INFO-4604, Applied Machine Learning
University of Colorado Boulder

August 28, 2018

Prof. Michael Paul

People

- Professor Michael Paul
  - 4th year at CU
  - 2nd year teaching this course
  - mpaul@colorado.edu
- Teaching Assistant: Arcadia Zhang
  - Currently getting PhD at CU
  - arcadia.zhang@colorado.edu

We will each hold 1 hour of office hours each week (2 hours total), time/place TBD; appointments possible

Information

- Public website
  - Lecture slides, readings, policies
  - http://cmci.colorado.edu/classes/INFO-4604/
- Canvas
  - Discussion, assignments, grades
  - https://canvas.colorado.edu/courses/22139
  - Make sure you have access!

What is machine learning?

Murphy:

- “a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data”

Essentially: learning from data

(Learning to do what? We’ll see examples.)

What makes this *applied?*

Compared to other courses…

- More emphasis on using existing tools than implementing algorithms
  - But you'll do a little bit of implementation too
- Less mathematical theory
  - But you'll still learn how the algorithms work
  - Math will be taught as needed
- More focus on creating systems/pipelines (data processing, design, evaluation)

Goals

After this course, you should be able to:

- identify when machine learning can help solve a problem and which approaches are appropriate;
- be comfortable doing machine learning in Python, and be familiar enough with the algorithms and parameters to easily adopt other toolkits;
- understand the underlying concepts well enough that you can read machine learning papers, and can modify implementations for your own needs.

Background

Programming background: Python

- Class time is mostly about concepts, not code; expect to spend some time learning on your own

Where to go for help?

- Ask questions on Canvas
  - Asking (not just answering) questions on Canvas helps your participation grade!
- Look at examples that come with the book; experiment with editing the code so that you understand it better

Background

Math background: nothing specifically assumed (but some math skills needed)

- Prior exposure to discrete math and probability is helpful (e.g., INFO-2301)
- Concepts will be taught as needed

Where to go for help?

- Ask questions on Canvas
- Review the free *OpenIntro Statistics* textbook

4604 vs 5604

Graduate students should be enrolled in 5604

5604 students will have to do additional problems on homework/quizzes/exams, and are assigned additional readings

- 4604 students can do the 5604 problems for extra credit

Attendance

If you need to miss a class, let me know **before** the lecture.

Attendance is required on days that we have quizzes or in-class problems

Laptop Policy

Laptops are not required (but helpful on occasion; I will announce which days)

If you use a laptop in class, please be respectful of your neighbors (nothing distracting on your screen)

Homework

- ~6 assignments
- Jupyter notebooks
  - Combination of programming and written answers (mostly making observations about your results)
- Expect to spend at least 10 hours on each assignment
  - Don’t procrastinate!

Homework

- 5 “late days”
- Once you’ve used up late days:
  - 80% credit within 1 day late
  - 60% credit after 1 day late
  - 0% credit after 7 days
- See website for more details

Homework

We will assign and grade the first programming assignment before the drop deadline (Wed, Sept 12)

- It will be based on the code in Chapter 2 of the textbook
  - First 3 chapters available in Canvas

Quizzes

6-7 quizzes, dates will be announced

- Worth 10% total