Nicolas Carpenter
CMPUT 466: Machine Learning

Mini Project

Question 1.

The accuracies of the baseline (number of the most frequent category, over the number of total samples) and of each model of the models on the dataset (with 1000 epochs and hyperparameter tuning):

Baseline = 0.45
Logistic Regression (Logistic) = 0.85
Polynomial Support Vector Machine (PSVM) = 0.9
Neural Network (NN) = 0.95

Figure 1 shows the decision boundaries of each classifier (From left to right: Logistic, PSVM, NN)
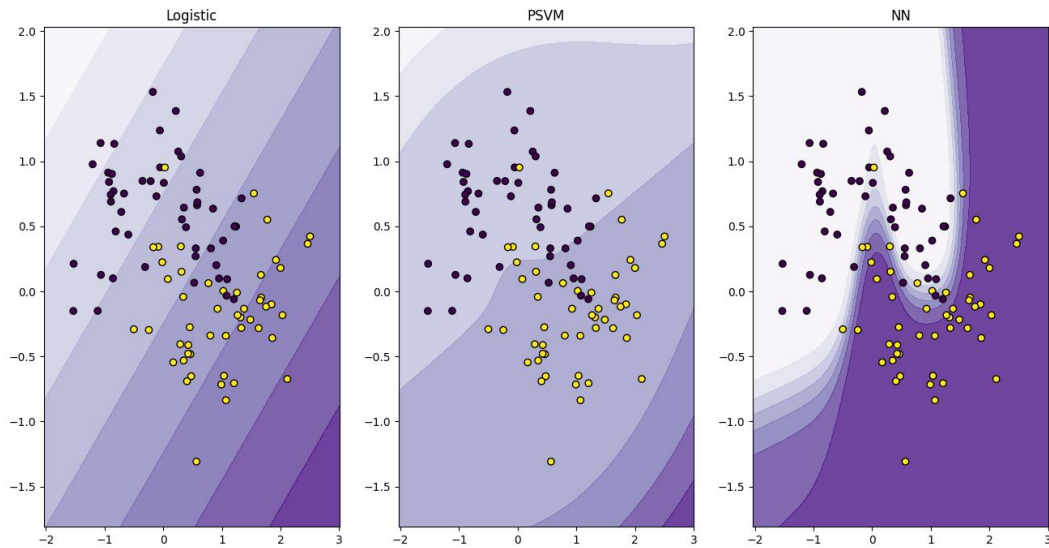

FIGURE 1. Decision Boundaries: varied classifiers

As we can see from the data for this set of data that cannot be exactly linearly separated, while all models perform better than the baseline, the models that can create a nonlinear decision line perform better, with the neural network having the greatest performance of the three.