

Assignment 3: Linear Models and Neural Networks
Due Date: April 19th 2020 @ 11:59 PM

Instructions:

1. This assignment can be done in groups of up to three students
2. You should submit your assignment to the email addresses:
 - inzamam.rahaman@sta.uwi.edu
 - shivramx@gmail.com

You should use the subject: “COMP3608 Assignment 3”, and you should submit a single zip folder named using the student ids of all members separated by commas. This **zip** folder should contain a README.md or readme.txt file with the names and id numbers of all group members.

3. You should separate your solutions to each component into separate sub-directories named P1 and P2

Part 1 - Price Prediction

In this part of the assignment, you are tasked with evaluating the performance of different variants of linear regression for price prediction on the diamonds data-set from the last assignment. You will be given two Numpy array files:

- *X.npy* - each row containing the encoded features for a diamond (i.e. a data-point). These are your input features
- *y.npy* - each row containing the price of a diamond. These are your targets.

The rows of *X.npy* and *y.npy* correspond to one another, i.e. row 1 of *X.npy* contains the features of the first data-point and row 1 of *y.npy* contains the price data of the first data-point.

Evaluate and compare the performance of standard linear regression, LASSO Regression, and Ridge Regression on predicting y from the features in X . You should use k -fold cross-validation where $k = 5$. Since prices should not be non-positive, you should use MAPE as your metric for the model's performance.

Part 2 - Breast Cancer Classification

The UCI Machine Learning repository contains a data-set for Breast Cancer detection. This data-set can be found at [https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)).

Using the input features specified on the data-set description, you are to do the following:

- Construct a Neural Network in PyTorch, Keras, or TensorFlow with a single hidden layer. This neural network should be designed in an attempt to classify a tumor as malignant or benign from the given input data. Recall that this is a binary classification problem.
- Using k -fold cross-validation where $k = 5$, evaluate the performance of your built model with Logistic regression. Decide which model is better for this problem.
- After deciding which model is better, write a short report conjecturing why this was the case.