2014057983 컴퓨터공학과 지소현

1. Summary of Algorithm

**-Constructing descision trees**

Strategy : top-down Recursive divide and conquer fashion

1. First, select attribute for root node. Create branch for each possibile attribute value.
2. Then, split instances into subsets. One for each branch extending from the node.
3. Finally, repeat recursively for each branch, using only instances that reach the branch.
4. Stop if all instances have the same class.

**-Computing purity : the information measures**

1) information is a measure of a reduction of uncertainty

2) It represents the expected amount of information that would be needed to “place” a new instance in the branch.

**-Criterion for attribute selcection**

1) Best attribute ? wants to get the smallest tree, heuristic : choose the attribute that proceduces the “purest” node

2) Information gatin: increases with the average purity if the subsets

3) Strategy: choose attribute that gives greatest information gain

-How to compute Information Gain : calculate Entropy

1) When the number of either yes or no is zero (that is the node is pure) the information is zero

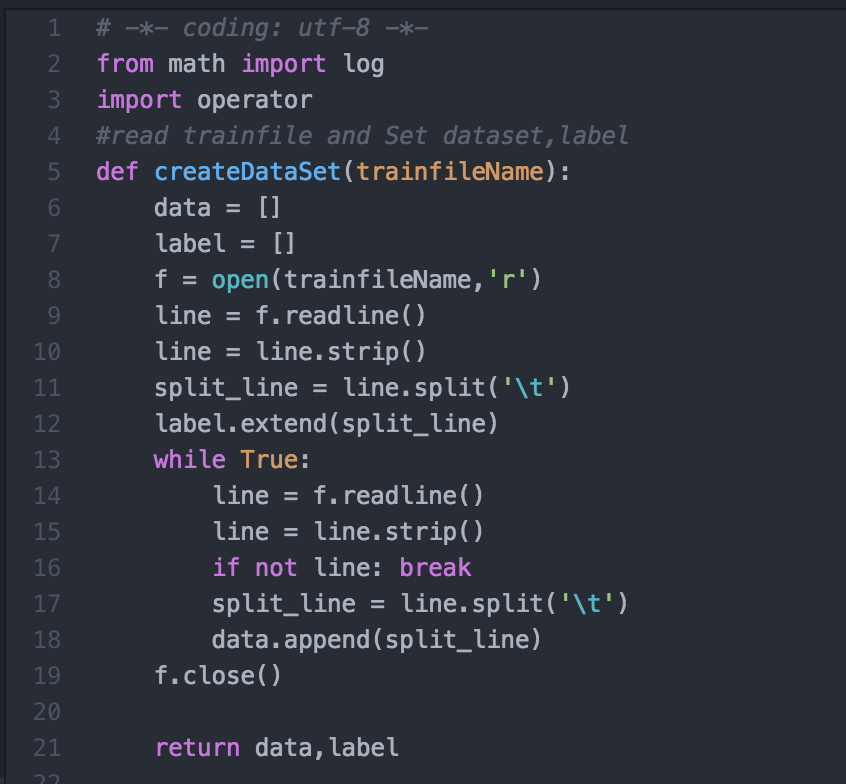
2) When the number of yes and no is equal, the information reaches its maximum because we are very uncertain about the outcome.

3) Complex scenarios : the measure should be applicable to a multiclass situation, where a multi-staged decision must be made.

1. Detailed description of codes

- createDataSet fuction is to read trainFile and setting dataset and label list.

