CSE17040 - FPGrowth

August 25, 2020

1 MLDM Lab 4 - FPGrowth

Manojkumar V K - CB.EN.U4CSE17040

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

1.1 Reading into dataframe

[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

1.2 Data preprocessing

[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']]
```

1.3 Using Transaction Encoder to encode

[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

1.4 Association Rule Mining - FP Growth

[7]:
```
frequent_itemsets = fpgrowth(df, min_support=0.2, use_colnames=True)
frequent_itemsets['length'] = frequent_itemsets['itemsets'].apply(lambda x:␣
 ↪len(x))
frequent_itemsets
```

[7]:

	support	itemsets	length
0	0.65	(BREAD)	1
1	0.35	(BISCUIT)	1
2	0.25	(MILK)	1
3	0.30	(CORNFLAKES)	1
4	0.35	(TEA)	1
5	0.20	(BOURNVITA)	1
6	0.25	(MAGGI)	1
7	0.40	(COFFEE)	1
8	0.30	(SUGER)	1
9	0.20	(BREAD, BISCUIT)	2
10	0.20	(MILK, BREAD)	2
11	0.20	(COFFEE, CORNFLAKES)	2
12	0.20	(TEA, BREAD)	2
13	0.20	(TEA, MAGGI)	2
14	0.20	(COFFEE, SUGER)	2
15	0.20	(BREAD, SUGER)	2

[8]:
```
lift = association_rules(frequent_itemsets, metric="lift", min_threshold=1)
lift
```

[8]:

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

	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	0.500000	1.666667	0.0800	1.400000
3	0.2	0.666667	1.666667	0.0800	1.800000
4	0.2	0.571429	2.285714	0.1125	1.750000
5	0.2	0.800000	2.285714	0.1125	3.250000

```
6      0.2    0.500000  1.666667    0.0800    1.400000
7      0.2    0.666667  1.666667    0.0800    1.800000
8      0.2    0.307692  1.025641    0.0050    1.011111
9      0.2    0.666667  1.025641    0.0050    1.050000
```

[9]:
```
frequent_itemsets = fpgrowth(df, min_support=0.2, use_colnames=True)
rules = association_rules(frequent_itemsets, metric="confidence",␣
 ↪min_threshold=0.5)
rules
```

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

   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    0.500000  1.666667    0.0800    1.400000
3      0.2    0.666667  1.666667    0.0800    1.800000
4      0.2    0.571429  0.879121   -0.0275    0.816667
5      0.2    0.571429  2.285714    0.1125    1.750000
6      0.2    0.800000  2.285714    0.1125    3.250000
7      0.2    0.500000  1.666667    0.0800    1.400000
8      0.2    0.666667  1.666667    0.0800    1.800000
9      0.2    0.666667  1.025641    0.0050    1.050000
```

[ ]: