

# CS 475/675 Machine Learning: Homework 1

Due: Tuesday, September 20, 2022, 11:59 pm

50 Points Total

Version 1.0

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

## 1 Introduction

The goal of the homework assignments in this course is to learn about how machine learning algorithms work through hands-on experience. This includes both implementation and analysis of various methods.

This semester we will have 5 homework assignments for a total of 200 points. The first and last homework will be worth 25 points, and the other homeworks will be worth 50 points. Homeworks will be worth 20% of your overall grade.

There are three different types of work that we will incorporate in the homework assignments.

1. **Analytical:** These questions will ask you to consider questions related to the topics covered by the assignment. You will be able to answer these questions without relying on programming. Many will require mathematical work.
2. **Lab:** In the lab you will utilize Python notebooks (Google colab) to explore data and algorithms, and answer questions. You can work on this as a standalone Python notebook (<https://jupyter.org>) or using Google Colab (<https://colab.research.google.com/>).
3. **Programming:** You will implement a machine learning algorithm in Python and evaluate it on a provided dataset. We will provide detailed instructions to guide your implementation, so behaviors will remain consistent across student submissions in the class.

Most assignments will focus on either analytical or programming work, in combination with a lab. For the first assignment, you will complete analytical work and a lab.

Each homework will have a master document (this document) that overviews all of the work in the assignment. For analytical work, there will be a separate L<sup>A</sup>T<sub>E</sub>X template that contains the actual questions. For labs, we will distribute a Python notebook for you to complete. For programming, we will (often) distribute starter code and data.

Each assignment will contain a version number at the top. While we try to ensure every homework is perfect when we release it, errors will happen. When we correct these, we'll update the version number, post a new PDF and announce the change. Each homework starts at version 1.0.

Let's get started!

## 2 Collaboration Policy

The course policy is that, unless otherwise specified, all work must be your own. See the course information Google document for more details.

**For this assignment, you may work in groups of 1 to 3 students.** You and your group will make a single submission to Gradescope. You will be able to indicate all students in your group when you submit. The entire group will receive the same grade, so please choose your group carefully.

You can only work in teams of one, two or three students; no more. Your group can include anyone from either section (01/02/03/04) or course (475 or 675) provided that all team members are enrolled in the class and taking it for credit (not audit). We *highly* recommend that you do every part of the assignment together instead of splitting it up. Our intention is to include questions on the exam that require an understanding of all parts of the homework. It is to your advantage to work together on all parts.

You can work on the same Overleaf document and think through the questions together. You probably want to work with the group for the semester (*only* for assignments where collaboration is allowed) but it is not a requirement.

Please see the course policy document for an explanation of how late hours apply in a group.

## 3 L<sup>A</sup>T<sub>E</sub>X

All solutions for the analytical problems must be PDFs compiled from a L<sup>A</sup>T<sub>E</sub>X template we will distribute for each assignment. Why learn L<sup>A</sup>T<sub>E</sub>X?

1. It is incredibly useful for writing mathematical expressions.
2. It makes references simple.
3. Many academic papers are written in L<sup>A</sup>T<sub>E</sub>X.

The list goes on. Additionally, it makes your assignments much easier to read than if you try to scan them in or complete them in Word.

We realize learning L<sup>A</sup>T<sub>E</sub>X can be daunting. Fear not. There are many tutorials on the Web to help you learn. We recommend using Overleaf.com. We have provided you with the tex source for this PDF. You **must use the template**. If you do not use this template, we will not grade your assignment. As the semester progresses, you'll no doubt become more familiar with L<sup>A</sup>T<sub>E</sub>X, and even begin to appreciate using it.

Be sure to check out this cool L<sup>A</sup>T<sub>E</sub>X tool for finding symbols. It uses machine learning! <http://detexify.kirelabs.org/classify.html>

## 4 Analytical (15 Points)

The analytical questions will be included in a separate template, where you can fill in your answers. Please open the file `homework1.template.tex` and respond to the analytical questions.

## 5 Lab (10 points)

In this assignment you will be exploring datasets for regression and classification. Open and complete all parts of the lab Python notebook.

You will hand in your completed Python notebook as a PDF.

## 6 What to Submit

For this assignment you will submit the following items to Gradescope.

1. **Analytical.** Your analytical solutions **must be compiled from L<sup>A</sup>T<sub>E</sub>X 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 and your partner's name in the writeup PDF and to use the provided L<sup>A</sup>T<sub>E</sub>X template for your answers following the distributed template. You will submit this to the assignment called "Homework 1: Analytical".
2. **Lab Python Notebook** You will submit your Python notebook as a PDF by going to File → Export via PDF or File → Export via PDF via LaTeX. Once you download the pdf, open the file to ensure that the plots show up. You will submit this to the assignment called "Homework 1: Lab".

You will need to create an account on [gradescope.com](https://www.gradescope.com) and signup for this class. The course is <https://www.gradescope.com/courses/427788>. Use entry code 8NKKGP. **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](https://www.youtube.com/watch?v=KMPoby5g_nE).

Seriously, this is important: **You must either use the email account associated with your JHED, or specify your JHED as your student ID.**

## 7 Questions?

Remember to submit questions about the assignment to the appropriate folder on Piazza.