**Decision Tree**

**- Programming assignment #2**

**2013011060**

**양승호**

**1. Environment**

Ubuntu Linux 18.04

Python 3.6.7

**2. How to run**

$ python3 dt.py [train\_file\_name] [test\_file\_name] [result\_file\_name]

(example: python3 dt.py dt\_train1.txt dt\_test1.txt dt\_result1.txt)

**3. Algorithm Summary**

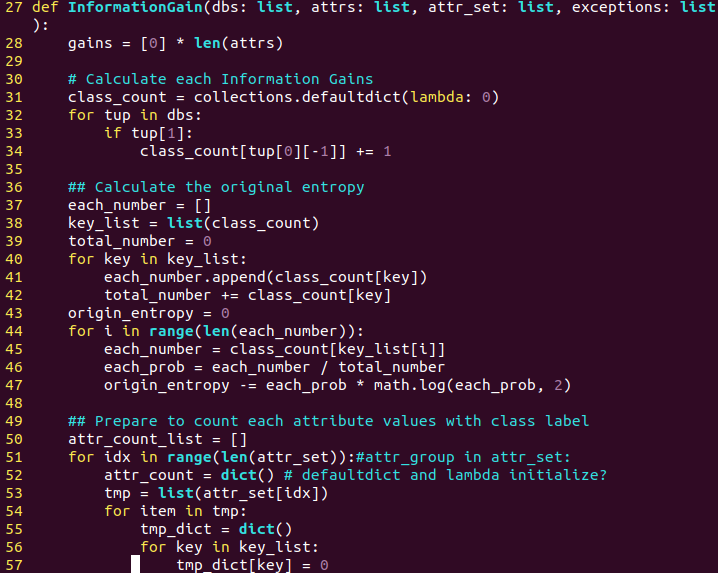
**- Decision Tree Constructing**

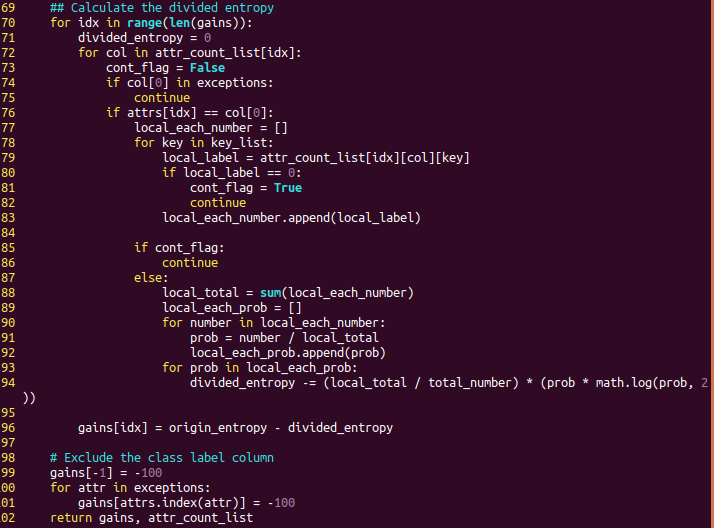
1. Find an attribute of maximum information gain from entire db.

2. Set the root node of decision tree using the information gain.

3. As moving to child node, modify the db following the attribute of parent node and repeat from [Procedure 1] to [Procedure 2] using the modified db until there’s neither samples nor attributes left, or all the remaining samples from the node are in same class.

**4. Details**





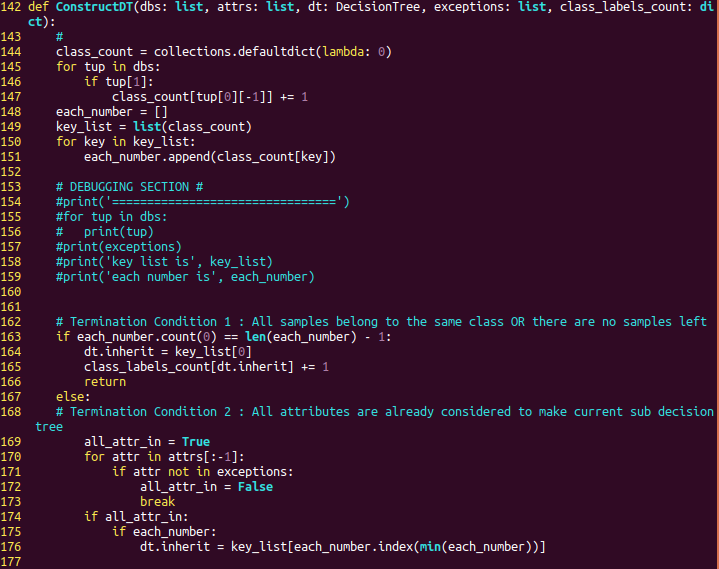
- InformationGain : return the list of information gain of each attributes using the modified dbs following the parents’ attributes.

