DEPARTMENT OF INFORMATION TECHNOLOGY, NITK SURATHKAL

LAB ASSIGNMENT-1

IT464: FOUNDATIONS OF MACHINE LEARNING

Write a Python program to answer the following

1. **Find the line of best fit (linear regression)** for the following data set and plot it.
   https://www.kaggle.com/datasets/tanuprabhu/linear-regression-dataset

   Find the predictions to x=[20,40,60,...,280,300] and **Compute the least squared error (LSE)**. Print/tabulate " x, predictions and LSE".

2. **Find the line of best fit (multiple linear regression - MLR)** for the California Housing Dataset and plot it.
   https://www.geeksforgeeks.org/dataset-for-linear-regression/
   Note: Exclude "longitude, latitude and ocean proximity" parameters/variables.

   Compute the price for the **"housing2" test data** using the MLR prediction.

3. **Perform MLR** and **Logistic regression** on the following data to predict heart disease.
   https://www.kaggle.com/datasets/dileep070/heart-disease-prediction-using-logistic-regression

   Predict heart disease for the **"heart2" test data**.

4. **Compute PCA components** for the heart disease data. **Predict heart disease with the PCA features** (consider **#PCA features = [1,2,3,4,5]**) and evaluate the performance in terms of confusion matrix. Note down your observations.

5. Load Cameraman image from python libraries and **reduce the dimensionality using SVD**, Check its visual appearances (**original versus new image**) for the different numbers of SVD components. Additionally, find the correlation between the original and reconstructed images from the **different numbers of SVD components (say 1,2,...,9,10)**.

6. Load Cocktail Party Problem dataset from kaggle to **perform ICA on separating the audios of different speakers**. **Test PCA and compare its performance with ICA's** in source separation problem.
   https://www.kaggle.com/datasets/anashamoutni/cocktail-party-problem-cities-of-the-us