CS167: Machine Learning

Syllabus
Logistics
Machine Learning: Introduction

Tuesday, January 30th, 2024

Md Alimoor **Reza**

- Assistant Professor of Computer Science, Dept. of Mathematics and Computer Science, Drake University
- Office: 323 Collier-Scripps
- Email: md.reza@drake.edu
- Phone: 515-271-1972
- Office hours: Mon/Wed: 1:00-3:30pm CDT
  additionally by appointment
  Zoom link

Course Logistics

Contacting

- I am here to help! To get a reply as quickly as possible:
  - For questions about the class, assignments, personal matters, etc: email md.reza@drake.edu

Course mechanics

- Syllabus, schedule, assignments, announcements, etc. on Drake Blackboard
  - https://drake.blackboard.com
  - Syllabus link

- Class meeting times and locations:

  - Location: Meredith Hall #201
  - Time: Tuesday/Thursday 9:30 am - 10:45 am (CRN#10638)
  - Time: Tuesday/Thursday 11:00 am - 12:15 pm (CRN#12201)

Course mechanics

- Office hour times and location:

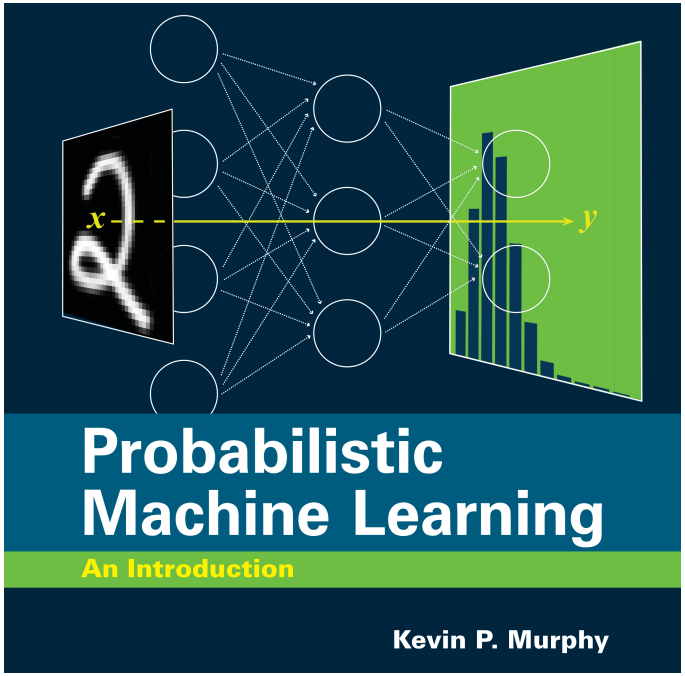
  - Office hour#1: Monday: 1:00 pm - 3:30 pm (tentative)
  - Office hour#2: Wednesday 1:00 pm - 3:30 pm (tentative)
  - Location: Collier-Scripps#323

Course mechanics

- Textbook:
  - We will not follow any particular textbook
  - Optional chapter reading will be assigned from:

    Probabilistic Machine Learning: An Introduction

    - Kevin P. Murphy (2022)



Grading and Requirement

- Notebook assignments (40%): *In-class and take home assignments*
- Quizzes (30%): *3 quizzes, 10% each*
- Projects (20%): *2 projects, each worth 10% of final grade*
- Attendance/Participation (10%): *Participation in polls, not based on correctness, physical attendance during lecture time*

Grading and Requirement

- Notebook assignments (40%): *In-class and take home assignments*.

Notebooks (40%) Throughout the course, we will be learning to apply machine learning principles using Python machine learning tools. Machine learning code is often developed in and communicated using an interactive integrated development environment called Jupyter Notebooks which support a natural interleaving of code, output/results, and mark-up documentation–what you're seeing right now is actually a Jupyter notebook. You will regularly submit notebook files (files with the extension .ipynb) to demonstrate your proficiency with the Python tools we are using. Given the long computation times of the programs you write, I will not usually be executing your code, so it is critical that the results from your executions are preserved in the notebook. You can expect to submit 6 notebooks throughout the course.

