AI Principles and Applications

Fall 2020 ------ Homework #3

Due Date: 9PM, Nov. 12th, 2020

Note: Submit all your answers included within a single .pdf file. Python code is to be used for listing only and it will not be executed from the submitted file.

Q1. Consider the description for using SciKit Learn utilities ([at this LINK](https://stackabuse.com/decision-trees-in-python-with-scikit-learn/)) for creating decision trees from data and analyzing their performance. Also, consider the data set described on a 2-D grid, given in the attached HW3Data.xlsx file. For each 2-D coordinate pair, the third column states the class to which this point belongs. The only possible classes are “0” and “1”. Perform the following tasks in the context of this tool for creating decision trees using the data in HW3Data.xls. For each part below submit the items shown in red font, **along with the Python code used for obtaining each result.**

1. (10) Select randomly 70% of the data points from HW3Data.xlsx. These points are to be used for learning the decision tree. Learn the decision tree for this training data, and making sure that each leaf of the tree is at least 90% pure. Show the decision tree obtained in either the tree or the list form, or in any other suitable display format. Your decision tree should be readable for grading purposes. It should include at each internal node of the tree the attribute test performed at that node. (There are many automated tools for various types of display)
2. (20) Use the remaining 30% of the data points in HW3Data.xlsx for testing. Test them with the decision tree to find the predicted class label for each test data point. Use the information of the actual and the predicted class labels for each data point to create a confusion matrix for your learned decision tree. (There are automated tools for creating the confusion matrix). Show the confusion matrix and the following performance metrics derived for the decision tree: accuracy of the tree, precision and recall for class “1”, and precision and recall for class “0”. Show your work for computing these performance metrics.
3. (20) Now consider all the points (natural numbers only) of the (70 X 70) 2-D grid as test points for the decision tree. Find the predicted class label for each of the 4900 data points. Use programming help to generate the 4900 points and do not hand-code them. Create and show a display of the 70 X 70 2-D grid in which each point is displayed in color1 (say red) if its predicted class is “1”, and is displayed in color2 (say black) if its predicted class is “0”.
4. (10) Interpret the decision tree plotted in #3 above and comment if you would have drawn some of the boundaries differently. Plot on a 70X70 grid all the data points of HW3Data.xls. Using these data points as your guide, Show the decision tree boundaries you would draw and give reasons for your suggestions. Compare your suggested boundaries with the ones obtained in Question#1 above.