1.1

a) AdaBoostM1(Weka) - Known as AdaBoost.M1 in popular literature -A multi-class extension of Discrete AdaBoost. In AdaBoost.M1 the weight of a base classifier is a function of the error rate.

b) LMT(Weka) - Known as logistic model tree, is a classification model with an associated supervised training algorithm that combines logistic regression and decision tree learning.

c) RandomForest(Weka) - Known as random forests are an ensemble learning method for classification (and regression) that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes output by individual trees. The algorithm for inducing a random forest was developed by Leo Breiman and Adele Cutler.

d)SimpleCart(Weka) - Classification and regression trees (CART) are decision tree learning techniques that produce classification or regression trees. The dependent variable defines which kind of tree is produced.

1.2 and 1.3

Note: Whenever we refer to error, we mean the ratio of the error with $error_{NB}$.

a) AdaBoostM1(Weka) - Parameter changed for tuning - Weight Threshold. Tried for multiples of 5 from 10-100. The default value is 100, we found better results from 55 to 65.

Weight Threshold	Average Error	Max Error
10	1.29996284293	2.70588235294
15	1.29996284293	2.70588235294
20	1.29996284293	2.70588235294
25	1.29996284293	2.70588235294
30	1.29996284293	2.70588235294
35	1.29996284293	2.70588235294
40	1.32850165571	2.70588235294
45	1.32850165571	2.70588235294
50	1.16646461868	2.70588235294
55	1.14045105842	2.70588235294
60	1.14045105842	2.70588235294
65	1.14045105842	2.70588235294
70	1.17327934124	2.70588235294
75	1.15829251647	2.70588235294
80	1.15893690027	2.70588235294
85	1.24290869892	2.70588235294
90	1.23536734671	2.70588235294
95	1.23359192874	2.70588235294
100	1.1910652123	2.70588235294

b) LMT(Weka) - We tried tuning numInstances (minimum number of instances at which a node can be split - default 15), but found NO significant difference between the different values tested. Maximum and average error rates were nearly unchanged when we varied it by using a CVParameterSelection.
Number of boosting Iterations is set to -1, which means that it doesn't stop until it converges, however this takes an extremely long amount of time(upwards of 10 seconds for the anneal test), so we decided to set the number of iterations to 10, as we think this provides a good balance between run time and robustness.

c) RandomForest(Weka) - Parameter changed for tuning - Num Trees. Tried for values from 5 to 24. The default value is 10. Better performance was attained at higher values of 16 and 17.

Num Trees	Average Error	Max Error
5	0.778362890319	1.16129032258
6	0.76978526674	1.3064516129
7	0.811430423439	1.20967741935
8	0.757306938741	1.24193548387
9	0.774998425857	1.14516129032
10	0.748494570827	1.20967741935
11	0.749582107554	1.09677419355
12	0.752637444991	1.11290322581
13	0.760228015768	1.08064516129
14	0.755485033387	1.08064516129
15	0.766496918997	1.16129032258
16	0.744010728485	1.11290322581
17	0.740043999557	1.12903225806
18	0.75285282416	1.12903225806
19	0.750311836935	1.13793103448
20	0.762843327946	1.13793103448
21	0.756864645693	1.11290322581
22	0.767627710138	1.14516129032
23	0.762580699823	1.13793103448
24	0.764078054712	1.12903225806

d) SimpleCart(Weka) - Parameter changed for tuning - Num Folds. Tried for values from 2 to 19. The default value is 5. We got the following average errors and max errors. 9 performs no worse than 5 for all datasets, while 16 is also better in terms of smaller average error.

numFolds	average error	max error
2	0.799706702575	1.1935483871
3	0.77754342326	1.1935483871
4	0.76450973681	1.17741935484
5	0.760874952683	1.1935483871
6	0.794561132469	1.83333333333
7	0.811710555805	1.83333333333
8	0.801022693959	1.83333333333
9	0.746828794176	1.17741935484
10	0.792491757174	1.5

11	0.860592115597	2
12	0.754154774871	1.1935483871
13	0.77787454581	1.5
14	0.771809449592	1.66666666667
15	0.790486295491	1.5
16	0.738896051187	1.22222222222
17	0.799777751537	1.66666666667
18	0.775950657976	1.66666666667
19	0.836310934827	1.5

FINAL CLASSIFER

Our best classifier out LMT, RandomForest, SimpleCart, and AdaBoostM1, measured in terms of average and maximum robustness was **LMT**. Our average error ratio was 0.6930028637520893, while the maximum error ratio was 1.0555555555555556. The average time per test was 2.804 s.