Grading and Requirement

- Quizzes (30%): *3 quizzes, 10% each*

Quizzes (30%) There will be 3 quizzes that will be administered via Blackboard. They will not be timed, and you will have a few days to complete them. Quizzes should be completed individually and will be due before class on the following Monday. There is no time limit on these quizzes, but they must be submitted before we convene for the next class. As in the real world, you will be allowed to use external resources like the class notes and the internet. You will be required to cite any sources that you used while completing these quizzes other than the class notes.

Grading and Requirement

- Projects (20%): *2 projects, each worth 10% of final grade*

Projects (20%): The two projects in this course will emphasize the design, execution, and interpretation of machine learning experiments. The grading emphasis will be on how well you explain your data and experiment as well as your written interpretation. For these, you will submit Jupyter Notebooks with more extensive writing in the mark-up cells than for your regular notebook assignments.

Grading and Requirement

- Attendance/Participation (10%): *Participation in polls, not based on correctness*

Attendance/Participation (10%): This class is highly interactive, meaning that active participation is both expected and the norm. You will receive credit for your participation, and it will be counted towards your final grade. I will keep track of your involvement using a signature sheet. Throughout the course, I will pose questions using polling software and conduct in-class Q&A sessions to better understand how the class is grasping the content. These responses will not be evaluated for correctness but rather for completion.

Grading Scale

- The tentative grading scale for this course would be as follows:

A (93%-100%)	A- (90%-92.9%)	B+ (87%-89.9%)
B (84%-86.9%)	B- (80%-83.9%)	C+ (77%-79.9%)
C (74%-76.9%)	C- (70%-73.9%)	D (60%-69.9%)
F (0%-59.9%)		

Course overview

- Brush up on necessary programming tools and libraries
  - Python basics, Pandas library, PyTorch, GitHub, Google Colab, …
- Classical Machine Learning
  - K-nearest-neighbors (k-NN)
  - decision trees
  - random forests (RF)
  - support vector machines (SVM)
  - perceptrons

