Assignment 6

Introduction to Machine Learning

Prof. B. Ravindran

1. Assume that you are given a data set and a neural network model trained on the data set. You are asked to build a decision tree model with the sole purpose of understanding/interpreting the built neural network model. In such a scenario, which among the following measures would you concentrate most on optimising?

   (a) Accuracy of the decision tree model on the given data set

   (b) F1 measure of the decision tree model on the given data set

   (c) Fidelity of the decision tree model, which is the fraction of instances on which the neural network and the decision tree give the same output

   (d) Comprehensibility of the decision tree model, measured in terms of the size of the corresponding rule set

2. Which of the following properties are characteristic of decision trees?

   (a) High bias

   (b) High variance

   (c) Lack of smoothness of prediction surfaces

   (d) Unbounded parameter set

3. To control the size of the tree, we need to control the number of regions. One approach to do this would be to split tree nodes only if the resultant decrease in the sum of squares error exceeds some threshold. For the described method, which among the following are true?

   (a) It would, in general, help restrict the size of the trees

   (b) It has the potential to affect the performance of the resultant regression/classification model

   (c) It is computationally infeasible

4. Which among the following statements best describes our approach to learning decision trees

   (a) Identify the best partition of the input space and response per partition to minimise sum of squares error

   (b) Identify the best approximation of the above by the greedy approach (to identifying the partitions)

   (c) Identify the model which gives the best performance using the greedy approximation (option (b)) with the smallest partition scheme

   (d) Identify the model which gives performance close to the best performance (option (a)) with the smallest partition scheme

   (e) Identify the model which gives performance close to the best greedy approximation performance (option (b)) with the smallest partition scheme

5. Having built a decision tree, we are using reduced error pruning to reduce the size of the tree. We select a node to collapse. For this particular node, on the left branch, there are 3 training data points with the following outputs: 5, 7, 9.6 and for the right branch, there are four training data points with the following outputs: 8.7, 9.8, 10.5, 11. What were the original responses for data points along the two branches (left & right respectively) and what is the new response after collapsing the node?

   (a) 10.8, 13.33, 14.48
   (b) 10.8, 13.33, 12.06
   (c) 7.2, 10, 8.8
   (d) 7.2, 10, 8.6

6. Given that we can select the same feature multiple times during the recursive partitioning of the input space, is it always possible to achieve 100% accuracy on the training data (given that we allow for trees to grow to their maximum size) when building decision trees?

   (a) Yes
   (b) No

7. Suppose on performing reduced error pruning, we collapsed a node and observed an improvement in the prediction accuracy on the validation set. Which among the following statements are possible in light of the performance improvement observed?

   (a) The collapsed node helped overcome the effect of one or more noise affected data points in the training set
   (b) The validation set had one or more noise affected data points in the region corresponding to the collapsed node
   (c) The validation set did not have any data points along at least one of the collapsed branches
   (d) The validation set did have data points adversely affected by the collapsed node

8. Consider the following data set:

price	maintenance	capacity	airbag	profitable
low	low	2	no	yes
low	med	4	yes	no
low	low	4	no	yes
low	high	4	no	no
med	med	4	no	no
med	med	4	yes	yes
med	high	2	yes	no
med	high	5	no	yes
high	med	4	yes	yes
high	high	2	yes	no
high	high	5	yes	yes

   Considering 'profitable' as the binary values attribute we are trying to predict, which of the attributes would you select as the root in a decision tree with multi-way splits using the cross-entropy impurity measure?

(a) price

(b) maintenance

(c) capacity

(d) airbag

9. For the same data set, suppose we decide to construct a decision tree using binary splits and the Gini index impurity measure. Which among the following feature and split point combinations would be the best to use as the root node assuming that we consider each of the input features to be unordered?

   (a) price - {low, med}|{high}

   (b) maintenance - {high}|{med, low}

   (c) maintenance - {high, med}|{low}

   (d) capacity - {2}|{4, 5}

10. Consider building a spam filter for distinguishing between genuine e-mails and unwanted spam e-mails. Assuming spam to be the positive class, which among the following would be more important to optimise?

