet (7.2)
    2. In-class demo code
12. Fairness
    1. Essay: "Physiognomy's New Clothes" (Blaise Agüera y Arcas, Margaret Mitchell, and Alexander Todorov)
13. Advanced Topics
14. Project Presentations

Grading

- Assignments (individual work): 45%
- Midterm: 20%
- Final project (group work): 30%
- Participation: 5%

Weekly homework assignments can be submitted up to 3 days late with a 10% (absolute) penalty per day. The lowest grade will be dropped.

Baseline grading range for this course is:
A for 93 or above,
A– for 90 or above,
B+ for 87 or above,
B for 83 or above,
B– for 80 or above,
C+ for 77 or above,
C for 73 or above,
C– for 70 and above,
D+ for 67 and above,
D for 63 and above,
D– for 60 and above, and
F for 59 and below.

Homework Assignments and Midterm

In Weeks 1–10, students will work on weekly assignments that connect concepts in the lectures with concrete tasks. In Week 6, students will complete a take-home midterm exam.

Final Project

In Weeks 11–14, students will work on a group project (ideally three to four people) that is more open ended. Groups can choose from a few recommended topics (hosted on Kaggle) or potentially select their own. Grading will reflect individual contributions, group size, and the application of concepts from the course to the selected task, including data analysis, modeling, experiments, and analysis of errors.

Diversity and Inclusion

Integrating a diverse set of experiences is important for a more comprehensive understanding of machine learning. We will make an effort to read papers and hear from a diverse group of practitioners, still, limits exist on this diversity in the field of machine learning. We acknowledge that it is possible that there may be both overt and covert biases in the material due to the lens with which it was created. We would like to nurture a learning environment that supports a diversity of thoughts, perspectives and experiences, and honors your identities (including race, gender, class, sexuality, religion, ability, veteran status, etc.) in the spirit of the UC Berkeley Principles of Community.

To help accomplish this, please contact your instructor or submit anonymous feedback through iSchool channels if you have any suggestions to improve the quality of the course. If you have a name and/or set of pronouns that you prefer we use, please let your instructor know. If something was said in class (by anyone) or you experience anything that makes you feel uncomfortable, please talk to your instructor about it. If you feel like your performance in the class is being impacted by experiences outside of class, please don't hesitate to talk with your instructor. We want to be a resource for you. Also, anonymous feedback is always an option, and may lead to your instructor to make a general announcement to the class, if necessary, to address your concerns.

As a participant in teamwork and course discussions, you should also strive to honor the diversity of your classmates.

If you prefer to speak with someone outside of the course, MICS Academic Director Lisa Ho, I School Assistant Dean of Academic Programs Catherine Cronquist Browning, and the UC Berkeley Office for Graduate Diversity are excellent resources. Also see the following link: https://www.ischool.berkeley.edu/about/community.

Attendance and Participation

We believe in the importance of the social aspects of learning: between students, and between students and instructors, and we recognize that knowledge-building is not solely occurring on an individual level, but that it is built by social activity involving people and by members engaged in the activity. Participation and communication are key aspects of this course that are vital to the learning experiences of you and your classmates.

Therefore, we like to remind all students of the following requirements for live class sessions:

- Students are required to join live class sessions from a study environment with video turned on and with a headset for clear audio, without background movement or background noise, and with an internet connection suitable for video streaming.

- You are expected to engage in class discussions, breakout room discussions and exercises, and to be present and attentive for your and other teams' in-class presentations.

- Keep your microphone on mute when not talking to avoid background noise. Do your best to minimize distractions in the background video, and ensure that your camera is on while you are engaged in discussions.

That said, in exceptional circumstances, if you are unable to meet in a space with no background movement, or if your connection is poor, make arrangements with your instructor (beforehand if possible) to explain your situation. Sometimes connections and circumstances make turning off video the best option. If this is a recurring issue in your study environment, you are responsible for finding a different environment that will allow you to fully participate in classes, without distraction to your classmates. Please contact Student Affairs if you have problems meeting these requirements.

Failure to adhere to these requirements will result in an initial warning from your instructor(s), followed by a possible reduction in grades or a failing grade in the course.