

CS 475/675 Machine Learning: Homework 1

Due: Friday, September 20, 2024, 11:59 pm US/Eastern
100 Points Total Version 1.0

Make sure to read from start to finish before beginning the assignment.

1 Homeworks

Homeworks will typically contain two parts:

1. **Analytical:** These analytical questions will consider topics from the course. These will include mathematical derivations and analyses. Your answers will be entirely based on written work, i.e. no programming.
2. **Practicum:** In the practicum portion of the assignment, you will apply machine learning concepts to gain experience working with data from different domains. Practicums could involve Python notebooks, applied explorations of topics covered in the class, or programming assignments. Please note that the use of any form of AI assistance is strictly prohibited in this assignment.

[Click here for the Practicum Google Colab Notebook](#)

The point total for each portion of the homework will be listed in the assignment. Written assignments will be submitted as PDFs. See below for more details about what to submit.

1.1 Collaboration Policy

The course policy is that, *unless otherwise specified*, all work must be your own.

1.2 What to Submit

For this assignment you will submit the following.

1. **Analytical.** You will submit your analytical solutions to Gradescope. **Your writeup must be compiled from L^AT_EX and uploaded as a PDF.** The writeup should contain all of the answers to the analytical questions asked in the assignment. Make sure to include your name in the writeup PDF and to use the provided L^AT_EX template for your answers following the distributed template. You will submit this to the assignment called “Homework 1: Analytical”.
2. **Practicum Python Notebook.** You will submit the notebook to the assignment titled “Homework 1: Practicum”.

You will need to create an account on gradescope.com and signup for this class. The course is <https://www.gradescope.com/courses/835426>. Use entry code BK5K8K. **You must either use the email account associated with your JHED, or specify your JHED as your student ID.** See this video for instructions on how to upload a homework assignment: https://www.youtube.com/watch?v=KMPoby5g_nE.

1.3 Questions?

Remember to submit questions about the assignment to courselore: counselore.org/courses/9582956601/.

2 Analytical (50 points)

Please see the accompanying `2024_homework1_analytical.tex` file for the analytical questions for this assignment. There is space provided in that file for you to type your answers in \LaTeX after each question. **Do not edit the file in any way except to add your answers.** Gradescope assumes that the PDF will exactly match our template except for your solutions.

In addition to completing the analytical questions, your assignment for this homework is to learn \LaTeX . All homework writeups must be PDFs compiled from \LaTeX . Why learn \LaTeX ?

1. It is incredibly useful for writing mathematical expressions.
2. It makes references simple.
3. Many academic papers are written in \LaTeX .

The list goes on. Additionally, it makes your assignments much easier to read than if they are written by hand or if you complete them in Word.

We realize learning \LaTeX can be daunting. Fear not. There are many tutorials on the Web to help you learn. We recommend using `pdflatex`. It's available for nearly every operating system. As the semester progresses, you'll no doubt become more familiar with \LaTeX , and even begin to appreciate using it.

Be sure to check out this cool \LaTeX tool for finding symbols. It uses machine learning! <http://detexify.kirelabs.org/classify.html>

For each homework analytical we will provide you with a \LaTeX template. You **must use the template**. The template contains detailed directions about how to use it.

Please open the template to view the analytical questions.

3 Practicum (50 points)

In this assignment, we will explore the **ID3 algorithm** and build a decision tree for classification tasks. Decision trees are a widely used machine learning model due to their simplicity, interpretability, and ability to handle both categorical and continuous features. You will learn to build a decision tree by recursively splitting the dataset based on the feature that provides the highest information gain. Each internal node in the resulting tree represents a decision based on a specific feature, while each leaf node represents a class label or final decision outcome. By the end of this assignment, you will have a solid understanding of how decision trees work, how to implement the ID3 algorithm, and how tree depth influences a model's ability to generalize to new data.

What You Will Do

Open the Jupyter notebook `CS475_homework1_practicum.ipynb`. This notebook will walk you through:

- Implement the ID3 algorithm from scratch, this exercise helps you reinforce key concepts such as entropy and information gain, which are essential for building a decision tree.
- Understand the procedure of training and evaluating a machine learning model.
- Understand the effect of tree complexity on model performance, you will explore how varying the depth of the tree affects both training and testing accuracy.

There are questions that should be answered inline within the notebook. You will hand in the Python notebook, which contains your implementation and the answers to the questions. Please see the instructions on how to submit the notebook.

[Click here for the Practicum Google Colab Notebook](#)