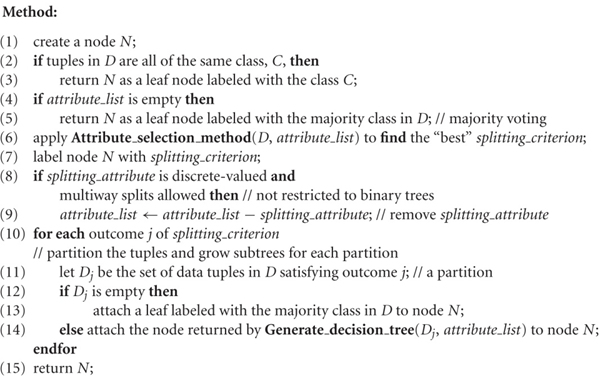
**Data mining**

**assignment 2**

**2011003865 박준형**

* **Summary of algorithm**

To generate decision tree, I refer to this page.



In this algorithm, I have to think about how to determine majority class and attribute.

To determine majority class, calculate weight of value of class. Suppose there are class value ‘Red’, ‘Green’, ‘Blue’ and the total number of class value is 5, 10, 15, respectively. Then, weight of the class value is 30/5, 30/10, 30/15 = 6, 3, 2. In some step, we have to decide majority class, and then number of class value of left data is ‘Red’: 1, ‘Green’:1. Although, the number of two class value is same, but weight of ‘Red’ is 6, and ‘Green’ is 3, then the value of the dataset is ‘Red’.

To determine attribute, I used gain ratio.

* **Detailed description of code**
  + def check\_same\_class(dataset, Class\_dict):

this function check the data whether all of data is belong to same class

if all data is in same class return True, if not return False

* + def vote\_class(dataset, Class\_dict, power\_of\_clss\_value):

this function determine which class is most proper to represent the dataset

return class name

* + def calculate\_how\_many(dataset, attribute\_name, attribute\_value):

In this function, calculate number of data that belongs to each attribute value.

return dictionary that contains {attribute\_value: number\_of\_data}

* + def calculate\_I(dataset, Class(dict):

this function return calculated value of info. (discussed in previous page)

* + def calculate\_I\_attribute(dataset, attribute\_name, attribute\_value, Class\_dict):

this function return calculated value of info\_A. (discussed in previous page)

* + def splitinfo(dataset, attribute\_name, value\_list, Class\_dict):

this function return value of splitinfo. (discussed in previous page)

* + def attribute\_selection\_method(dataset, attribute\_set, Class\_dict):

this function return what attribute used in this step of generation decision tree

using some index (in my code, I used gain ratio), calculate what attribute is the best to classify dataset.

* + def generate\_decision\_tree(dataset, attribute\_dict, Class\_dict, power\_of\_class\_value):

In this function, decision tree is generated. Implement the algorithm discussed in first page.

+) Additionally, more detail in comments in code file.

* **Instructions for execution**

