TD Association Mining

You need to install firstly the package ***mlxtend*** under ***Python3.6.***
The package ***mlxtend*** doesn't work with the pre-installed *python3* in some systems, since it uses print(f " ") in the package. Check the version of the python3 before the installation.
You can find more information related to the package *mlxtend* from
http://rasbt.github.io/mlxtend/user_guide/frequent_patterns/association_rules/

Part I Examples

Write the comments to explain how the following programs work.

Example I.

```
import pandas as pd
from mlxtend.preprocessing import TransactionEncoder

dataset = [['f', 'a', 'c', 'd', 'g', 'i', 'm', 'p'],
           ['a', 'b', 'c', 'f', 'l', 'm', 'o'],
           ['b', 'f', 'h', 'j', 'o'],
           ['b', 'c', 'k', 's', 'p'],
           ['a', 'f', 'c', 'e', 'l', 'p', 'm', 'n']]

te = TransactionEncoder()
te_ary = te.fit(dataset).transform(dataset)
df = pd.DataFrame(te_ary, columns=te.columns_)
df

from mlxtend.frequent_patterns import apriori

#apriori(df, min_support=0.5, use_colnames=True)
frequent_itemsets = apriori(df, min_support=0.5, use_colnames=True)
frequent_itemsets['length'] = frequent_itemsets['itemsets'].apply(lambda x: len(x))
frequent_itemsets

frequent_itemsets[ (frequent_itemsets['length'] == 2) &
                   (frequent_itemsets['support'] >= 0.6) ]
frequent_itemsets[ frequent_itemsets['itemsets'] == {'f', 'm'} ]
frequent_itemsets[frequent_itemsets['itemsets'].str.contains('m', regex=False)]
```

Example II.

```
import pandas as pd
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules

Jogging_data=pd.read_csv('JoggingTitre.csv', sep=',')
x_dum=pd.get_dummies(Jogging_data)
x_dum

frequent_itemsets02=apriori(x_dum, min_support=0.2, use_colnames=True)
frequent_itemsets02
```

```
ARules02=association_rules(frequent_itemsets02, metric="confidence", min_threshold=0.8)
ARules02
ARules02.iloc[0:, 0:4]
ARules02.iloc[0:, [0, 1, 4, 5]]
ARules02[(ARules02['consequents'] == {'Jogging_Oui'})|(ARules02['consequents'] ==
{'Jogging_Non'})]
ARules02[(ARules02['consequents'] == {'Jogging_Oui'})].iloc[0:, [0, 1, 4, 5]]
```

Part II Exercises

Execute the following analyses and observations with at least 3 of the following data sets :

car.data,
flag.data,
agaricus-lepiota.data,
tic-tac-toe.data, and
zoo.data

It's better to analyze the same data sets used in the last TP in order to make some comparaisons

Exercise1

- Find the frequent patterns and association rules from the data set with *minSup*=10% and *minConf*=80%. You have to give your attention to the following points.
  a) It's not necessary to use all attributes in the data set, for example, the attributes as identifier should not be used in your analyses.
  b) You may need to do some preprocessing steps to preparer your data.
  c) You should use categorical data, at least ordinal data, in your analyses. You can try some kinds of discrimination to convert your numerical data to categorical data.
  d) The first line in the data set should be the attribute name for your data when you use *mlxtend* package.
- Mine the 3 most interesting phenomena from your association rule set.

Exercise2

- A decision tree can be transformed to a set of " If...Then..." rules. In this exercise, you are asked firstly to transform the highest accuracy decision tree to a such set of rules for each analyzed data set in last TP.
- Compare and observe the rules from the last TP and the association rules found in Exercise1.

Final Report

In your report, you have to return
- the programs with comments of Part I;
- Python programs with comments used for the exercises in Part II;
- descriptions of the interesting phenomena found in Exercise1 of Part II;
- observations from Exercise 2 of Part II