CS 6364–002 Homework 2

August 30, 2022

Requirements:

- Deadline for submission: **Sep-12-2022**.

- All assignments **MUST** have your name, student ID, course name/number at the beginning of your documents.

- For each of the four questions, please write all the codes in one Jupyter notebook and run the codes to display the results in the notebook before you save the notebook (ipynb file). Please use markdown cells (`https://www.tutorialspoint.com/jupyter/jupyter_notebook_markdown_cells.htm`) to write text to explain or discuss your codes/results.

- About how to create and save a jupyter notebook with anaconda navigator, check the following youtube video: `https://www.youtube.com/watch?v=-MyjGO0la2k`

- As there are four questions in total, you will need to submit four Jupyter notebooks. Please zip the four notebooks into one zipped file and submit the zipped file.

This homework assignment has four questions stated as follows:

Q1 (Linear Regression): Write python codes in a Jupyter notebook that use the python library (sklearn.linear_model) to train a linear regression model for the Boston housing dataset:

`https://towardsdatascience.com/linear-regression-on-boston-housing-dataset-f409b7e4a155`.

Split the dataset to a training set (70% samples) and a testing set (30% samples). Print the root mean squared errors (RMSE) on the training and testing sets in the Jupyter notebook.

Q2 (Linear Regression) Write python codes in a Jupyter notebook that implement the gradient descent algorithm to train a linear regression model for the Boston housing data set

`https://towardsdatascience.com/linear-regression-on-boston-housing-dataset-f409b7e4a155`

Split the dataset to a training set (70% samples) and a testing set (30% samples). Print the root mean squared errors (RMSE) on the training and testing sets in the Jupyter notebook.

Q3 (Logistic Regression): Write python codes in a Jupyter notebook that use the python library (sklearn.linear_model) to train a logistic regression model for the Titanic dataset:

`https://blog.goodaudience.com/machine-learning-using-logistic-regression-in-python-with-code-ab3c7f5f3bed`.

Split the dataset to a training set (80% samples) and a testing set (20% samples). Print the overall classification accuracies on the training and testing sets and report the precision, recall, and F-measure scores for each of the two classes on the training and testing sets in the Jupyter notebook.

Q4 (Logistic Regression): Write python codes in a Jupyter notebook that implement the gradient descent algorithm to train a logistic regression model for the Titantic dataset:

`https://blog.goodaudience.com/machine-learning-using-logistic-regression-in-python-with-code-ab3c7f5f3bed`.

Split the dataset to a training set (80% samples) and a testing set (20% samples). Print the overall classification accuracies on the training and testing sets and print the precision, recall, and F-measure scores for each of the two classes on the training and testing sets in the Jupyter notebook.