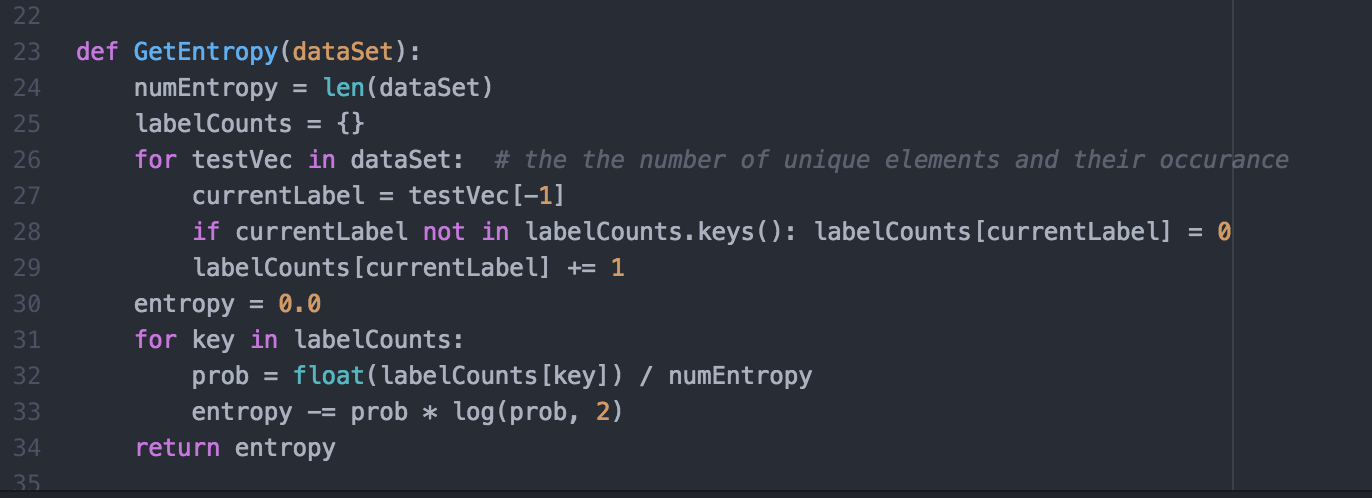
And then returns them.These datas are called by main function.



SplitDataSet And GetBestSplitCriterion is choosing best split criterion to make decision tree. To choose it, calurate entropy.

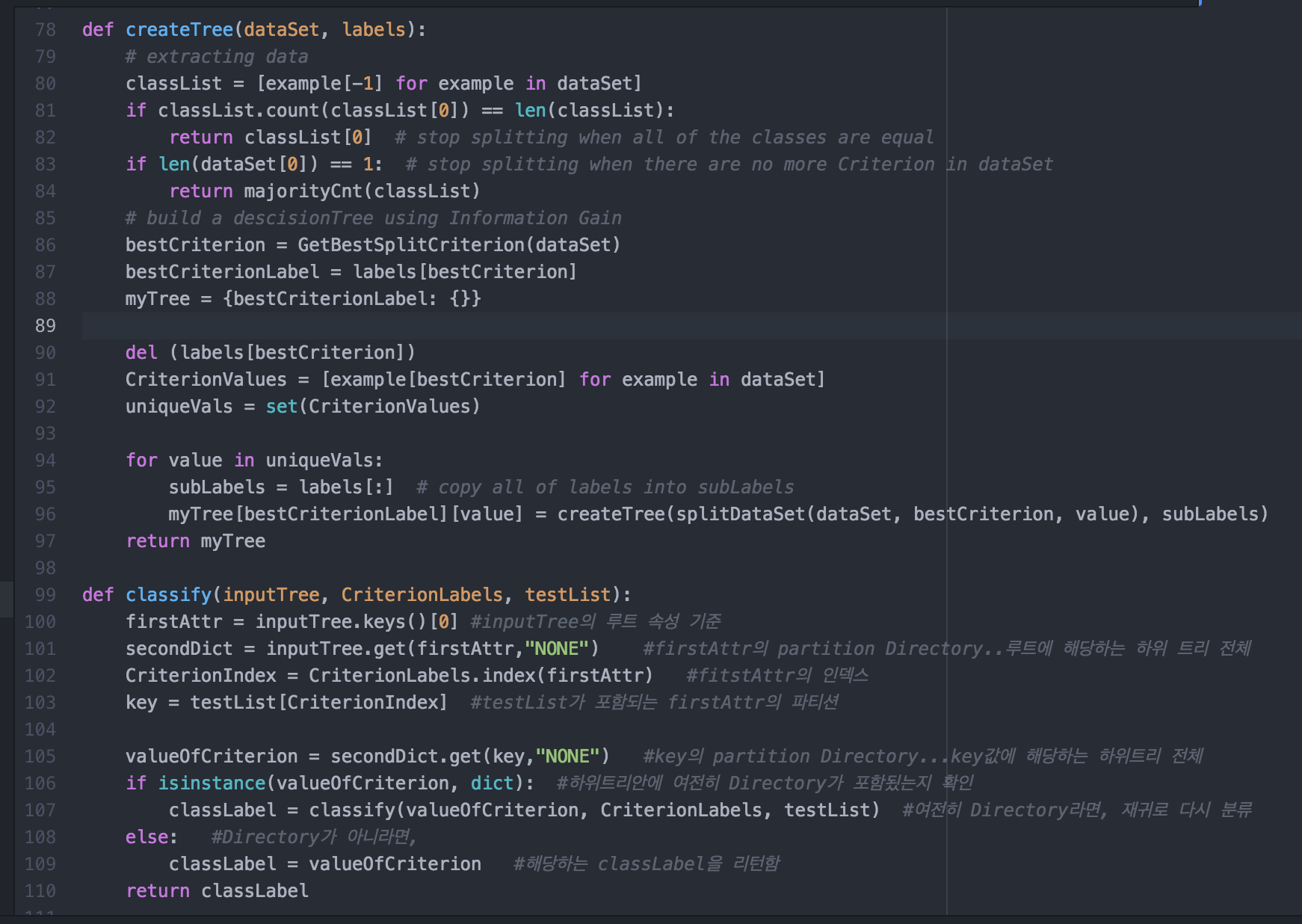
Best split criterion is attribute which has Max entropy.





