Assignment 2.2 SVM and Decision Trees

Troy Dean - 0222566

Wednesday 19th September, 2018

0.1. Introduction

For this assignment we were required to run an experiment on three supplied datasets to compare four different simple linear classifiers. The data sets used were the pima diabetes, ionosphere and sonar datasets from datasets. The experiment compared Fischer's Linear Discriminant Analysis (FLDA), Simple Logistic, Multilayer Perceptron and LibSVM functions using the WEKA Experimenter.

0.2. Analysis of Experimenter Results

Running the experimenter accross the supplied datasets using FLDA, Simple Logistic, Multilayer Perceptron and LibSVM functions produced the output seen in the appendix A.

0.3. Analysis of Visual Boundaries

Figure 0.1.: Simple Logistic Boundary Visualization - X: plas, Y: mass

Figure 0.2.: Multilayer Perceptron Boundary Visualization - X: plas, Y: mass

Figure 0.3.: LibSVM Boundary Visualization - X: plas, Y: mass

Running the J48 classifier on the diabetes dataset and looking at the resulting tree showed that the two most informative features (according to J48) are plas and mass. Figure 1 shows the results of using these to do a boundary visualisation on the diabetes dataset. We can see that the boundary is a slightly arched line sloping from the top down to the X axes, splitting the graph roughly in half along the diagonal. There is quite a wide band of overlap into the green (tested positive) region.

0.4. Choosing a Classifier

The experiment and boundary visualisation have shown that accuracy will not help to decide on a classifier for these data sets. As such looking at other factors is necessary. FLDA and Simple logistic classifiers are fast and efficient and produce relatively easy to

interpret and explain models. This makes them ideal for datasets such as the diabetes data where a clear explanation of the models decisions is a crucial feature of the model.

A. Experimenter Output

ExperimenterOutput.txt

```
Tester:     weka.experiment.PairedCorrectedTTester -G 4,5,6 -D 1 -R 2 -S 0.05 -V
                       -result-matrix "weka.experiment.ResultMatrixPlainText -mean-prec 2
                       -stddev-prec 2 -col-name-width 0 -row-name-width 25 -mean-width 2
                       -stddev-width 2 -sig-width 1 -count-width 5 -show-stddev -print-col-names
                       -print-row-names -enum-col-names"
Analysing:  Percent_correct
Datasets:   3
Resultsets: 8
Confidence: 0.05 (two tailed)
Sorted by:  -
Date:       19/09/18 8:35 PM


Dataset          (1) bayes.NaiveBay | (2) functions.F (3) functions.Q (4) functions.L (5) functions.Li
-----------------------------------------------------------------------------------------------------
pima_diabetes   (100)   75.75(5.32) |   76.13(5.53)     73.95(4.86)     76.81(4.54)     53.20(13.21) *
ionosphere      (100)   82.17(6.14) |   87.04(5.26) v   89.92(4.09) v   88.07(5.37) v   88.35( 4.44) v
sonar           (100)   67.71(8.66) |   74.38(8.61)     81.79(8.39) v   76.60(8.27) v   53.38( 1.63) *
-----------------------------------------------------------------------------------------------------
                            (v/ /*) |         (1/2/0)         (2/1/0)         (2/1/0)          (1/0/2)


Dataset          (1) bayes.NaiveBay |(6) functions.Li (7) functions.L (8) trees.J48 '
-----------------------------------------------------------------------------------
pima_diabetes   (100)   75.75(5.32) |  46.48(15.76) *   65.11(0.34) *   74.49(5.27)
ionosphere      (100)   82.17(6.14) |  64.10( 1.36) *   93.05(3.95) v   89.74(4.38) v
sonar           (100)   67.71(8.66) |  53.38( 1.63) *   64.99(7.66)     73.61(9.34)
-----------------------------------------------------------------------------------
                            (v/ /*) |         (0/0/3)         (1/1/1)         (1/2/0)


Key:
(1) bayes.NaiveBayes '' 5995231201785697655
(2) functions.FLDA '-R 0.001' -9212385698193681291
(3) functions.QDA '-R 0.001' -9113383498193689291
(4) functions.LibSVM '-S 0 -K 0 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1
                -Z -model \"C:\\\\Program Files\\\\Weka-3-8\" -seed 1' 14172
(5) functions.LibSVM '-S 0 -K 1 -D 2 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1
                -model \"C:\\\\Program Files\\\\Weka-3-8\" -seed 1' 14172
(6) functions.LibSVM '-S 0 -K 1 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1
                -model \"C:\\\\Program Files\\\\Weka-3-8\" -seed 1' 14172
(7) functions.LibSVM '-S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1
                -model \"C:\\\\Program Files\\\\Weka-3-8\" -seed 1' 14172
(8) trees.J48 '-C 0.25 -M 2' -217733168393644444
```