    (a) Precision

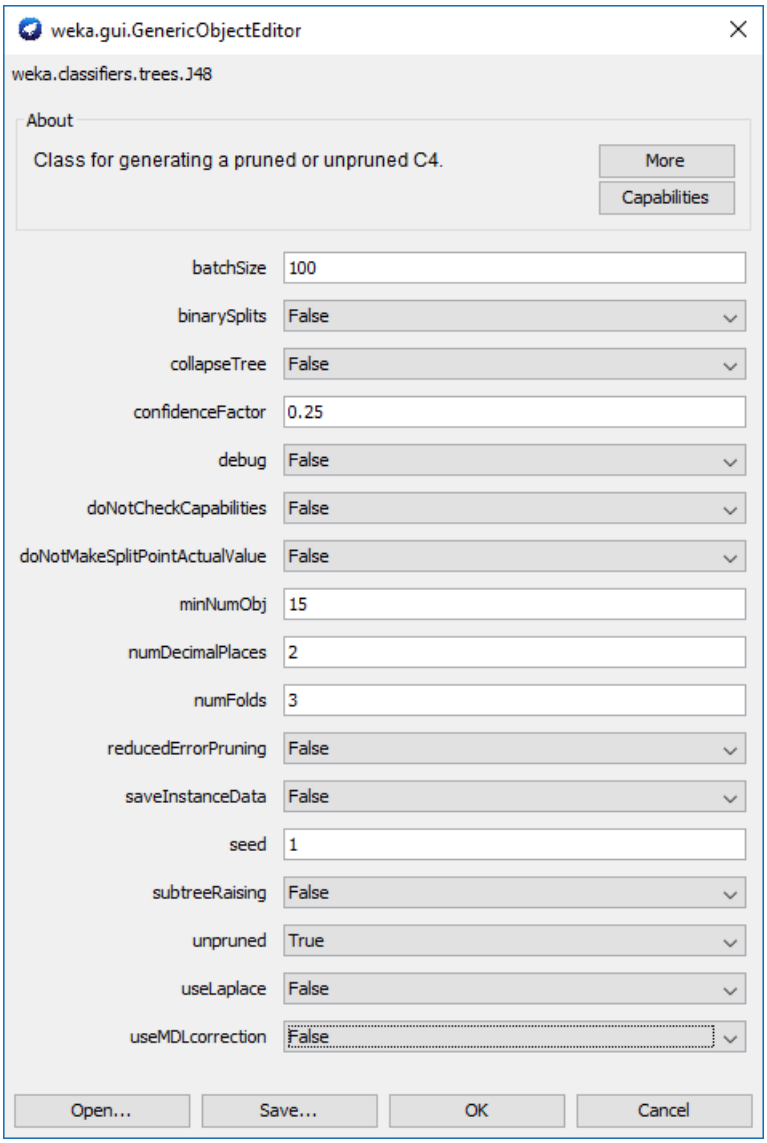
    (b) Recall

Weka-based assignment questions

In this assignment, we will use the UCI Mushroom data set available here.

We will be using the J48 decision tree algorithm which can be found in Weka under classifiers/trees.

We will consider the following to be the default parameter settings:



Note the following:

The class for which the prediction model is to be learned is named 'class' and is the first attribute in the data.

We will use the default 'Cross-validation' test option with 'Folds = 10'.

Once a decision tree model has been built, you can right click on the corresponding entry in the 'Result list' pane on the bottom left, and select 'Visualize tree' to see a visual representation of the learned tree.

11. How many levels does the unpruned tree contain considering multi-way and binary splits respectively, with the other parameters remaining the same as above?

(a) 6, 8

(b) 6, 7

(c) 5, 7

(d) 5, 8

12. How many levels does the pruned tree (unpruned = false, reducedErrorPruning = false) contain considering multi-way and binary splits respectively?

    (a) 6, 6

    (b) 6, 7

    (c) 5, 7

    (d) 5, 6

13. Consider the effect of the parameter minNumObj. Try modifying the parameter value and observe the changes in the performance values. For example, considering respectively, multi-way and binary splits with the default parameters as discussed before, at what (maximum) value of the parameter do we start to see zero error?

    (a) 11, 7

    (b) 11, 6

    (c) 10, 6

    (d) 10, 7

14. Which among the following pairs of attributes seem to be the most important for this particular classification task?

    (a) population, gill-spacing

    (b) stalk-surface-below-ring, cap-color

    (c) gill-spacing, ring-number

    (d) odor, spore-print-color