Transaction 1 for Enron1 Data

Script executed to convert the strings in ARFF into words:

java weka.filters.unsupervised.attribute.StringToWordVector-R-W1000-prune-rate-1.0-N0-stemmerweka.core.stemmers.NullStemmer-stopwords-handlerweka.core.stopwords.Null-M1-tokenizerweka.core.tokenizers.WordTokenizer -delimiters \" \\r\\n\\t.,;:\\\'\\\"()?!\ -b -i data.arff(train) -o data_normalized.arff -r data.arff(test) -s data_normalized(test)

In each .arff segment in the above code appropriate address has to be provided and also the destination file address needs to be provided as well. Above script is more of a generic statement.

Script executed to execute Multi Layer Perceptron:

java -cp "%CLASSPATH%;C:\Users\Snow_Leopard\workspace\tmp\Weka\weka-3-7-12\weka.jar" weka.classifiers.functions.MultilayerPerceptron -L 1.0 -M 0.5 -N 100 -E 20 -H 10 -B -C -I -t " train\data_normalized.arff" -T " test\data_normalized.arff

As like the first script the .arff file statement requires correct file address and both the train and test file address are supposed to be provided. The above is more generic and is repetitive for other data set results.

Time taken to build model: 599.26 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 97.05 seconds

=== Summary ===

Correctly Classified Instances	307	67.9204 %
Incorrectly Classified Instances	145	32.0796 %
Kappa statistic	0	
Mean absolute error	0.3765	
Root mean squared error	0.4942	
Relative absolute error	88.4885 %	
Root relative squared error	105.6693 %	
Coverage of cases (0.95 level)	100 %	
Mean rel. region size (0.95 level)	100 %	
Total Number of Instances	452	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.000	0.000	0.000	0.000	0.000	0.000	0.521	0.344	spam
	1.000	1.000	0.679	1.000	0.809	0.000	0.521	0.675	ham
Weighted Avg.	0.679	0.679	0.461	0.679	0.549	0.000	0.521	0.569	

=== Confusion Matrix ===

a	b	<-- classified as
0	145	a = spam
0	307	b = ham

Transaction 2 for Enron4 dataset

The scripts above are generic and require file addresses as replacements.

Time taken to build model: 989.91 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 452.51 seconds

=== Summary ===

Correctly Classified Instances	386	71.7472 %
Incorrectly Classified Instances	152	28.2528 %
Kappa statistic	0	

Mean absolute error 0.3894

Root mean squared error 0.4503

Relative absolute error 99.2695 %

Root relative squared error 99.7929 %

Coverage of cases (0.95 level) 100 %

Mean rel. region size (0.95 level) 100 %

Total Number of Instances 538

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	1.000	0.717	1.000	0.835	0.000	0.756	0.867	spam
	0.000	0.000	0.000	0.000	0.000	0.000	0.756	0.550	ham
Weighted Avg.	0.717	0.717	0.515	0.717	0.599	0.000	0.756	0.778	

=== Confusion Matrix ===

a	b	<-- classified as
386	0	a = spam
152	0	b = ham

Transaction 3

Scripts are generic and the file addresses need to replace.

Time taken to build model: 619.77 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 87.95 seconds

=== Summary ===

Correctly Classified Instances	340	73.5931 %
Incorrectly Classified Instances	122	26.4069 %
Kappa statistic	0	
Mean absolute error	0.4794	
Root mean squared error	0.481	
Relative absolute error	123.1936 %	
Root relative squared error	109.0998 %	
Coverage of cases (0.95 level)	100 %	
Mean rel. region size (0.95 level)	100 %	
Total Number of Instances	462	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1.000	1.000	0.736	1.000	0.848	0.000	0.566	0.737	ham

	0.000	0.000	0.000	0.000	0.000	0.000	0.566	0.362	spam
Weighted Avg.	0.736	0.736	0.542	0.736	0.624	0.000	0.566	0.638	

=== Confusion Matrix ===

a	b	<-- classified as
340	0	a = ham
122	0	b = spam