MSE 1065 Lab 9 – Dimensionality reduction and effects on underlying predictions

In this lab, we will perform dimensionality reduction through Principal Component Analysis (PCA) on the same dataset as Lab 8. Your submission will be a single jupyter notebook containing all your code and analysis.

Part 1: Load the dataset and perform normalization by computing mean and standard deviation for each feature. Create train and test split with test sample size of 20%.

Part 2 (10 points): Fit a PCA model on training set. Use only 2 principal components by setting 'n_components' equal to 2. On your PCA model, fit a ridge regression model for alpha equals 0.1. Compute Mean absolute error (MAE), r-squared value and root mean squared error, corresponding to both training and test set. Comment on the error variation as compared to the ridge regression implemented in Lab -8.

Part 3 (5 points): Fit a PCA model on training set. Use only 4 principal components by setting 'n_components' equal to 4. On your PCA model, fit a ridge regression model for alpha equals 0.1. Compute Mean absolute error (MAE), r-squared value and root mean squared error, corresponding to both training and test set. Comment on the error variation as compared to the ridge regression implemented in Lab -8.

Part 4 (5 points): Fit a PCA model on training set. Use only 8 principal components by setting 'n_components' equal to 8. On your PCA model, fit a ridge regression model for alpha equals 0.1. Compute Mean absolute error (MAE), r-squared value and root mean squared error, corresponding to both training and test set. Comment on the error variation as compared to the ridge regression implemented in Lab -8.

Part 5 (5 points): On your best performing PCA model in Part 2-4, fit a neural network and compute Mean absolute error (MAE), r-squared value and root mean squared error, corresponding to both training and test set. Let's say 4 principal components give you lowest test set MAE, then fit a neural network on 4 principal components as computed in Part 3. The architecture of neural network is two layers with 6 hidden units. Comment on the error variation as compared to the neural network implemented in Lab -8.

Part 6 (5 points): On your best performing PCA model in Part 2-4, fit your favourite model and compute Mean absolute error (MAE), r-squared value and root mean squared error, corresponding to both training and test set. Comment on the error variation as compared to the models implemented in Lab -8.