B-659 – Applied Machine Learning– PA- 2

Team –Prashanth Balasubramani (prasbala), Siddharth Jayasankar (sidjayas)

RESULTS:

Bagging :

a) depth 3, bags 5

```
C:\Users\Narmi\Desktop\AML\PA2>python pa2_template.py bag 3 5 C:/Users/Narmi/Desktop/AML/PA2
*************************************************
        Algorithm : Bagging
        Decision Tree for Depth : 3
        Number of Bags : 5
*************************************************
Total Count : 2125
Correctly Classified : 1629
Misclassified Count : 496
                CONFUSION MATRIX
         Predicted-             Predicted+
Actual-   1597                  496
Actual+   0                     32
Accuracy : 0.766588235294
```

b) depth 3, bags 10

```
C:\Users\Narmi\Desktop\AML\PA2>python pa2_template.py bag 3 10 C:/Users/Narmi/Desktop/AML/PA2
*************************************************
        Algorithm : Bagging
        Decision Tree for Depth : 3
        Number of Bags : 10
*************************************************
Total Count : 2125
Correctly Classified : 1629
Misclassified Count : 496
                CONFUSION MATRIX
         Predicted-             Predicted+
Actual-   1597                  496
Actual+   0                     32
Accuracy : 0.766588235294
```

c) depth 5, bags 5

```
C:\Users\Narmi\Desktop\AML\PA2>python pa2_template.py bag 5 5 C:/Users/Narmi/Desktop/AML/PA2
*************************************************
        Algorithm : Bagging
        Decision Tree for Depth : 5
        Number of Bags : 5
*************************************************
Total Count : 2125
Correctly Classified : 1629
Misclassified Count : 496
                CONFUSION MATRIX
         Predicted-             Predicted+
Actual-   1597                  496
Actual+   0                     32
Accuracy : 0.766588235294
```

d) depth 5, bags 10

```
C:\Users\Narmi\Desktop\AML\PA2>python pa2_template.py bag 5 10 C:/Users/Narmi/Desktop/AML/PA2
*************************************************
        Algorithm : Bagging
        Decision Tree for Depth : 5
        Number of Bags : 10
*************************************************
Total Count : 2125
Correctly Classified : 1629
Misclassified Count : 496
                CONFUSION MATRIX
         Predicted-             Predicted+
Actual-    1597                 496
Actual+    0                    32
Accuracy : 0.766588235294
```

AdaBoost:

a) depth 1, trees 5

```
C:\Users\Narmi\Desktop\AML\PA2>python pa2_template.py boost 1 5 C:/Users/Narmi/Desktop/AML/PA2
*************************************************
        Algorithm : Boosting
        Decision Tree for Depth : 1
        Number of Trees : 5
*************************************************
Total Count : 2125
Correctly Classified : 1592
Misclassified Count : 533
                CONFUSION MATRIX
         Predicted-             Predicted+
Actual-    1560                 533
Actual+    0                    32
Accuracy : 0.749176470588
```

b) depth 1, trees 10

```
C:\Users\Narmi\Desktop\AML\PA2>python pa2_template.py boost 1 10 C:/Users/Narmi/Desktop/AML/PA2
*************************************************
        Algorithm : Boosting
        Decision Tree for Depth : 1
        Number of Trees : 10
*************************************************
Total Count : 2125
Correctly Classified : 1592
Misclassified Count : 533
                CONFUSION MATRIX
         Predicted-             Predicted+
Actual-    1560                 533
Actual+    0                    32
Accuracy : 0.749176470588
```

c) depth 2, trees 5

```
C:\Users\Narmi\Desktop\AML\PA2>python pa2_template.py boost 2 5 C:/Users/Narmi/Desktop/AML/PA2
**************************************************
        Algorithm : Boosting
        Decision Tree for Depth : 2
        Number of Trees : 5
**************************************************
Total Count : 2125
Correctly Classified : 1629
Misclassified Count : 496
                CONFUSION MATRIX
          Predicted-              Predicted+
Actual-    1597                   496
Actual+    0                      32
Accuracy : 0.766588235294
```

d) depth 2, trees 10

```
C:\Users\Narmi\Desktop\AML\PA2>python pa2_template.py boost 2 10 C:/Users/Narmi/Desktop/AML/PA2
**************************************************
        Algorithm : Boosting
        Decision Tree for Depth : 2
        Number of Trees : 10
**************************************************
Total Count : 2125
Correctly Classified : 1629
Misclassified Count : 496
                CONFUSION MATRIX
          Predicted-              Predicted+
Actual-    1597                   496
Actual+    0                      32
Accuracy : 0.766588235294
```

Weka Results

The values of the target variable is changed from 0, 1 to N, P respectively to convert it to nominal values.

