AI Principles and Applications

Fall 2019 ------ Homework #4

Due Date: 9PM, Nov. 25th, 2019

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 P4Data.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 decision trees and the data in P4Data.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 P4Data.xlsx. These points are to be used for learning the decision tree. Learn the decision tree for the 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. (There are many automated tools for various types of display)
2. (20) Use the remaining 30% of the data points in P4Data.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”.
3. (20) Now consider all the points (natural numbers only) of the (50 X 50) 2-D grid as test points for the decision tree. Find the predicted class label for each data point. Create and show a display of the 50 X 50 2-D grid in which each point is plotted in color1 (say red) if its predicted class is “1”, and is plotted 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. Point out the boundaries you would draw differently and give reasons for your suggestions.

|  |  |  |  |
| --- | --- | --- | --- |
| W | X | Y | Class |
| 3 | 7 | 0 | 0 |
| 12 | 4 | 4 | 1 |
| 7 | 1 | 3 | 0 |
| 8 | 2 | 9 | 1 |
| 9 | 6 | 5 | 1 |

Q2: (40) Consider the 3-D data points and their class labels shown in the table here. Use (1 2 3 4) as the initial set of weights for the perceptron. Show all the steps of the perceptron learning algorithm for two full epochs of the training data. That is, you train with the five data points once, and then 2nd time again the same five points are used to adjust the weights. Show the final weights obtained. How many data points are correctly classified after the first two epochs?