CS 475/675 Machine Learning: Homework 2
Introduction

Due: Friday, September 30, 2022, 11:59 pm

50 Points Total Version 1.0

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

1 Introduction

Assignments in this course can consist of three parts. The collaboration policy for this assignment (same as last homework).

1. **Analytical (15 points):** 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.
2. **Programming (20 points):** The goal of the programming assignments in this course is to learn about the hands-on experiences on logistic regression algorithms. In each homework assignment you will first implement an algorithm and then run it on the testing dataset.
3. **Lab (15 points):** Download the following iPython notebook: iPython notebook for HW-2, and open it in Google Colab as per the instructions. You will solve all of the programming problems using that iPython notebook.

2 Analytical (15 Points)

In addition to completing the analytical questions, your assignment for this homework is to learn LaTeX.

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

3 Programming (20 points)

We will first describe how programming assignments in general and the framework we will use to run your assignment. We will then describe the details of this assignment.

All files needed for this assignment are available online at:

- Colab: `https://colab.research.google.com/drive/1hp6wc6aLNkUT5Jy0RBlbOStgwjJ_woV-#scrollTo=pjPfIJ5G52It`
- Dataset: `https://drive.google.com/file/d/1SqG9PL0lnY7XXQRVbR44mZECh0fFAUKv/view?usp=sharing`

In the programming assignments you will implement machine learning algorithms, and test them on provided datasets.

3.1 Logistic Regression (10 points)

In this section, we are going to implement the logistic regression model with L2 regularization using gradient descent.

You will fill in the details. Search for comments that begin with `TODO`; these sections need to be written. You may change the internal code as you see fit but **do not change the names of any of the files or command-line arguments that have already been provided.**

3.2 Logistic Regression Adagrad (10 points)

Start by duplicating your implementation of logistic regression from the previous exercise into the class `LogisticRegressionAdagrad`. First, adjust your implementation to use stochastic gradient descent instead of regular gradient descent. The training should run multiple loops over the entire training set, selecting each instance in turn. The gradient is then computed for just that one instance, and the model parameters are updated accordingly based on that one gradient (not the average gradient).

Once you have stochastic gradient descent tested and working, do the exploration on the Adagrad (more instructions showed on the Colab). Adagrad adapts the learning rate to individual features, and changes it over time. Modify the gradient computation to use Adagrad instead of the constant learning rate alpha, allowing it to adapt the learning rate for each feature. Test your implementation thoroughly.

4 Lab (15 points)

- Download and open the Colab. All the requirements are shown on the Colab.
- You will need complete all `TODO` in the Colab and write a report for the lab. **Generally, in the lab report, it should include all plotting results and the interpretation results.**

5 What to Submit

In each assignment you may submit three different things.

1. **Programming: Submit your .ipynb to Gradescope**. **Your .ipynb must complete all the TODO sections and with all output results**. You will submit this to the assignment called "Homework 2: Supervised Classifiers 2: Programming".
2. **Analytical: Submit your writeup to Gradescope**. **Your writeup must be compiled from LaTeX 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 LaTeX template for your answers following the distributed template. You will submit this to the assignment called "Homework 2: Supervised Classifiers 2: Analytical".

3. **Lab: Submit your PDF to Gradescope**. **Your PDF must complete all the TODO sections and with all output results**. You will submit this to the assignment called "Homework 2: Supervised Classifiers 2: Lab".

You will need to create an account on 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`.

6 Questions?

Remember to submit questions about the assignment to the appropriate group on Piazza: `https://piazza.com/class/l7542wgbgfu7a8`.