Assignment-2
Statistical Machine Learning

Q1. Implement Mahalanobis distances and LOF (local outlier factor) methods from scratch and find the outliers through otsu thresholding (implementation from scratch) on the following dataset: [Dataset for Q1]
[6 Marks]

Q2. Implement Logistic Regression from scratch and perform binary classification using the following dataset: [Dataset for Q2]. Try both PCA+FDA and just FDA pre-processing steps before the classification and compare the results. Implement FDA yourself.
NOTE: PCA+FDA means PCA followed by FDA.
[5 Marks]

Q3. Study the Normal Equation approach to Linear Regression yourself, implement it and develop a regression model using real estate price prediction dataset. Evaluate the model using the evaluation metrics discussed in the class. Please provide the derivation of the normal equation in your report in your own words to demonstrate your understanding of the approach. What are the limitations of this approach? **[4 Marks]**

Q4. Implement LDA from scratch and check if using it as a pre-processing step can improve kNN's accuracy on IRIS dataset. Take k=5. **[3 Marks]**

Q5. How can we apply Logistic Regression in a multi-class classification problem? Extend your Logistic Regression's implementation in Q2 to such a scenario and apply it on the IRIS dataset. Does it provide better results than the ones reported in Q4?
[2 Marks]