CS440 : Introduction to Artificial Intelligence **Spring 2014**

Problem Set 2 - Part 1

Dan McQuillan *Handed In: February 9, 2014*

1. **Solution to problem 1**

   (a) **Training on data set 1**

   a) The final weights

   $Threshhold = 0.0$

   Weights:
   $w_0 \rightarrow -748.0$
   $w_1 \rightarrow 4775.0$
   $w_2 \rightarrow 17744.0$
   $w_3 \rightarrow 780.0$
   $w_4 \rightarrow -6.0$
   $w_5 \rightarrow 657.0$
   $w_6 \rightarrow 3041.0$
   $w_7 \rightarrow -2008.0$
   $w_8 \rightarrow 5501.0$
   $w_9 \rightarrow 9662.399999998896$
   $w_{10} \rightarrow 3846.0$
   $w_{11} \rightarrow 14912.0$
   $w_{12} \rightarrow 21058.0$

   b) The number of training epochs required
   The training took 1975 epochs.

   c) The margin
   $\gamma \rightarrow 2.206549570044712 * 10^{-5}$

   (b) **Test on data set 1**

   a) A confusion matrix

54	0
0	63

   b) Two lists of example indices
   No errors were found with this dataset since the weight vectors were calculated from this dataset.

   c) The total loss summed over the misclassified examples
   The loss for this dataset is 0.0 since it is the training data set.

(c) **Test on data set 2**

a) A confusion matrix

12	1
0	20

b) Two lists of example indices
False Negatives:
Index: 22
Inputs: $(60.0, 1.0, 4.0, 140.0, 293.0, 0.0, 2.0, 170.0, 0.0, 1.2, 2.0, 2.0, 7.0)$

False Positives:
There were no false positives

c) The total loss summed over the misclassified examples
Total loss: 448.1200000013341

(d) **Application to data set 3**
Classifications:
$0 \rightarrow 0.0$
$1 \rightarrow 1.0$
$2 \rightarrow 0.0$
$3 \rightarrow 1.0$
$4 \rightarrow 1.0$
$5 \rightarrow 0.0$
$6 \rightarrow 0.0$
$7 \rightarrow 0.0$
$8 \rightarrow 0.0$
$9 \rightarrow 1.0$
$10 \rightarrow 1.0$
$11 \rightarrow 0.0$
$12 \rightarrow 1.0$
$13 \rightarrow 0.0$
$14 \rightarrow 0.0$
$15 \rightarrow 0.0$
$16 \rightarrow 1.0$
$17 \rightarrow 0.0$
$18 \rightarrow 0.0$
$19 \rightarrow 0.0$
$20 \rightarrow 0.0$

Which feature is the most important?
In order to determine the most influential property I have used the correlation coefficients to find the most correlation between each property and the classification. The source code to find this property is also within the project.

Label	CorrelationCoefficient
age	0.576399638726158
sex	0.15811388300841894
chest	0.4934637712198269
resting blood pressure	0.08008953852726183
serum cholestoral	−0.32382553481251514
fasting blood sugar	0.31622776601683766
resting electrocardiographic results	0.1386750490563073
maximum heart rate achieved	−0.5353426981014223
exercise induced angina	0.685994340570035
oldpeak	0.8086701966434094
slope	0.6123724356957945
number of major vessels	0.8152133857595864
thal	0.43905703995876144

Property With Maximum Correlation
Label: number of major vessels
Correlation Coefficient: 0.8152133857595864