Bagging:

```
=== Run information ===
Scheme:       weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 10 -W weka.classifiers.trees.REPTree -- -M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0
Relation:     mashroomtrain-test-weka.filters.unsupervised.attribute.Reorder-R1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,22,23,24,25,26,27,28,29,
30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,
85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100,101,102,103,104,105,106,107,108,109,110,111,112,113,114,115,116,117,118,119,120,121,122,123,124,125,126,21
Instances:    8125
Attributes:   126
              [list of attributes omitted]
Test mode:    split 73.85% train, remainder test

=== Classifier model (full training set) ===

Bagging with 10 iterations and base learner

weka.classifiers.trees.REPTree -M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0

Time taken to build model: 4.37 seconds

=== Evaluation on test split ===

Time taken to test model on training split: 0.01 seconds

=== Summary ===

Correctly Classified Instances        1593               74.9647 %
Incorrectly Classified Instances       532               25.0353 %
Kappa statistic                          0
Mean absolute error                      0.2504
Root mean squared error                  0.5004
Relative absolute error                 44.4957 %
Root relative squared error             88.9286 %
Total Number of Instances             2125

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.000    0.250    0.000      0.000    0.000      0.000    ?         ?         P
                 0.750    0.000    1.000      0.750    0.857      0.000    ?         1.000     N
Weighted Avg.    0.750    0.000    1.000      0.750    0.857      0.000    0.000     1.000

=== Confusion Matrix ===

   a    b   <-- classified as
   0    0 |    a = P
 532 1593 |    b = N
```

