EEE 419/591 Project 2 Report

Taman Truong

October 20th, 2023

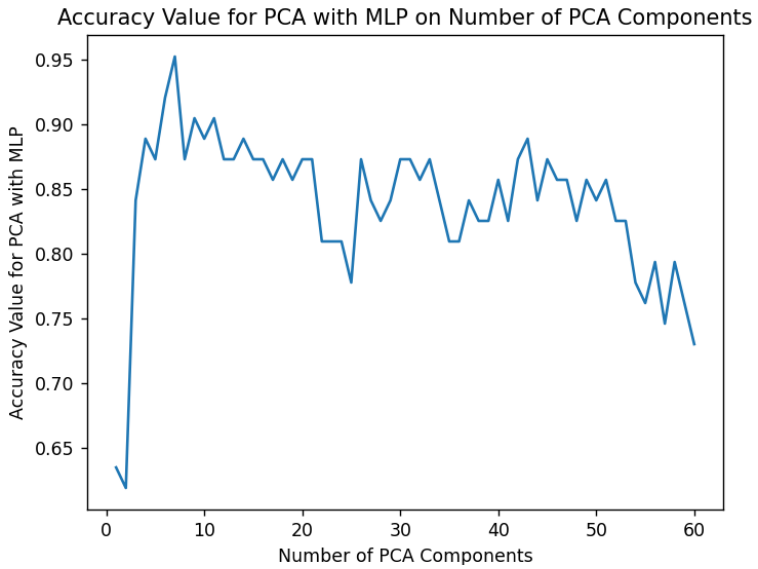
Introduction

For operating the SeaView Naval submarine safely across an old mine field that has never been cleared, a principal component analysis with a multi-level Perceptron machine learning model was done on data of classifying mines and rocks tracked over 60 time samples. We begin by defining machine learning terms and algorithms to aid the understanding of the reader in our analysis.

Principal Component Analysis (PCA) is a statistical procedure that utilizes and orthogonal transformation that maps correlated elements to uncorrelated elements. **Multi-layer Perceptron (MLP)** is a generalized Perceptron model consisting of multiple hidden neural network layers.

PCA Analysis of Sonar Mine/Rock Data

For the PCA, all components were tested up to the maximum number of time samples present in the dataset (60). To set up the training and testing datasets for the MLP classifier machine learning algorithm, the database was split so that 70% of the data served as the training dataset that the model would use to train for patterns and 30% of the data served as the testing dataset to test whether or not the model can detect patterns seem in the training data to the testing data to a sufficient accuracy. Each model had optimized parameters that generated the accuracies in predicting whether the objects were either mines or rocks. The graph of the accuracy versus the number of PCA components used and the output for the maximum accuracy for PCA with MLP, the number of PCA components requirements to achieve maximum accuracy, and the confusion matrix from this analysis is shown below.




```
Maximum Accuracy for PCA with MLP, Test Data: 0.95
Number of PCA Components to Achieve Maximum Accuracy: 7
[[28  1]
 [ 2 32]]
```

We see that from the graph and from the output, the maximum accuracy achieved with PCA and MLP is 95% with 7 PCA components. The parameters for the MLP classifier was chosen so that the runtime for the algorithm was fast enough to receive rapid solutions while maintaining a high accuracy and performance for this model. The confusion matrix tells us how well the machine learning model classified whether or not an object with 60 time samples as its features is either a mine or a rock correctly or not correctly. From the confusion matrix above, the model predicted that the object was a mine when it was actually a mine (true positive) 28/63 times, the model predicted that the object was a mine when it was actually a rock (false positive) 2/63 times, the model predicted that the object was a rock when it was actually a rock (true negative) 32/63 times, and the model predicted that the object was a mine when it was actually a rock (false negative) 1/63 times. The reason why 7 PCA components (a small number of components) achieved the maximum accuracy for this machine learning model is because PCA trades performance and accuracy in the model for simplicity in the model; adding more PCA components, adds more features, either relevant or irrelevant, to the analysis, inherently adding more noise to the model and effectively decreasing the performance and accuracy of the model as a result. From this analysis, we can conclude that within 7 time intervals, one can determine whether or not an object is a mine or a rock and can avoid danger whenever possible within that time frame.