

CSE17040 - Apriori

August 25, 2020

1 MLDM Lab 5 - Apriori

(Without libraries)

Manojkumar V K - CB.EN.U4CSE17040

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

1.1 Reading into dataframe

```
[2]: df = pd.read_csv('zoo.csv')
      df.head()
```

```
[2]: animal_name  hair  feathers  eggs  milk  airborne  aquatic  predator  \
0    aardvark     1         0     0     1         0         0         1
1    antelope     1         0     0     1         0         0         0
2         bass     0         0     1     0         0         1         1
3         bear     1         0     0     1         0         0         1
4         boar     1         0     0     1         0         0         1

      toothed  backbone  breathes  venomous  fins  legs  tail  domestic  catsize  \
0          1          1          1          0     0     4     0          0          1
1          1          1          1          0     0     4     1          0          1
2          1          1          0          0     1     0     1          0          0
3          1          1          1          0     0     4     0          0          1
4          1          1          1          0     0     4     1          0          1

      class_type
0              1
1              1
2              4
3              1
4              1
```

1.2 Data preprocessing

```
[3]: df = df.set_index('animal_name')
df = df.drop(['class_type'],axis=1)
df.head()
```

```
[3]:
```

	hair	feathers	eggs	milk	airborne	aquatic	predator	toothed	\
animal_name									
aardvark	1	0	0	1	0	0	1	1	
antelope	1	0	0	1	0	0	0	1	
bass	0	0	1	0	0	1	1	1	
bear	1	0	0	1	0	0	1	1	
boar	1	0	0	1	0	0	1	1	

	backbone	breathes	venomous	fins	legs	tail	domestic	catsize	
animal_name									
aardvark	1	1	0	0	4	0	0	1	
antelope	1	1	0	0	4	1	0	1	
bass	1	0	0	1	0	1	0	0	
bear	1	1	0	0	4	0	0	1	
boar	1	1	0	0	4	1	0	1	

1.3 Encoding

Instead of using Transaction Encoder, we directly apply a simple logic to encode the dataframe. This is because the given dataset is already encoded.

```
[4]: df = df > 0
df.head()
```

```
[4]:
```

	hair	feathers	eggs	milk	airborne	aquatic	predator	\
animal_name								
aardvark	True	False	False	True	False	False	True	
antelope	True	False	False	True	False	False	False	
bass	False	False	True	False	False	True	True	
bear	True	False	False	True	False	False	True	
boar	True	False	False	True	False	False	True	

	toothed	backbone	breathes	venomous	fins	legs	tail	\
animal_name								
aardvark	True	True	True	False	False	True	False	
antelope	True	True	True	False	False	True	True	
bass	True	True	False	False	True	False	True	
bear	True	True	True	False	False	True	False	
boar	True	True	True	False	False	True	True	

	domestic	catsize
animal_name		
aardvark	False	True
antelope	False	True
bass	False	False
bear	False	True
boar	False	True

1.4 Association Rule Mining - Apriori

```
[5]: frequent_itemsets = apriori(df, min_support=0.2, use_colnames=True)
frequent_itemsets['length'] = frequent_itemsets['itemsets'].apply(lambda x:
↪len(x))
frequent_itemsets[:5]
```

```
[5]:
```

	support	itemsets	length
0	0.425743	(hair)	1
1	0.584158	(eggs)	1
2	0.405941	(milk)	1
3	0.237624	(airborne)	1
4	0.356436	(aquatic)	1

```
[6]: lift = association_rules(frequent_itemsets, metric="lift", min_threshold=1)
lift[:5]
```

```
[6]:
```

	antecedents	consequents	antecedent support	consequent support	support \
0	(milk)	(hair)	0.405941	0.425743	0.386139
1	(hair)	(milk)	0.425743	0.405941	0.386139
2	(toothed)	(hair)	0.603960	0.425743	0.376238
3	(hair)	(toothed)	0.425743	0.603960	0.376238
4	(hair)	(backbone)	0.425743	0.821782	0.386139

	confidence	lift	leverage	conviction
0	0.951220	2.23426	0.213312	11.772277
1	0.906977	2.23426	0.213312	6.386139
2	0.622951	1.46321	0.119106	1.523031
3	0.883721	1.46321	0.119106	3.405941
4	0.906977	1.10367	0.036271	1.915842

```
[7]: frequent_itemsets = apriori(df, min_support=0.2, use_colnames=True)
rules = association_rules(frequent_itemsets, metric="confidence",
↪min_threshold=0.5)
rules[:5]
```

```

[7]: antecedents consequents antecedent support consequent support support \
0      (milk)      (hair)      0.405941      0.425743 0.386139
1      (hair)      (milk)      0.425743      0.405941 0.386139
2  (toothed)      (hair)      0.603960      0.425743 0.376238
3      (hair)  (toothed)      0.425743      0.603960 0.376238
4      (hair)  (backbone)      0.425743      0.821782 0.386139

      confidence      lift  leverage conviction
0      0.951220  2.23426  0.213312   11.772277
1      0.906977  2.23426  0.213312    6.386139
2      0.622951  1.46321  0.119106    1.523031
3      0.883721  1.46321  0.119106    3.405941
4      0.906977  1.10367  0.036271    1.915842

```

```

[ ]:

```