1 Results

The results will be guided mainly by the drugs script in that it will mainly look at the FISHER and LOG_REGRESS flags (mentioned in the Method section of this report)

Figure 1 shows a histogram that illustrates the separation of the drug dataset with the raw inputs. Figure 2 shows a simple line plot of the data for each class.

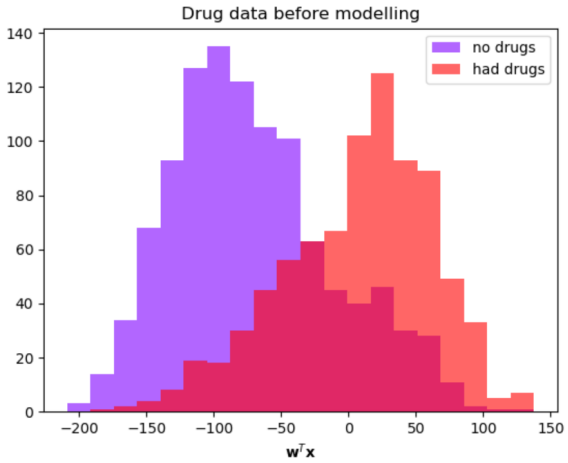

Figure 1: Drug Dataset: Histogram of the classes before modelling

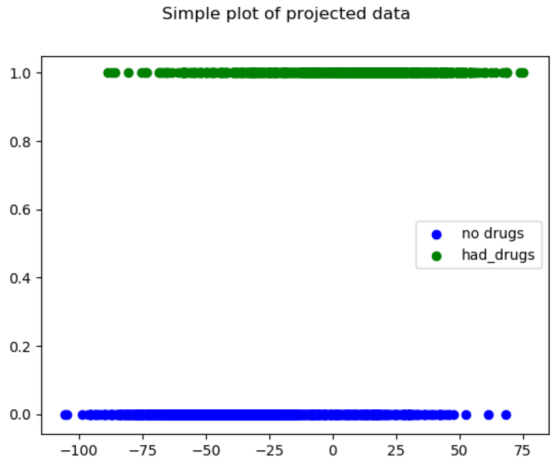

Figure 2: Fishers (Drug Dataset) simple line plot to show class separation

Figure 3 shows compares the prediction ability of a model when the raw inputs are used to generate weights from both the Fisher and maximum separation methods. A ROC curve is used to demonstrate this. These results can be obtained from setting the EXPLORATORY flags to true.

When set the FISHER flag to true, it ran fishers discriminant analysis on raw data, using cross validation and then testing on held out data. This resulted in 73% accuracy. The next analysis involved applying radial basis functions to all the inputs and then only on Gaussian looking predictors. This resulted in a 62% and 48% respectively prediction accuracy on unseen data. The functions that calculate these accuracies also check that the cross validation results comply with a specified tolerance of 10%. The output to console messages indicate that when all the data is modified with RBF, a stable model results. However, when applied to Gaussian looking predictors, the model is not stable. The ROC curves in figure 6 compares these results.

The drug classification script next tests the outcome of moving the threshold discriminant. With the default discriminant line set to zero, the line is shifted according to the minimum and maximum projected inputs. In hindsight, the authors now realise that this is perhaps not the correct calculation for the threshold and

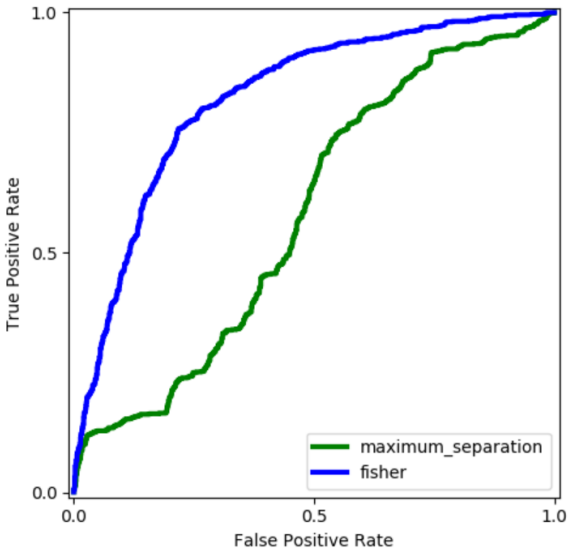

Figure 3: Fishers and Maximum Separation (Drug Dataset): ROC with all inputs

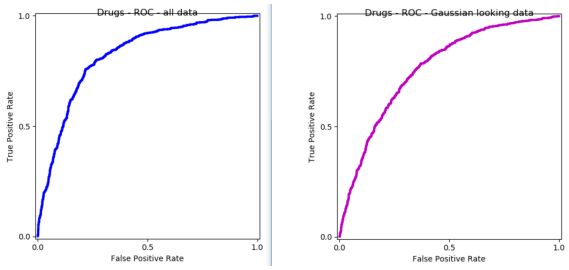

Figure 4: Drugs dataset: Fishers - All data vs Gaussian only looking data (pre modelling)

should have instead used the weights bias in an attempt to test the prediction accuracy. With this said, the following results will be wrong. Using the method that was used, the minimum threshold value was -105 and the maximum was 74. The best threshold was very close to 0 giving 78% prediction accuracy. A multiple ROC plot was created based on this method's results and this can be seen in figure 7.

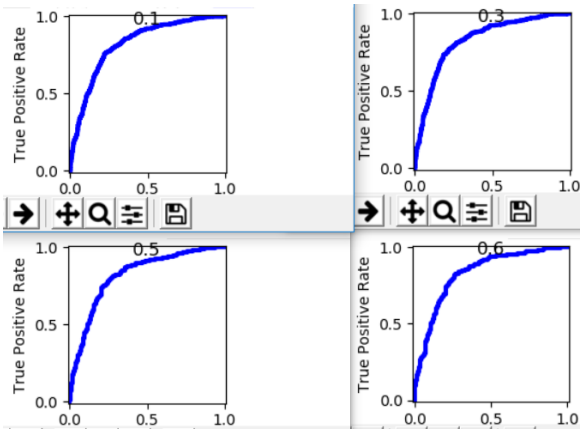

Figure 5: Drug Dataset: Fishers - ROC from different test fractions

The final part of the FISHER section for the drug script makes a call to a function that uses different portions of input for train and test data. Using splits of train/test 90/10, 70/30, 50/50, 40/60 gives predictions for the Fisher discriminant analysis of 74-77%.