The City College of New York

CSc 59929 – Introduction to Machine Learning
Spring 2018
Professor Erik K. Grimmelmann, Ph.D.

Syllabus

Overview

This course will provide a theoretical and hands-on introduction to the basics of machine learning and its application to various real-world problems.

Prerequisites

You are expected to have a basic knowledge of computer science including object-oriented programming, algorithms, data structures, and data analysis and a working knowledge of Python. In addition, you are expected to have a rudimentary understanding of probability, statistics, and basic linear algebra.

The formal prerequisites are

- CSc 212: Data Structures
- CSc 217: Probability and Statistics for Computer Science
- CSc 220: Algorithms
- CSc 221: Software Design Laboratory
- Math 346: Elements of Linear Algebra

If you have not taken (and done well in) all of these courses (or their equivalents), please check with me prior to continuing in the course.

Class Meetings

We will meet **Tuesdays, 11:00 am to 1:30 pm in NAC 7/312**. Two and a half hours is a long time to sit in one place, so we'll take one or two short breaks during each class.

Please note that per the City College calendar, no classes for this course are scheduled on February 20 and on April 3.

Please **arrive promptly**. Arriving late is better than not arriving at all, but please allow adequate time to get to campus, especially if you come by public transit.

Attendance in class is required. If you have an unavoidable conflict (e.g., a job interview) or are ill, please let me know via email, preferably in advance. If you're sick with a cold or the flu, stay home and recover. Consistent unexcused absences are not okay.

Readings

The required textbook is

Raschka, Sebastian, and Vahid Mirjalili. *Python machine learning: machine learning and deep learning with Python, scikit-Learn, and TensorFlow*. Packt, 2017. Note that this is the second edition of this book; the first edition had only Raschka as the author and a slightly different title.

In general, the lectures will not be drawn from the textbook but rather from a number of other works in the field. I'll be letting you know which sources I've used for the topics we cover.

A bibliography with suggested other references will be posted on the Blackboard site.

Guest Lecture

Depending on their availability, I will be trying to bring some of my friends and mentors in to speak to the class. I'll let you know when a guest lecture is scheduled. It's unlikely that any guest lecture will go for the full two and a half hours.

Programming

All programming in the course will be in Python and its relevant add-ins and libraries (such as NumPy, MatPlotLib, Pandas, and Scikit-Learn) and TensorFlow. You are free to work in any environment that supports Python (e.g., Windows, Max, Unix, Linux). We'll be using Jupyter notebooks throughout the course.

Blackboard

We will be using Blackboard as our online environment. Once you're enrolled in the course and the course has started, you should have access to the Blackboard course site. We will use the course site for

- This syllabus
- Links to reference materials
- Announcements
- Posting and submission of assignments
- Classroom presentations (typically within a few days of the class session.
- Datasets
- Sample code
- Assignment grades

Course grades will not be posted on Blackboard but rather on CUNYfirst.

Course Policies

Except where I tell you otherwise, you are free to collaborate freely with each other and to consult any sources you wish to in your work for this class.

I expect you to act professionally and respectfully to your classmates (and our occasional guests) at all times. Harassment will not be tolerated.

If for any reason your preferred name is not the one that appears on the course roster, please let me know how you would like to be addressed. I am not good at remembering names so I will asking you to submit headshots on Blackboard.   These headshots will be for my use only; submitting a headshot is optional.

Grades

Your grades will be based on the following factors:

- Class attendance 10%
- Class participation 10%
- Individual programming exercises 40%
- Final project presentation 10%
- Final project submission 30%

Integrity

Just to refresh your memory, here's the City College statement on academic integrity:

> Academic integrity is an essential part of the pursuit of truth, and of your education. We are all are all responsible for maintaining academic integrity at City College – it is the rock on which the value of your degree is built.
>
> If you cheat on a test or plagiarize by using someone else's work or ideas, you defeat the purpose of your education. In addition, academic dishonesty is prohibited in the City University of New York, and is punishable by failing grades, suspension and expulsion.

Here's a link to a list of City College and CUNY policies (and links to them), https://www.ccny.cuny.edu/about/policies

If you use code from any source other than your own imagination for any coding assignment, be sure to list the source(s).

Your feedback

I welcome your feedback at all points in the course. If something is unclear, please speak up. If you find an error in my lectures, code examples, assignments, or in anything else, please let me know.

My Contract Information

The best (and fastest) way to reach me is via email at egrimmelmann@ccny.cuny.edu.

Office Hours

My office hours will be shortly after class each week, from 1:45 pm to 3:00 pm. If you haven't let me know that you'll be coming to office hours and no one is there at 2:15 pm, I may leave.

Office hours will be held in NAC 8/207, the conference room directly across the hall form the Computer Science department office.

In special cases, I can arrange to meet you at another time. Occasionally I will have a conflict with the standard time for office hours; when this is the case I'll let you know in advance.

Course Schedule

The schedule below is only approximate. We may end up going faster or slower on some of the topics, so we could end up being ahead or behind of the schedule at varying points in the course.

Week 1 (January 30)
- Course introduction
- Context for the growth in Machine Learning
- Introduction to Machine Learning

Week 2 (February 6)
- The Perceptron

Week 3 (February 13)
- Adaptive Linear Neuron
- Polynomial curve fitting

Week 4 (February 27)
- Multiclass Classification
- Logistic Regression

Week 5 (March 6)
- Support Vector Machines

Week 6 (March 13)
- Decision Trees
- Pruning Decision Trees

Week 7 (March 20)
- Strong and weak learners
- Gradient boosting and AdaBoost

Week 8 (March 27)
- Bootstrapping
- Random Forests

Week 9 (April 10)

- Principal Component Analysis
- K-Nearest Neighbors
- Regression
- Project Previews

Week 10 (April 17)

- Neural Nets
- Deep Neural Nets

Week 11 (April 24)

- Convolutional Nets
- Recurrent and Recursive Nets

Week 12 (May 1)

- Reinforcement Learning
- Project Previews

Week 13 (May 8)

- Project Presentations

Week 14 (May 15)

- Ethics and Fairness in Machine Learning