CreateTree function makes descision tree.Leaf node is class label and internal node can classify class label with selection attribute using bestCriterion which has max Entropy.

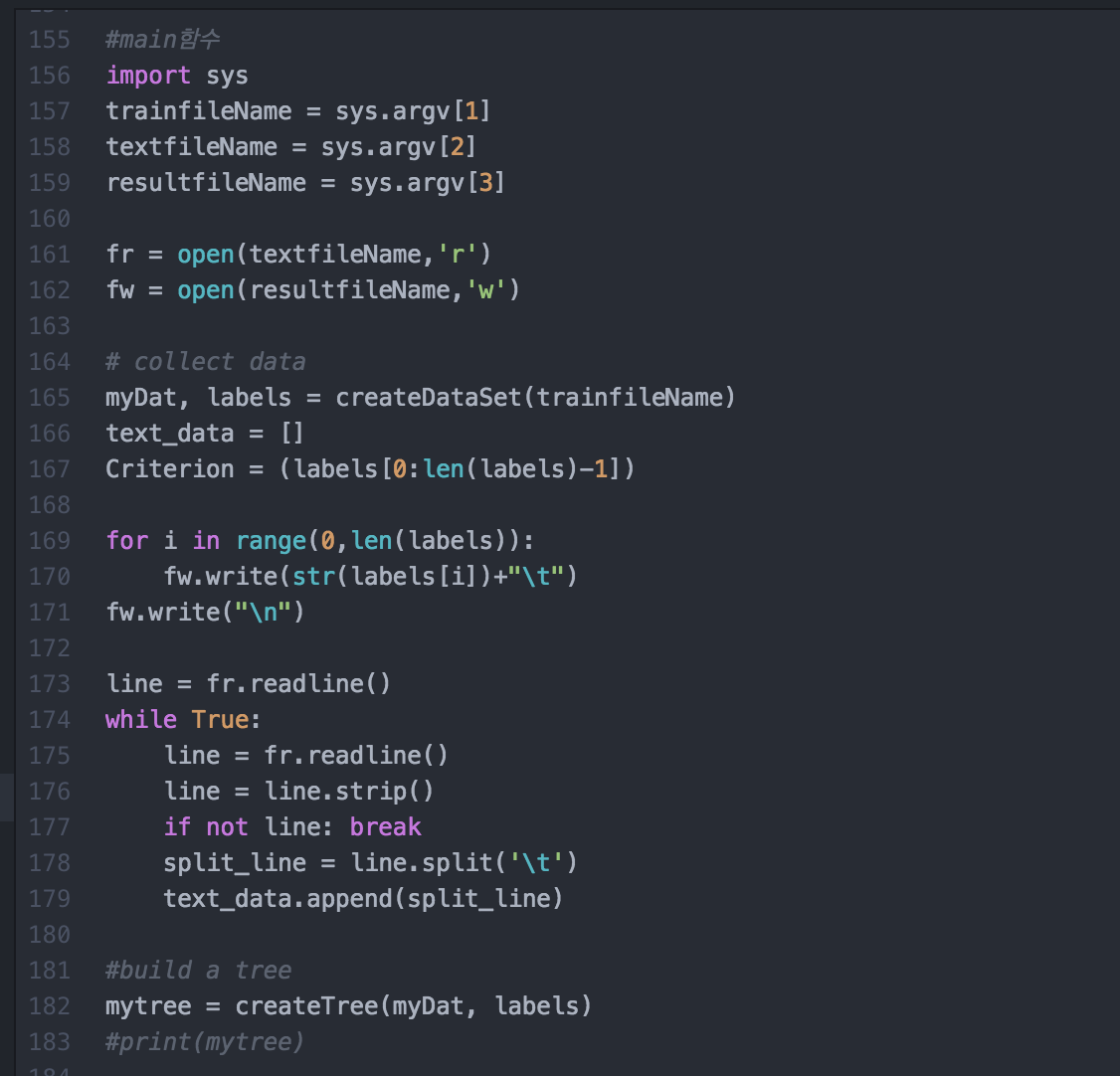
Classify function get class label using top-down Recursive divide and conquer fashion.

