

CS 475/675 Machine Learning: Homework 2

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 2: Analytical”.
2. **Practicum Python Notebook.** You will submit the notebook (*.ipynb) to the assignment titled “Homework 2: 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: courselore.org/courses/9582956601/.

2 Analytical (50 points)

Please see the accompanying `2024_homework2_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.

3 Practicum (50 points)

In this assignment, we will implement **Polynomial Regression**, and **Logistic Regression**, while exploring the impact of **Regularization** on these models. The notebook is structured as two parts:

Part 1: Exploring Regularization with Polynomial Regression

Polynomial Regression is a form of regression analysis in which the relationship between the independent variable X and the dependent variable y is modeled as an n th-degree polynomial. This allows for more complex relationships between the variables to be captured, but can also lead to overfitting. To mitigate this risk, we will utilize Ridge Regression, a regularization technique that helps reduce the impact of multicollinearity and is particularly effective when combined with polynomial regression.

Part 2: Implement Logistic Regression using Newton-Raphson Method

Logistic Regression is used for predicting the outcome of a categorical dependent variable based on one or more predictor variables. It estimates the probability of an event occurring (e.g., 1/0, yes/no) based on the input variables. We will use the Newton-Raphson Method, an iterative optimization algorithm, to find the maximum likelihood estimates of the parameters in the logistic regression model. This method starts with an initial guess and updates the parameters until convergence is achieved.

By the end of this assignment, you will have a solid understanding of these regression techniques and the impact of regularization.

What You Will Do

Open the Jupyter notebook `CS475_homework2_practicum.ipynb`. This notebook will walk you through:

- Implement the Ridge regression algorithm from scratch, this exercise helps you reinforce key concepts such as polynomial feature generation, coefficient calculation, and regularization.

- Implement Newton-Raphson Method for Logistic Regression.
- Understand the impact of model complexity and regularization on the regression models

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](#)