Course overview

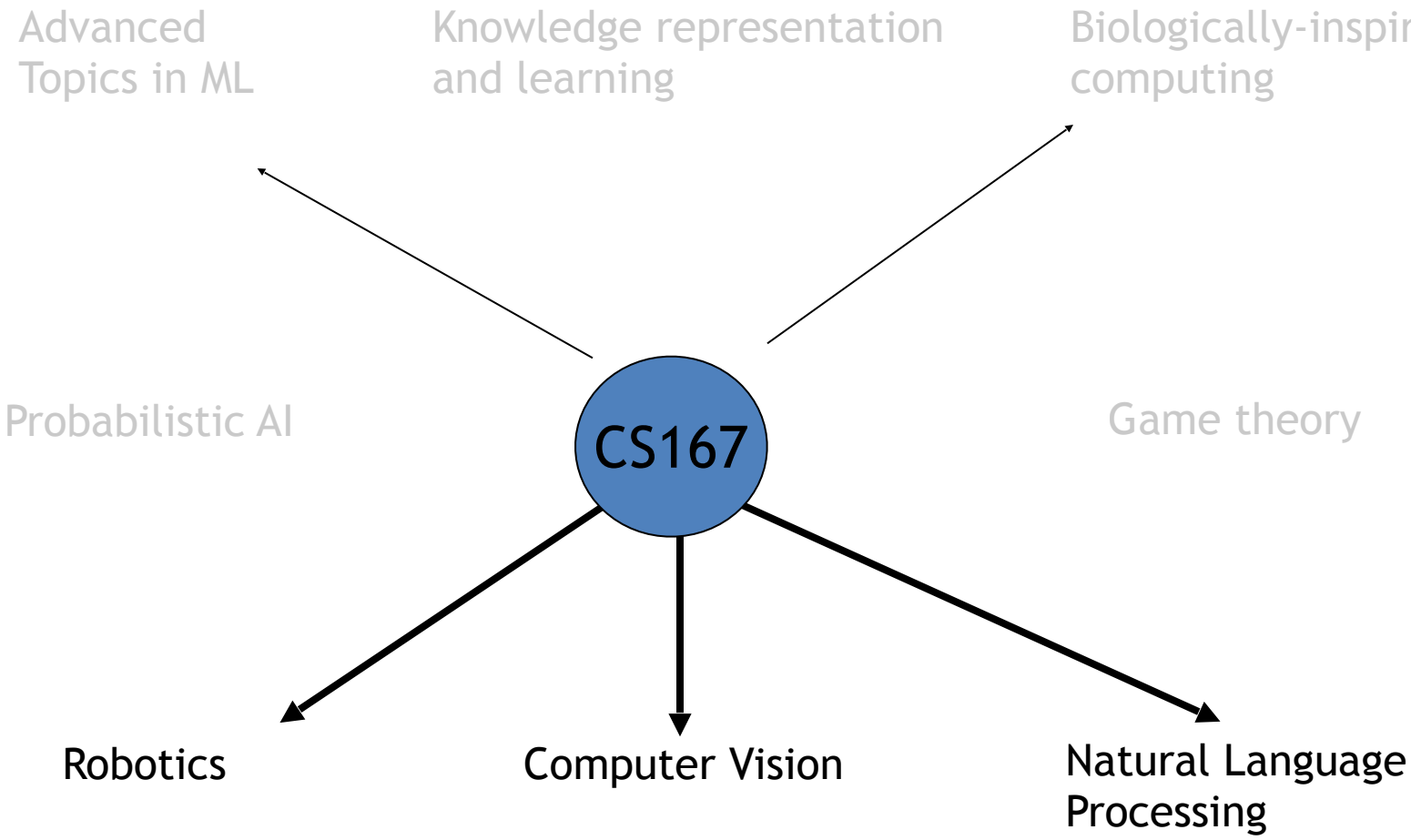
- Deep Machine Learning
  - Artificial neural networks
  - Convolutional Neural Network (CNN)
  - Recurrent Neural Network (RNN)
  - Long Short-Term Memory (LSTM)
  - Generative Adversarial Networks (GAN)
  - Transformers
- Applications
  - Computer vision, natural language processing, robotics, audio/speech analysis

Why take this class?

- Learn a lot about Machine Learning

- Apply a variety of modeling techniques to classification, regression, and unsupervised learning problems using data in different formats (such as typical structured data, text, and images).
- Create software that utilizes machine-learning programming libraries in order to conduct machine-learning-based data analysis.
- Develop and conduct machine-learning-based data analysis experiments, and they will be able to interpret and explain the results.
- Feel comfortable with using industry-standard tools such as Google Colab, GitHub, etc
- Understand fundamentals of machine learning
- Gain an understanding of the advantages and disadvantages of different learning paradigms so that students can choose appropriate solutions given a problem description
- Receive hands-on experience with commonly used algorithms and software tools within machine learning

Why take this class?



Careers in ML/AI

- 'Pure' ML/AI
  - Academia, industry labs
- Applied ML/AI
  - Almost any area of CS
  - NLP, vision, robotics
  - Economics
- Cognitive Science

LOCKHEED MARTIN

Willow Garage

iRobot

SRI International

Microsoft Research

evolution robotics

Google

Prerequisites

- CS 66: Introduction to Computer Science II
- Practically, this means:
  - Proficiency in a general-purpose programming language
  - Some level of mathematical maturity will be helpful, esp. with calculus, linear algebra, statistics
  - Willingness to learn some programming and/or math on your own if necessary

Blackboard Tour

Participation Cards

Participation Cards

- It is a goal of mine to make my classrooms as *inclusive and equitable as possible*. Towards this goal, we will be using **participation cards**.