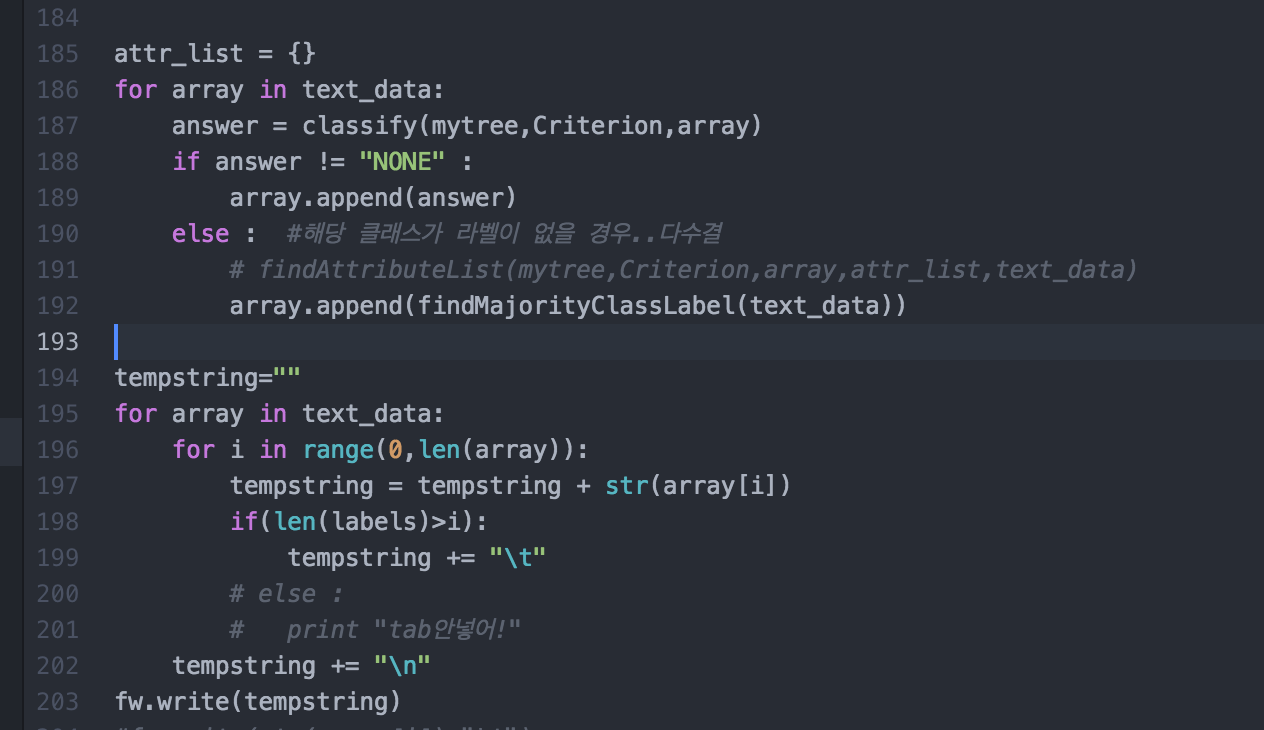


findMajorityClassLabel function get classLabel which has Majority instances.



Main Function... Make Descision tree and then applicate descision tree into text file to classify class labels.





1. Insrtuctions for compiling my source code(e.g screenshot)

Use this command…

python dt.py TrainFileName TextFileName ResultFileName

python dt.py dt\_train1.txt dt\_test1.txt dt\_result1.txt  
