CS5228: Assignment One

September 12, 2018

1 Introduction

In this assignment, you are going to implement the association rule mining algorithm and DBscan algorithm. The deadline for this project is 6pm on 1/Oct/2018.

2 Association Rule Mining

In this part, you are going to implement the association rule mining algorithm, we have provided the data, the test cases, the template output and template code to you. What you need to do in this part is to implement the generate_frequent_itemset(...) function and generate_association_rules(...) function in the template file. Please follow the readme.txt file in the "Association Rule Mining" folder to finish this task.

3 DBscan

In this part, you are going to implement the DBSCAN algorithm, we have provided the data, the test cases, the template output and template code to you. What you need to do in this part is to implement the dbscan(...) function in the template file. Please follow the readme.txt file in the "DBScan" folder to finish this task.

4 Library

It is okay for you to use some packages in python, you are okay to use some basic data structures and some basic functions which can help you implement your function. However, you are not allowed to use some too high level functions, like a function call to directly output the frequent itemsets to you or association rules or cluster labels to you. In this project, you are tested on implementing the frequent itemset generation, rule generation and dbscan algorithm. As these algorithms are already written in the lecture slides, however, in pesudo code, if you follow pesudo code to do your implementation, then you will do the right amount of work for this project.

5 Cooperation

It is okay for you to discuss implementation ideas with your classmates, however, you need to write the code yourself, plagiarism will not be tolerated in this module.

6 Error Reporting

If you spot any bugs for the template, and if the bug is minor, you are encouraged to modify it yourself and report in the forum to share it with your classmates.

7 Marking Criteria

Accuracy is our first marking criteria for this project. We will generate several test cases to test the accuracy of your code. In addition, we will also consider the efficiency of your code, if the running time of your code is too slow, corresponding marks will be deducted from you. Finally, if you write your own templates for this project, you need to make sure your templates satisfy the requirements described in the "Free Policy" section of the readme.txt file.

8 Submission Format

When you submit your code, please zip your assignment folder and name your folder in the following format "YourNameInIVLE_YourUserIDInIVLE.zip" (You can check the displayed information from the Class & Groups tab in the module CS5228 from IVLE). An example may be "XYZ ABCD_e1234567.zip". In addition, you need to make sure that if we unzip your folder, the "Association Rule Mining" and "DBScan" folders resides in the "XYZ ABCD_e1234567" folder. Please follow the submission format tightly, as we will apply this format to automatically mark your code. So if you fail to obey this format, we will not be able to mark your code and corresponding marks will be deducted from you.