**Pattern Recognition and Anomaly Detection**

**Assignment – II**

**1) Linear Regression**

Download the Abalone dataset. It contains 9 variables in which the last column is the output variable and the other 8 are input variables.

I. Perform Linear regression (implemented from scratch) on this dataset, use 8:2 train:test split, set random seed to 0. Report the RMSE on training and testing data.

II. Apply regularization techniques Ridge Regression and Lasso Regression (Sk-learn's implementation can be used) here. You have a hyperparameter alpha that you can modify to regulate how much the coefficients are restricted.

(a) Plot a graph to show the effect of alpha value (at least 10 different alphas must be used) on the testing data's RMSE and report best model's coefficients/parameters for both Ridge and Lasso Regression respectively.

(b) Use Sk-learn's Grid search function to find the best alpha value and report the best model coefficient for both Ridge and Lasso Regression respectively? Compare the best model coefficients with reported in Q1.II.a.

**2) Logistic Regression**

Download the Diabetes Dataset, Readme for this data. This dataset contains 9 columns, use the ``Outcome” column as the target value and the other columns as the features.

You need to implement gradient descent (both SGD and BGD) from scratch (You may use NumPy, but other libraries like Sk-learn, Keras are not allowed)

I. Perform Logistic Regression (implemented from scratch) on this dataset, use a 7:2:1 train:val:test split. Show all your pre-processing steps and mention them in your report.

(a) Include loss plots between training loss v/s iterations and validation loss v/s iterations. Comment on the convergence of the model. Compare and give your analysis between the plots.

(b) Re-Run your implementation for different learning rates 0.01, 0.0001, 10. Compare and give your analysis.

(c) Make the confusion matrix and report the accuracy, precision, recall and f1 score obtained.

II. Choose an appropriate learning rate, number of epochs from part Q2. Part I. Using

the same learning rate, run Sk-learn's Logistic Regression on the dataset above and

compare the following:

(a) Loss plots of Sk-learn and your implementation.

(b) Number epochs to converge of Sk-learn and your implementation.

(c) Compare performance of Sk-learn's implementation and your implementation. Report accuracy, precision, recall and f1 score.