Diabetes Dataset - https://drive.google.com/drive/folders/185pTXHW-ze1SwwCRpSFtEXo08wcfXoLf?usp=sharing

A. Linear Regression -

Diabetes dataset is one of the datasets available in sklearn. The diabetes dataset consists of 10 physiological variables (age, sex, weight, blood pressure etc) measure on 442 patients, and an indication of disease progression after one year.

You are given a Training dataset csv file with X train and Y train data. As studied in lecture, your task is to come up with Linear Regression training algorithm and thus predictions for the test dataset given.

Use Linear Regression (in sklearn) as a training algorithm and submit results predicted by that. Print feature importance and different error metrics used.

Submit a csv file with only predictions for X test data. File should not have any headers and should only have one column i.e. predictions. Also prediction values in file should be upto 5 decimal places.

B. Classification - Convert above Linear Regression problem to Binary Classification Problem by considering patients to be diabetic if Y>=138, else non-diabetic.

Report performance metrics (Accuracy/Precision/Recall/F1) upon applying Multilayer Neural Networks. Vary the number of layers between [2, 4] and plot the accuracy with the number of layers in x axis. Each layer should have fewer nodes than the previous layer and the total number of neural nodes should be the same while increasing the number of layers.

Submit the csv with metric values and the required plot.

Note: While calculating different performance metrics, consider diabetic to be a positive class.