Adaboost:

```
=== Run information ===
Scheme:       weka.classifiers.meta.AdaBoostM1 -P 100 -S 1 -I 10 -W weka.classifiers.trees.DecisionStump
Relation:     mashroomtrain-test-weka.filters.unsupervised.attribute.Reorder-R1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,22,23,24,25,26,
27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,
81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100,101,102,103,104,105,106,107,108,109,110,111,112,113,114,115,116,117,118,119,120,121,122,123,124,125,126,21
Instances:    8125
Attributes:   126
              [list of attributes omitted]
Test mode:    split 73.85% train, remainder test

=== Classifier model (full training set) ===

AdaBoostM1: Base classifiers and their weights:

Decision Stump

Classifications

ring-type-pendant <= 0.5 : N
ring-type-pendant > 0.5 : P
ring-type-pendant is missing : N

Class distributions

ring-type-pendant <= 0.5
P   N
0.04618715419773876 0.9538128458022612
ring-type-pendant > 0.5
P   N
0.8024193548387096  0.1975806451612903
ring-type-pendant is missing
P   N
0.4155076923076923  0.5844923076923076


Weight: 1.99

Decision Stump

Classifications

habitat-waste <= 0.5 : N
habitat-waste > 0.5 : P
habitat-waste is missing : N
```

```
Class distributions

habitat-waste <= 0.5
P   N
0.246981777489534   0.753018222510466
habitat-waste > 0.5
P   N
1.0000000000000018  -1.8379144360271516E-15
habitat-waste is missing
P   N
0.32104914363810294 0.6789508563618971


Weight: 1.25

Decision Stump

Classifications

stalk-surface-above-ring-silky <= 0.5 : P
stalk-surface-above-ring-silky > 0.5 : N
stalk-surface-above-ring-silky is missing : P

Class distributions

stalk-surface-above-ring-silky <= 0.5
P   N
0.6609063013262418  0.33909369867375805
stalk-surface-above-ring-silky > 0.5
P   N
-1.287519121347168E-16  1.0000000000000002
stalk-surface-above-ring-silky is missing
P   N
0.5632697793655754  0.43673022063442457


Weight: 0.9

Decision Stump

Classifications

gill-spacing-close <= 0.5 : N
gill-spacing-close > 0.5 : P
gill-spacing-close is missing : N
```

```
Class distributions

gill-spacing-close <= 0.5
P   N
0.046720987123565996    0.9532790128764339
gill-spacing-close > 0.5
P   N
0.509991831907344   0.49000816809265607
gill-spacing-close is missing
P   N
0.39611037578889835 0.6038896242111017


Weight: 0.49

Decision Stump

Classifications

stalk-root-missing <= 0.5 : N
stalk-root-missing > 0.5 : N
stalk-root-missing is missing : N

Class distributions

stalk-root-missing <= 0.5
P   N
0.48564178970732047 0.5143582102926796
stalk-root-missing > 0.5
P   N
0.08914201053384678 0.9108579894661533
stalk-root-missing is missing
P   N
0.32577310380532687 0.674226896194673


Weight: 0.73

Decision Stump

Classifications

stalk-root-missing <= 0.5 : P
stalk-root-missing > 0.5 : N
stalk-root-missing is missing : N
```

```
Class distributions

stalk-root-missing <= 0.5
P   N
0.6614845561894526  0.3385154438105475
stalk-root-missing > 0.5
P   N
0.1684306338222013  0.8315693661777988
stalk-root-missing is missing
P   N
0.4999999999999835  0.5000000000000164


Weight: 0.93

Decision Stump

Classifications

gill-size-broad <= 0.5 : N
gill-size-broad > 0.5 : P
gill-size-broad is missing : N

Class distributions

gill-size-broad <= 0.5
P   N
0.1130296085558805  0.8869703914441195
gill-size-broad > 0.5
P   N
0.5725609472539196  0.4274390527460805
gill-size-broad is missing
P   N
0.4076528203349045  0.5923471796650955


Weight: 0.78

Decision Stump

Classifications

habitat-waste <= 0.5 : N
habitat-waste > 0.5 : P
habitat-waste is missing : N
```

```
Class distributions

habitat-waste <= 0.5
P   N
0.281113160131391   0.7188868398686089
habitat-waste > 0.5
P   N
1.000000000000234   -2.3402829581471575E-13
habitat-waste is missing
P   N
0.33226088234708223 0.6677391176529178


Weight: 1.04

Decision Stump

Classifications

stalk-surface-below-ring-silky <= 0.5 : P
stalk-surface-below-ring-silky > 0.5 : N
stalk-surface-below-ring-silky is missing : P

Class distributions

stalk-surface-below-ring-silky <= 0.5
P   N
0.6212635232846437  0.37873647671535626
stalk-surface-below-ring-silky > 0.5
P   N
-1.4770949859835752E-14 1.0000000000000147
stalk-surface-below-ring-silky is missing
P   N
0.5481456859441176  0.45185431405588233


Weight: 0.69

Decision Stump

Classifications

cap-surface-scaly <= 0.5 : N
cap-surface-scaly > 0.5 : P
cap-surface-scaly is missing : N
```

```
Class distributions

cap-surface-scaly <= 0.5
P   N
0.30357127922199695 0.6964287207780031
cap-surface-scaly > 0.5
P   N
0.7354687891098726  0.2645312108901273
cap-surface-scaly is missing
P   N
0.4116210130870912  0.5883789869129087

Weight: 0.88

Number of performed Iterations: 10

Time taken to build model: 2.36 seconds

=== Evaluation on test split ===
Time taken to test model on training split: 0.01 seconds

=== Summary ===

Correctly Classified Instances        1611               75.8118 %
Incorrectly Classified Instances       514               24.1882 %
Kappa statistic                          0
Mean absolute error                      0.2454
Root mean squared error                  0.489
Relative absolute error                 43.6197 %
Root relative squared error             86.906  %
Total Number of Instances             2125

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.000    0.242    0.000      0.000    0.000      0.000    ?         ?         P
                 0.758    0.000    1.000      0.758    0.862      0.000    ?         1.000     N
Weighted Avg.    0.758    0.000    1.000      0.758    0.862      0.000    0.000     1.000

=== Confusion Matrix ===

   a    b   <-- classified as
   0    0 |    a = P
 514 1611 |    b = N
```