Assignment 1

AI503 & CSE613: AI Toolkits
Spring Semester, 2024, UNIST

March 25, 2024

1 Introduction

In this assignment, your task is to implement models for linear regression and logistic regression. To accomplish this, follow the given instructions and complete the designated "TODO" sections in the provided notebook. It is important to include comprehensive comments to explain the reasoning behind your implementation strategy.

Key Requirement: Use only NumPy to develop your solutions. Do not use built-in functions from external libraries such as PyTorch or scikit-learn, as this will result in a score of zero. Ensure that the submitted code executes without any errors. Failure to follow these guidelines will result in a score of zero. Each task will be evaluated on a pass/fail basis, and partial credit will not given.

- **Release Date**: 26, March, 2024
- **Due Date**: 12, April, 2024
- **Given Files**
  - **Assignment1.ipynb**: this notebook includes "TODO" sections that requires completion.
  - **height_weight_genders.csv**: an input file.
  - **helpers.py**: this file contains some useful functions such as `load_data()` and `sample_data()`.
  - **plots.py**: this file contains a `visualization()` function.
  - **test_utils.py**: this file contains some testing modules for sanity check.
- **TAs**: Heelim Hong (heelim@unist.ac.kr) and Sungho Jeon (sungho@unist.ac.kr)
  - Be sure to cc Prof. Yeon-Chang Lee (yeonchang@unist.ac.kr) when you email a question to the two TAs.

2 Task 1. Linear Regression

2.1 Least Squares

Your task is to complete the notebook function `least_squares(y, tx)`, which solves the normal equations for linear mean squared error (MSE). This function should return the optimal weights and the mean-squared error. Do not the `numpy.linalg.inv` function to solve a linear system $Ax = b$ in this section.

2.2 Classification Using Linear Regression

We will use linear regression for classification. To this end, we will use the given `height-weight data` to predict the binary-valued gender (See 'height_weight_genders.csv' file). For better visualization, we will randomly sample 200 data points from the dataset.

- Utilize the `least_squares(y, tx)` function implemented in **Subtask 1-1** to compute the weights w on the `height-weight data`.
- Visualize the data points and the decision boundary using `visualization()` as shown in Figure 1.

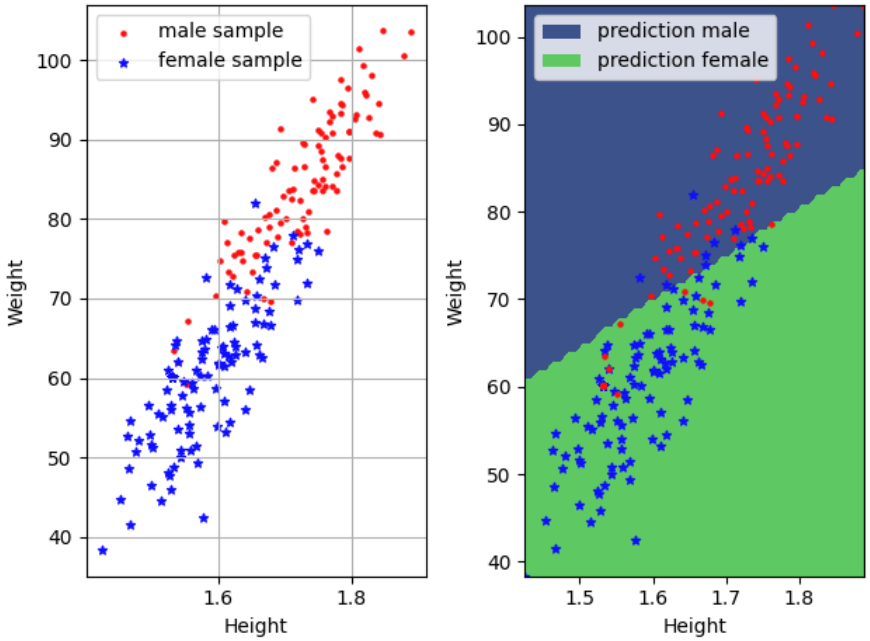

Figure 1: Classification by least square.

3 Task 2. Logistic Regression

Your task is to first complete the notebook functions `sigmoid()`, `calculate_loss()`, and `calculate_gradient()`.

- The `calculate_loss()` function should return the *negative log-likelihood loss (NLL)*
- The `calculate_gradient()` function should return the *gradient* of this loss with respect to the parameters w.

Then, complete the notebook function `learning_by_gradient_descent()`.

- You should calculate the loss and its gradient with respect to w.
- Now, you can update the weights w, and the function should return the loss and these updated weights w.

Finally, plot the predictions to get a visualization similar to the right part of Figure 1. Check if you get similar or different results.

4 How to submit your assignment

- Create A1-[your student id].zip (e.g., A1-20241111.zip), which should contain the following files:
  - Assignment1.ipynb: this should contain your implementation. Do not modify the code outside of the "TODO" regions.
  - height_weight_genders.csv, helpers.py, plots.py, and test_utils.py
- All files should be included in a single folder.
- Make sure that no other files are included in the zip file.
- Submit the zip file at Blackboard. The submission section will be set up soon.

Note that submitting work after the deadline is not acceptable.