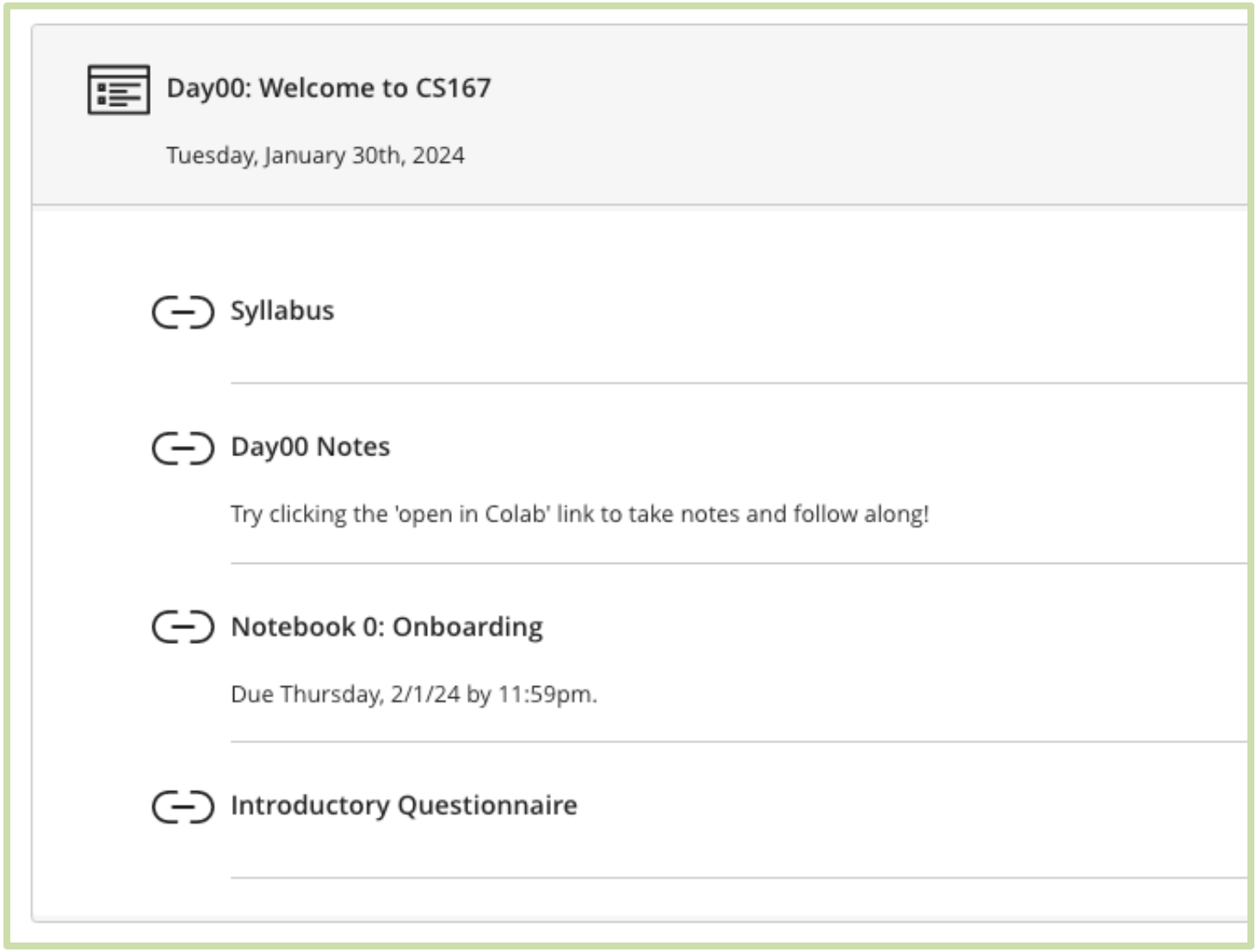
- How it's going to work:
  - Today, in class, we're going to make **participation cards**
  - At the start of class, I'll shuffle and draw a card. The person whose name is on that card becomes the **card bearer** for the day.
  - The **card bearer** is responsible for flipping cards and calling out the person whose name is on the card when I ask them to.

Participation Cards

- If your name is called out by the card bearer you have a few options:
  1. Answer the question
  2. Ask what the question was (you'll be surprised how often, even I, forget what the question was).
  3. Say *'Pass'*, or *'I'm not sure'*
  4. Ask for a group *'huddle'*, where you can talk to your neighbors for a solution.

- Participation cards are a solution to help the class participation be more equitable. They're not meant to induce anxiety, and I hope you'll find I use them in a way that is more fun than anxiety inducing.

