Lab - 4
CSL2010: Introduction To Machine Learning
AY 2022-23
Linear Regression

General Instructions

1. You need to upload a zip *<Lab4_Your_Roll_No>.zip*, which contains one file for the task in *<Lab4_Your_Roll_No>.py* format and the report for the entire assignment in *<Lab4_Your_Roll_No>.pdf* format.
2. Provide your colab file link in the report. **Make sure that your file is accessible**.
3. Submit a single report mentioning your observations for all the tasks.
4. Report/Cite any resources you have used while attempting the assignment.

Dataset
Download the dataset from the link below.
https://drive.google.com/file/d/1s2IhEwbbSAGEtVuPpLwLq_wQ8P3Svf0l/view? usp=sharing

Q1) Linear Regression [30 marks]

a) Take "Age" as the target variable and perform a Linear Regression fit on the above dataset using inbuilt sklearn library. Perform the train-test split in a ratio of your choice. Also perform the necessary data cleaning and preprocessing steps. [15]

b) Implement the Linear Regression model from scratch, using Numpy only for calculating matrix inverse. Use RMS error as the evaluation metric. Compare with the results obtained in (a). [15]

Q2) Analysis [30 marks]: Perform the following analyses on both (a) and (b) in Q1.

a) Implement any two feature normalization techniques and analyze the difference in the RMS error. [10]

b) Try three different ratios of splitting when training the linear regression model and analyze the difference in the RMS error. [10]

c) As done in Q1(a), learn a model (the 'w' vector) using any one feature and the target variable, and create two separate 2-D scatter plots of the points in the training and testing sets, and also show the learned 'w' in each. Perform the same for any one other attribute and compare with the previously obtained plots. [10]