**Data Mining III**

**CSE 40977**

**Assignment III**

1 – Use three other meta learners from the long list of weka’s meta learners (besides bagging, boosting and stacking) on the IMAGE\_AssignmentII.arff dataset and report your results.

2 – Using the same dataset, and starting with the Decision Tree method (J48), apply the bagging, boosting and stacking on the IMAGE data. Report the results.

3 - Using the same dataset, and starting with the Grafting Decision Tree method (J48graft), apply the bagging, boosting and stacking on the IMAGE data. Report the results tested on the unseen data found in ImageSegmentationData\_TestSet\_199attr.arff.

4 - Finally, in preparations for your final assignment, I wanted to reiterate what makes an adequate model. A good model not only has a high % correctly classified instances (for nominal class data), but most of the instances are on the main diagonal of the confusion matrix, the TP rate is high for all of the class values, and the model is not trivial, i.e. M = 0, or always a class value, or always 1 etc. It is possible to have a high overal % of correctly classified instances and not an adequate model (i.e. one class is almost never predicted, but on average, the % cc is high).  Here is an example of a good model: === Stratified cross-validation === === Summary === Correctly Classified Instances 2244 97.1429 % Incorrectly Classified Instances 66 2.8571 %Kappa statistic 0.9667 Mean absolute error 0.0093 Root mean squared error 0.0894 Relative absolute error 3.7783 % Root relative squared error 25.5346 % Total Number of Instances 2310 === Detailed Accuracy By Class === TP Rate FP Rate Precision Recall F-Measure ROC Area Class 0.982 0.005 0.973 0.982 0.977 0.988 brickface 1 0.001 0.994 1 0.997 0.999 sky 0.942 0.009 0.945 0.942 0.944 0.975 foliage 0.952 0.004 0.975 0.952 0.963 0.974 cement 0.927 0.014 0.916 0.927 0.922 0.957 window 1 0.001 0.997 1 0.998 1 path 0.997 0 1 0.997 0.998 0.998 grass Weighted Avg. 0.971 0.005 0.972 0.971 0.971 0.984 === Confusion Matrix === a b c d e f g <-- classified as 324 0 3 2 1 0 0 | a = brickface 0 330 0 0 0 0 0 | b = sky 2 1 311 1 15 0 0 | c = foliage 3 1 0 314 12 0 0 | d = cement 4 0 15 5 306 0 0 | e = window 0 0 0 0 0 330 0 | f = path 0 0 0 0 0 1 329 | g = grass For your final assignment, make sure you read the web site guidelines (of how to transform the area variable). You will need several iterations, and patience and enough time to complete this final project, of our highest level data mining class.  Please start looking at it as soon as possible, as it will take you longer amount of time than the other assignments.

**Workflow**

(1) The following meta learners were implemented on the IMAGE\_AssignmentII.arff file and the associated performance summary reported. Each meta learner was used at default parameters with 10-fold cross-validation applied.

|  |  |  |  |
| --- | --- | --- | --- |
| Classifier | Correctly classified instances (%) | Mean absolute error | Are majority of instances in main diagonal? |
| MultiClassClassifier | 92.6 | 0.234 | Yes |
| AttributeSelectedClassifier | 95.6 | 0.0142 | Yes |
| END | 97.3333 | 0.0165 | Yes |

(2) The J48 Decision Tree was applied in line with bagging, boosting, and stacking for the same dataset and the results and depicted below:

|  |  |  |  |
| --- | --- | --- | --- |
| Classifier | Correctly classified instances (%) | Mean absolute error | Are majority of instances in main diagonal? |
| weka.classifiers.meta.Bagging -P 100 -S 1 -I 10 -W weka.classifiers.trees.J48 -- -C 0.25 -M 2 | 96.7333 | 0.0171 | Yes |
| weka.classifiers.meta.AdaBoostM1 -P 100 -S 1 -I 10 -W weka.classifiers.trees.J48 -- -C 0.25 -M 2 | 97.4667 | 0.0071 | Yes |
| weka.classifiers.meta.StackingC -X 10 -M "weka.classifiers.functions.LinearRegression -S 1 -R 1.0E-8" -S 1 -B "weka.classifiers.trees.J48 -C 0.25 -M 2" | 95.7333 | 0.0227 | Yes |

(3) J48graft was applied in line with bagging, boosting, and stacking for the IMAGE data as training data at 10-fold cross validation and for the ImageSegmentationData\_TestSet\_199attr.arff as testing data. The results tested on the unseen data found in ImageSegmentationData\_TestSet\_199attr.arff are reported below:

|  |  |  |  |
| --- | --- | --- | --- |
| Classifier | Correctly classified instances (%) | Mean absolute error | Are majority of instances in main diagonal? |
| weka.classifiers.meta.Bagging -P 100 -S 1 -I 10 -W weka.classifiers.trees.J48graft -- -C 0.25 -M 2 | 96.9136 | 0.0163 | Yes |
| weka.classifiers.meta.AdaBoostM1 -P 100 -S 1 -I 10 -W weka.classifiers.trees.J48graft -- -C 0.25 -M 2 | 98.0247 | 0.0059 | Yes |
| weka.classifiers.meta.StackingC -X 10 -M "weka.classifiers.functions.LinearRegression -S 1 -R 1.0E-8" -S 1 -B "weka.classifiers.trees.J48graft -C 0.25 -M 2" | 95.9259 | 0.0206 | Yes |

(4) -