EC 414: Introduction to Machine Learning
Fall, 2021

Lecture Time

MW 2:30-4:15pm, CAS B20

Discussion Sections

Two discussion sections are offered, one Friday 10:10-11am and the other 12:20-1:10pm. The 10:10am discussion is in PHO 202 and the 12:20pm discussion is in PHO 201. Both discussions should be attended either in-person.

Staff Information

Instructor: Prof. Brian Kulis
Office: PHO 441
Email: bkulis@bu.edu (please use EC414 in the subject line when emailing)
Office hours: Monday 4:30-5:30pm, PHO 441

GTAs/Graders:
Hoang Tran (tranhp@bu.edu)
Office hours: TBD

Junyu Liu (jyliu@bu.edu)
Office hours: TBD

Zeyu Gu (zgu@bu.edu)
Office hours: TBD

Benyamin Trachtenberg (btt@bu.edu)
Office hours: TBD

Tao Zhang (mtao@bu.edu)
Office hours: TBD

Course Description

This course will introduce students to modern techniques of data analysis and machine learning. The development and deployment of such tools has become an increasingly integral component in many domains. For instance, Google now fundamentally relies on data analysis and machine learning to drive their businesses (as do many other companies); other enterprises such as those in bioinformatics, engineering, and finance have also been reshaped by the emergence of such tools. It is increasingly important to equip students entering the workforce with the knowledge to be able to understand and employ these tools in a variety of contexts. As such, this course will overview many of the basic tools of data analysis and machine learning---regression, classification, data visualization, and clustering---as well as touch upon some of the recent progress in this area, most notably deep learning and large-scale neural networks. Further, the course will expose students to programming assignments involving real-world data.

Prerequisites

Linear Algebra (EK 102 / EK 103)
Probability and Statistics (EK 381 or equivalent)
Programming: EK 125 (additional exposure to programming at the level of EC 330 is useful but not required)

Topics

Introduction to machine learning (regression, classification, and unsupervised learning), statistical tools for data analysis (maximum likelihood, MAP estimation), data visualization, and an overview of recent topics such as deep learning:

Review of Linear Algebra and Probability [1 week]
Linear Regression (Least Squares) [1 week]
Machine Learning Basics and Simple Classification Techniques (Curse of Dimensionality, Overfitting, k-NN) [1 week]
Maximum Likelihood and MAP estimation (with Linear Regression revisited) [1 week]
Linear Classification (Naive Bayes, Linear SVM, Logistic Regression) [2 weeks]
Data Visualization and Feature Extraction (Projections, SVD, PCA) [2 weeks]
Clustering (k-means, Gaussian mixtures) [2 weeks]
Intro to Deep Learning (Backpropagation, MLP, overview of modern Deep Learning) [2 weeks]
Other topics (Kernel methods, Recommendation Systems; Applications to Vision, Speech and/or Natural Language Processing; Sequential Data; etc.) [1-2 weeks]

Webpage

Announcements, course material, readings, and an updated schedule will be posted on Blackboard. We also have set up piazza for this course. Grading will be done on gradescope.

Textbooks

K. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012. This book is exhaustive, and covers nearly all of the topics in the course (and many more that we will not have time to cover). It is an excellent reference to own but it is more probability-heavy than we will be. I will not assign readings from it, so it's not really required per se, but it is a great resource.

Another good reference is C. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006. This is also more probabilistic (Bayesian) than we will be.

Homework

There will be weekly homework during the course. We will be using gradescope for submitting all homework in this class. *Homeworks are released on Wednesday and due on the following Wednesday at the beginning of class, unless otherwise specified.* The homeworks will be a mix of pen-and-paper exercises and programming assignments. **Programming will be done in Python**; we will do some review of Python at the beginning of the semester. For most assignments, we will encourage students to use Google Colab.

Important notes on homework:

Homeworks can be done in groups (and in fact this is highly encouraged), but homework exercises need to be written up individually. Also, write down the names of any students with whom you collaborated. You must be able to fully explain your answers on demand, if necessary. You may not use resources outside of class including web-based services, etc; failure to do so may be considered plagiarism and will be taken very seriously.

Regrades can be requested via gradescope. The graders will take a look at these requests periodically. We will typically not look at regrade requests that are only arguing about the amount of partial credit.

The late policy for homeworks is that every day that the homework is late, there is a 15 percent penalty; after 3 days no credit will be given. So, up to one day late, you can receive a maximum of 85%; up to two days late, a maximum of 70%, and up to three days late, a maximum of 55%. After three days, you will receive a zero. Note: late penalties will be shown under the first question of each HW on Gradescope.

Exams

There will be two midterm exams during the class time (see schedule for dates). There will be a final exam during the normal final exam slot; this exam will be cumulative.

Grade Breakdown

Midterms: 45% (22.5% each)
Final: 30%
Homework: 25%