Task: Posted on Blackboard



Introduction to Machine Learning

An Example: Classifying my dog

- Imagine we want to classify which image depicts a specific dog we want to identify
- Our training data might look something like this:

An Example: Classifying my dog

- Then, when we have some new pictures of my dogs, the idea is that we can make a **prediction** based on previous data as to whether it is **Zoey** or **Ace** in the photo.

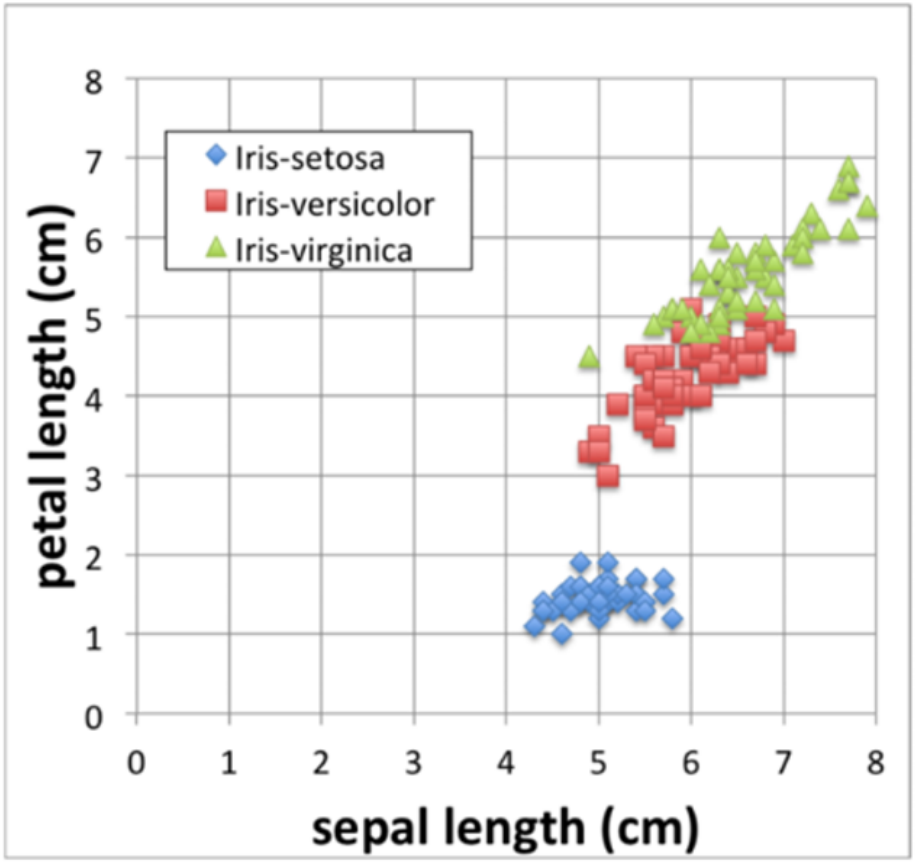
Another Example: What species of Iris is this?

- Use the poll link to answer this question

Imagine you found this beautiful flower while on a walk and took the following measurements:

5.1 cm petal length
7.2 cm sepal length

What species do you think it is?



Machine Learning Variations

- We are going to learn about a lot of different types of machine learning in CS167. Here are a few categories to look out for:

  - **classification:** identify which category it goes in, eg, *'Spam or ham?', 'Eric or Tim?', 'Fish, amphibian, reptile, bird, or mammal'*
  - **regression**: real-valued labels eg, price of Bitcoin, tomorrow's temperature, etc.
  - **supervised learning:** data has labels, goal is to predict the labels of new instance
  - **unsupervised learning:** data does not have a label, the goal is to analyze/ cluster the examples
  - **other issues:** missing data, sequential data, outlier anomaly detection, and many more

Group Exercise

- Group Exercise:
  - Take 2 minutes to talk to your neighbors about the following:
    - Come up with as many examples as you can of ways you interact with machine learning on a day-to-day basis.
    - Submit your answers to Google Form

ML interactions