CSE17040 - Apriori

August 18, 2020

1 MLDM Lab 4 - Apriori

Manojkumar V K - CB.EN.U4CSE17040

[1]:
```
import pandas as pd
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules
```

[2]:
```
df = pd.read_csv('GroceryStoreDataSet.csv', sep=',',header=None,index_col=False)
df.head()
```

[2]:

	0
0	MILK,BREAD,BISCUIT
1	BREAD,MILK,BISCUIT,CORNFLAKES
2	BREAD,TEA,BOURNVITA
3	JAM,MAGGI,BREAD,MILK
4	MAGGI,TEA,BISCUIT

[3]:
```
items = []
for x in df[0]:
    items.append(x)
df = items
```

[4]:
```
data = []
for i in df:
    data.append(i.split(','))
data
```

[4]:
```
[['MILK', 'BREAD', 'BISCUIT'],
 ['BREAD', 'MILK', 'BISCUIT', 'CORNFLAKES'],
 ['BREAD', 'TEA', 'BOURNVITA'],
 ['JAM', 'MAGGI', 'BREAD', 'MILK'],
 ['MAGGI', 'TEA', 'BISCUIT'],
 ['BREAD', 'TEA', 'BOURNVITA'],
 ['MAGGI', 'TEA', 'CORNFLAKES'],
 ['MAGGI', 'BREAD', 'TEA', 'BISCUIT'],
```

```
['JAM', 'MAGGI', 'BREAD', 'TEA'],
['BREAD', 'MILK'],
['COFFEE', 'COCK', 'BISCUIT', 'CORNFLAKES'],
['COFFEE', 'COCK', 'BISCUIT', 'CORNFLAKES'],
['COFFEE', 'SUGER', 'BOURNVITA'],
['BREAD', 'COFFEE', 'COCK'],
['BREAD', 'SUGER', 'BISCUIT'],
['COFFEE', 'SUGER', 'CORNFLAKES'],
['BREAD', 'SUGER', 'BOURNVITA'],
['BREAD', 'COFFEE', 'SUGER'],
['BREAD', 'COFFEE', 'SUGER'],
['TEA', 'MILK', 'COFFEE', 'CORNFLAKES']]
```

[5]:
```
te = TransactionEncoder()
te_try = te.fit(data).transform(data)
```

[6]:
```
df = pd.DataFrame(te_try, columns=te.columns_)
df.head()
```

[6]:

	BISCUIT	BOURNVITA	BREAD	COCK	COFFEE	CORNFLAKES	JAM	MAGGI	MILK	SUGER	TEA
0	True	False	True	False	False	False	False	False	True	False	False
1	True	False	True	False	False	True	False	False	True	False	False
2	False	True	True	False	False	False	False	False	False	False	True
3	False	False	True	False	False	False	True	True	True	False	False
4	True	False	False	False	False	False	False	True	False	False	True

[20]:
```
frequent_itemsets = apriori(df, min_support=0.2, use_colnames=True)
frequent_itemsets['length'] = frequent_itemsets['itemsets'].apply(lambda x:␣
↪len(x))
frequent_itemsets
```

[20]:

	support	itemsets	length
0	0.35	(BISCUIT)	1
1	0.20	(BOURNVITA)	1
2	0.65	(BREAD)	1
3	0.40	(COFFEE)	1
4	0.30	(CORNFLAKES)	1
5	0.25	(MAGGI)	1
6	0.25	(MILK)	1
7	0.30	(SUGER)	1

```
8      0.35                (TEA)       1
9      0.20       (BISCUIT, BREAD)       2
10     0.20          (MILK, BREAD)       2
11     0.20         (SUGER, BREAD)       2
12     0.20           (TEA, BREAD)       2
13     0.20  (COFFEE, CORNFLAKES)       2
14     0.20        (SUGER, COFFEE)       2
15     0.20           (TEA, MAGGI)       2
```

[21]:
```
lift = association_rules(frequent_itemsets, metric="lift", min_threshold=1)
lift
```

[21]:
```
    antecedents   consequents  antecedent support  consequent support  \
0        (MILK)       (BREAD)                0.25                0.65
1       (BREAD)        (MILK)                0.65                0.25
2       (SUGER)       (BREAD)                0.30                0.65
3       (BREAD)       (SUGER)                0.65                0.30
4      (COFFEE)  (CORNFLAKES)                0.40                0.30
5  (CORNFLAKES)      (COFFEE)                0.30                0.40
6       (SUGER)      (COFFEE)                0.30                0.40
7      (COFFEE)       (SUGER)                0.40                0.30
8         (TEA)       (MAGGI)                0.35                0.25
9       (MAGGI)         (TEA)                0.25                0.35

   support  confidence      lift  leverage  conviction
0      0.2    0.800000  1.230769    0.0375    1.750000
1      0.2    0.307692  1.230769    0.0375    1.083333
2      0.2    0.666667  1.025641    0.0050    1.050000
3      0.2    0.307692  1.025641    0.0050    1.011111
4      0.2    0.500000  1.666667    0.0800    1.400000
5      0.2    0.666667  1.666667    0.0800    1.800000
6      0.2    0.666667  1.666667    0.0800    1.800000
7      0.2    0.500000  1.666667    0.0800    1.400000
8      0.2    0.571429  2.285714    0.1125    1.750000
9      0.2    0.800000  2.285714    0.1125    3.250000
```

[19]:
```
frequent_itemsets = apriori(df, min_support=0.2, use_colnames=True)
rules = association_rules(frequent_itemsets, metric="confidence",␣
 ↪min_threshold=0.5)
rules
```

[19]:
```
    antecedents   consequents  antecedent support  consequent support  \
0     (BISCUIT)       (BREAD)                0.35                0.65
1        (MILK)       (BREAD)                0.25                0.65
2       (SUGER)       (BREAD)                0.30                0.65
3         (TEA)       (BREAD)                0.35                0.65
4      (COFFEE)  (CORNFLAKES)                0.40                0.30
```

```
5  (CORNFLAKES)      (COFFEE)                0.30                0.40
6       (SUGER)      (COFFEE)                0.30                0.40
7      (COFFEE)       (SUGER)                0.40                0.30
8         (TEA)       (MAGGI)                0.35                0.25
9       (MAGGI)         (TEA)                0.25                0.35

   support  confidence      lift  leverage  conviction
0      0.2    0.571429  0.879121   -0.0275    0.816667
1      0.2    0.800000  1.230769    0.0375    1.750000
2      0.2    0.666667  1.025641    0.0050    1.050000
3      0.2    0.571429  0.879121   -0.0275    0.816667
4      0.2    0.500000  1.666667    0.0800    1.400000
5      0.2    0.666667  1.666667    0.0800    1.800000
6      0.2    0.666667  1.666667    0.0800    1.800000
7      0.2    0.500000  1.666667    0.0800    1.400000
8      0.2    0.571429  2.285714    0.1125    1.750000
9      0.2    0.800000  2.285714    0.1125    3.250000
```

[ ]: