**Lovemore Tenha**

**Project3 Report**

In this report, we explore the performance of various classifiers on 2 different datasets; the digits dataset and the EEG\_Eye\_State dataset from UCI, which has time series instances. We use the classifiers from the scikit-learn library in python. The classifiers are Perceptron, Linear and Non-Linear Support Vector Machine, Logistic Regression, K Nearest Neighbor and Decision Tree.

The following Results were obtained after applying the classifiers on the digits dataset.

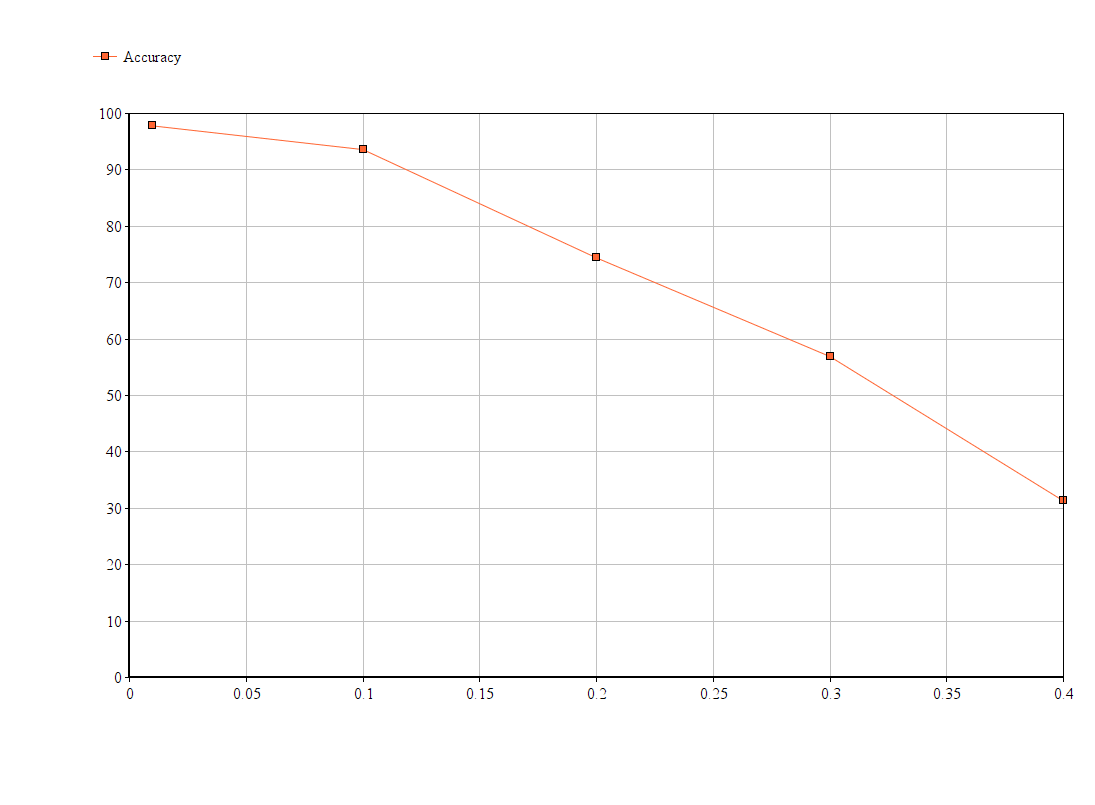
|  |  |  |
| --- | --- | --- |
| **Classifier** | **Accuracy (%)** | **Runtime(seconds)** |
| Perceptron | 92.5 | 0.03 |
| Linear SVM | 95.3 | 0.39 |
| Non-Linear SVM | 92.8 | 0.54 |
| KNNeighbor | 97.5 | 0.01 |
| Logistic Regression | 96.7 | 0.46 |
| Decision Tree | 83.3 | 0.02 |

**Accuracy and Runtime for each classifier on digits dataset**

K Nearest Neighbor and Logistic regression had the highest accuracies. However, K Nearest Neighbor achieved such high accuracy with the fastest runtime and K = 5. The perceptron and both SVM classifiers also had high accuracies. The perceptron, however, had the faster runtime compared to the runtimes of the SVMs. The decision tree classifier had the lowest accuracy and its maximum depth was set at 20.

**Effect of changing the value of the gamma parameter for the Non-Linear SVM**

The plot below shows the effect on the accuracy of the Non-Linear SVM of changing the value of the gamma parameter. The initial value of gamma was set at 0.01. The results show that increasing the value of gamma decreases the accuracy of the classifier.

****

For the K Nearest Neighbor classifier, altering the value of parameter k does not change the accuracy by much. However, increasing the value of K slightly increases the runtime. Increasing the maximum depth for the decision tree classifiers increases the accuracy of the classifier. As the depth becomes larger, the rate of increase of accuracy becomes small.

**The following results were obtained after applying the classifiers on the EEG\_Eye\_State dataset.**

|  |  |  |
| --- | --- | --- |
| **Classifier** | **Accuracy (%)** | **Runtime(seconds)** |
| Perceptron | 62.5 | 0.001 |
| Linear SVM | 57.5 | 0.06 |
| Non-Linear SVM | 82.5 | 0.002 |
| KNNeighbor | 67.5 | 0.001 |
| Logistic Regression | 67.5 | 0.02 |
| Decision Tree | 65.0 | 0.002 |

The maximum iterations was set at 1000. The results show that Non-Linear SVM and Logistic regression had the highest accuracy. The Linear SVM and perceptron had relatively lower accuracies, which is quite different from the results for the digits dataset. The EEG Eye State dataset has 2 classes and it appears the data is not linearly separable. That explains the better performance for the non-linear classifiers.

The accuracy for the Non-Linear SVM was highest when gamma was set at 0.5. Decreasing or increasing the value of gamma resulted in lower accuracies.

From the analysis of the classifiers on the 2 datasets, it is evident that there is no universally better classifier. In the EEG Eye dataset, nonlinear classifiers performed better than linear classifiers because the data was not linearly separable (only 2 class). On the other hand, linear classifiers perform relatively good on the digits dataset, which had 10 classifiers.