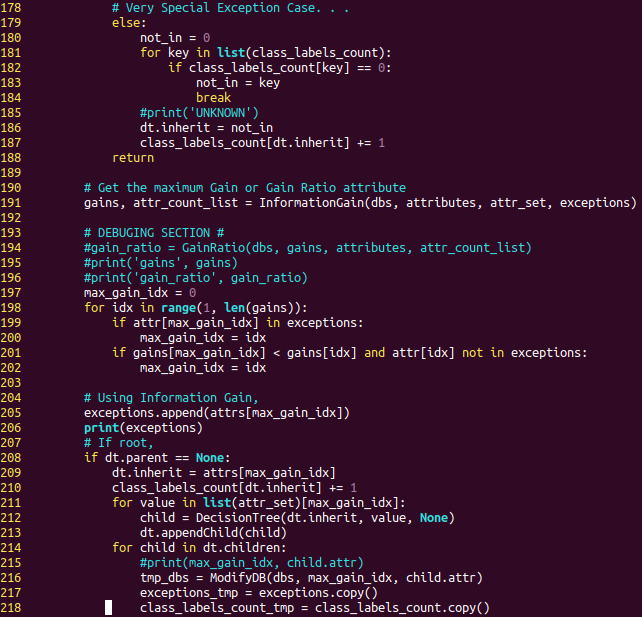
- First, get the number of class labels of db and calculate the original entropy

- and then calculate the information gain of each attribute except for attributes already chosen.

- Exclude the attributes already chosen by setting them the minimum value not to be chosen again.

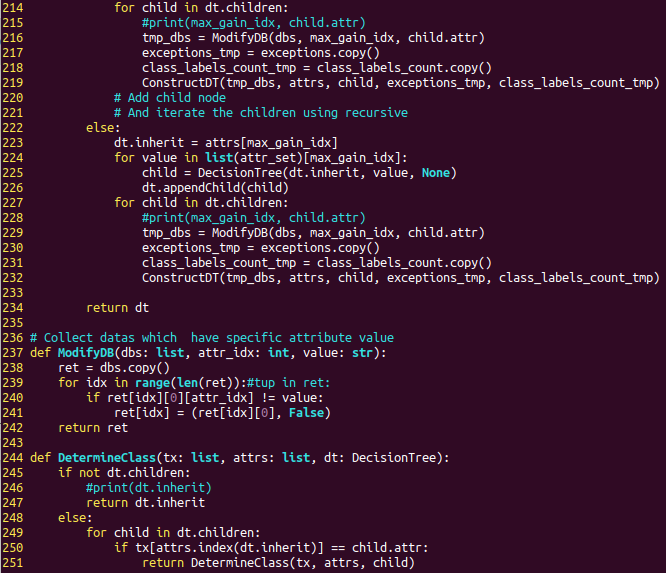
- GainRatio : fx that return gain ratio for studying(not used in this project)



- ConstructDT: Construct the Decision Tree using information gain

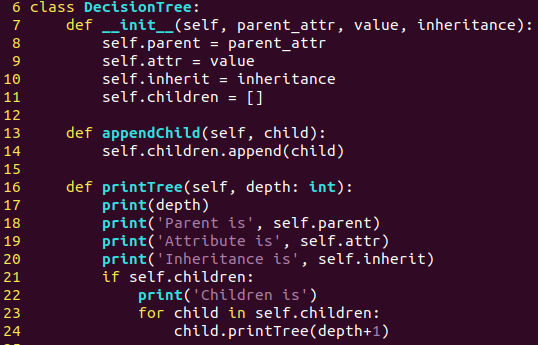
- if there are neither samples nor attributes or all the samples belong to same class or over the half of them does, then stop constructing the sub decision tree(Termination Condition)

- else, recursively keep constructing

- ModifyDB: modify the dbs using specific attributes not to make corresponding datas to be considerted to Decision Tree making.

- set True, False to each samples

- DetermineClass: Determine the class of test samples using decision tree already made.



- Decision Tree data structure

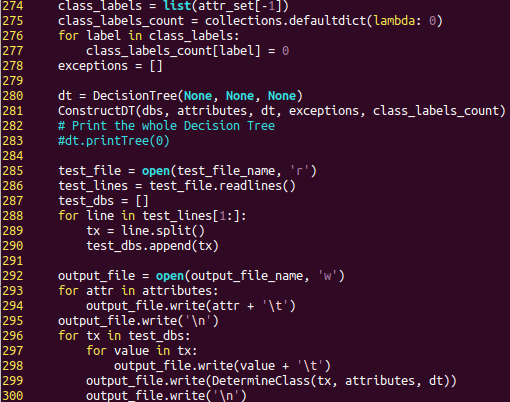
- parent stands for the attribute of parent, which current attr belongs to

- attr stands for the vale of parent

- inherit stnads for the attribute of children

- children is a list of child nodes of current node

- printTree: for debugging, check the decision tree made.



- main function

**5. Testing Result(result1.txt) T^T**

