APPLIED MACHINE LEARNING

PROGRAMMING ASSIGNMENT 2

TEAM MEMBERS: GANESH NAGARAJAN

DIPANWITA MALLICK

Q1. For bagging, try two depths 3 and 5 and two sets of bags 5 and 10.

Ans.

For depth 3, bags 5:

```
[gnagaraj@hulk ~]$ python3 gnagaraj.py bag 3 5 "."
,0,1
0,1561,532
1,0,32
0.7496470588235294
[gnagaraj@hulk ~]$
```

For depth 5, bags 5:

```
[gnagaraj@hulk ~]$ python3 gnagaraj.py bag 5 5 "."
,0,1
0,1597,496
1,0,32
0.7665882352941177
[gnagaraj@hulk ~]$
```

For depth 3, bags 10:

```
[gnagaraj@hulk ~]$ python3 gnagaraj.py bag 3 10 "."
,0,1
0,1597,496
1,0,32
0.7665882352941177
[gnagaraj@hulk ~]$
```

For depth 5, bags 10:

```
[gnagaraj@hulk ~]$ python3 gnagaraj.py bag 5 10 "."
,0,1
0,1553,540
1,0,32
0.7458823529411764
[gnagaraj@hulk ~]$
```

Q2. For Adaboost, try two depths 1 and 2. Try two bags 5 and 10 trees.

Ans.

For depth 1, bags 5:

```
[gnagaraj@hulk ~]$ python3 gnagaraj.py boost 1 5 "."
,-1,1
-1,1597,496
1,0,32
0.7665882352941177
[gnagaraj@hulk ~]$
```

For depth 2, bags 5:

```
[gnagaraj@hulk ~]$ python3 gnagaraj.py boost 2 5 "."
,-1,1
-1,1553,540
1,0,32
0.7458823529411764
[gnagaraj@hulk ~]$
```

For depth 1, bags 10:

```
[gnagaraj@hulk ~]$ python3 gnagaraj.py boost 1 10 "."
,-1,1
-1,1597,496
1,0,32
0.7665882352941177
[gnagaraj@hulk ~]$
```

For depth 2, bags 10:

```
[gnagaraj@hulk ~]$  python3 gnagaraj.py boost 2 10 "."
,-1,1
-1,1597,496
1,0,32
0.7665882352941177
[gnagaraj@hulk ~]$
```

Q3. Report the confusion matrix for these four settings.

Ans.

The confusion matrices have been provided already in the above two questions.

Q4. Now, use Weka's default Adaboost and bagging and present their results.

Ans.

For Bagging:

No. of iterations: 5

```
Correctly Classified Instances        1593               74.9647 %
Incorrectly Classified Instances       532               25.0353 %
Kappa statistic                          0.0812
Mean absolute error                      0.2504
Root mean squared error                  0.5004
Relative absolute error                 45.0609 %
Root relative squared error             90.0298 %
Total Number of Instances             2125

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.746    0.000    1.000      0.746    0.854      0.206    0.873     0.996     0
                 1.000    0.254    0.057      1.000    0.107      0.206    0.873     0.057     1
Weighted Avg.    0.750    0.004    0.986      0.750    0.843      0.206    0.873     0.982

=== Confusion Matrix ===

    a    b   <-- classified as
 1561  532 |    a = 0
   0   32 |    b = 1
```

No. of iterations: 10

```
=== Summary ===

Correctly Classified Instances        1593               74.9647 %
Incorrectly Classified Instances       532               25.0353 %
Kappa statistic                          0.0812
Mean absolute error                      0.2504
Root mean squared error                  0.5004
Relative absolute error                 45.0609 %
Root relative squared error             90.0298 %
Total Number of Instances             2125

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.746    0.000    1.000      0.746    0.854      0.206    0.873     0.996     0
                 1.000    0.254    0.057      1.000    0.107      0.206    0.873     0.057     1
Weighted Avg.    0.750    0.004    0.986      0.750    0.843      0.206    0.873     0.982

=== Confusion Matrix ===

    a    b   <-- classified as
 1561  532 |    a = 0
   0   32 |    b = 1
```

For Adaboost:

No. of iterations: 5

```
=== Summary ===

Correctly Classified Instances        1629               76.6588 %
Incorrectly Classified Instances       496               23.3412 %
Kappa statistic                          0.0884
Mean absolute error                      0.2363
Root mean squared error                  0.4799
Relative absolute error                 42.5389 %
Root relative squared error             86.3464 %
Total Number of Instances             2125

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.763    0.000    1.000      0.763    0.866      0.215    0.966     0.999     0
               1.000    0.237    0.061      1.000    0.114      0.215    0.966     0.182     1
Weighted Avg.  0.767    0.004    0.986      0.767    0.854      0.215    0.966     0.987

=== Confusion Matrix ===

   a    b   <-- classified as
 1597  496 |    a = 0
   0   32 |    b = 1
```

No. of iterations: 10

```
=== Summary ===

Correctly Classified Instances        1611               75.8118 %
Incorrectly Classified Instances       514               24.1882 %
Kappa statistic                          0.0847
Mean absolute error                      0.2452
Root mean squared error                  0.4889
Relative absolute error                 44.1395 %
Root relative squared error             87.9716 %
Total Number of Instances             2125

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.754    0.000    1.000      0.754    0.860      0.210    0.994     1.000     0
               1.000    0.246    0.059      1.000    0.111      0.210    0.994     0.571     1
Weighted Avg.  0.758    0.004    0.986      0.758    0.849      0.210    0.994     0.993

=== Confusion Matrix ===

   a    b   <-- classified as
 1579  514 |    a = 0
   0   32 |    b = 1
```
