Association Rules

Example

Association rules範例- weather

- 載入資料集(以weather.csv為例)

```
In [20]: import pandas as pd
         df = pd.read_csv('weather.csv')
         df
```

Out[20]:

	outlook	temperature	humidity	windy	play
0	sunny	hot	high	False	no
1	sunny	hot	high	True	no
2	overcast	hot	high	False	yes
3	rainy	mild	high	False	yes
4	rainy	cool	normal	False	yes
5	rainy	cool	normal	True	no
6	overcast	cool	normal	True	yes
7	sunny	mild	high	False	no
8	sunny	cool	normal	False	yes
9	rainy	mild	normal	False	yes
10	sunny	mild	normal	True	yes
11	overcast	mild	high	True	yes
12	overcast	hot	normal	False	yes
13	rainy	mild	high	True	no

Association rules範例- weather

- 轉換資料型態

欄位中有boolean，先轉換成string

轉成list格式

```
In [29]: df = df.astype(str)
         data = df.values.tolist()
```

Association rules範例- weather

- 安裝apyori

```
~/ECT_hw pip3 install apyori
```

```
Collecting apyori
  Downloading https://files.pythonhosted.org/packages/5e/62/5ffde5c473ea4b033490617ec5caa80d59
804875ad3c3c57c0976533a21a/apyori-1.1.2.tar.gz
Building wheels for collected packages: apyori
  Building wheel for apyori (setup.py) ... done
  Stored in directory: /Users/daisy/Library/Caches/pip/wheels/5d/92/bb/474bbadbc8c0062b9eb168f
69982a0443263f8ab1711a8cad0
Successfully built apyori
Installing collected packages: apyori
Successfully installed apyori-1.1.2
```

Association rules範例- weather

- 建立關聯式規則

In [22]:
```
from apyori import apriori

#建立rule, 設定參數
#變成list
rules = list(apriori(data, min_support= 0.2, min_confidence= 0.9))
rules
```

Out[22]:
```
[RelationRecord(items=frozenset({'normal', 'cool'}), support=0.2857142857142857, ordered_statistics=[OrderedStatist
ic(items_base=frozenset({'cool'}), items_add=frozenset({'normal'}), confidence=1.0, lift=2.0)]),
 RelationRecord(items=frozenset({'yes', 'overcast'}), support=0.2857142857142857, ordered_statistics=[OrderedStatis
tic(items_base=frozenset({'overcast'}), items_add=frozenset({'yes'}), confidence=1.0, lift=1.5555555555555554)]),
 RelationRecord(items=frozenset({'normal', 'False', 'yes'}), support=0.2857142857142857, ordered_statistics=[Ordere
dStatistic(items_base=frozenset({'normal', 'False'}), items_add=frozenset({'yes'}), confidence=1.0, lift=1.55555555
55555554)]),
 RelationRecord(items=frozenset({'False', 'yes', 'rainy'}), support=0.21428571428571427, ordered_statistics=[Ordere
dStatistic(items_base=frozenset({'False', 'rainy'}), items_add=frozenset({'yes'}), confidence=1.0, lift=1.555555555
5555554), OrderedStatistic(items_base=frozenset({'yes', 'rainy'}), items_add=frozenset({'False'}), confidence=1.0,
lift=1.75)]),
 RelationRecord(items=frozenset({'normal', 'cool', 'yes'}), support=0.21428571428571427, ordered_statistics=[Ordere
dStatistic(items_base=frozenset({'cool', 'yes'}), items_add=frozenset({'normal'}), confidence=1.0, lift=2.0)]),
 RelationRecord(items=frozenset({'high', 'no', 'sunny'}), support=0.21428571428571427, ordered_statistics=[OrderedS
tatistic(items_base=frozenset({'high', 'sunny'}), items_add=frozenset({'no'}), confidence=1.0, lift=2.8), OrderedSt
atistic(items_base=frozenset({'no', 'sunny'}), items_add=frozenset({'high'}), confidence=1.0, lift=2.0)])]
```

Association rules範例- weather

- 抓出每一條rule的 rule/support/confidence，整理成 dataframe

In [23]:
```
result = pd.DataFrame()
for item in rules:
    series = pd.Series({"Rule":item[0],"Support":item[1],"Confidence":item[2][0][2]})
    result = result.append(series, ignore_index=True)
```

Association rules範例- weather

- 依confidence降冪排列

```
In [24]: result.sort_values(by= ['Confidence'], ascending=False)
```

Out[24]:

	Confidence	Rule	Support
0	1.0	(normal, cool)	0.285714
1	1.0	(yes, overcast)	0.285714
2	1.0	(normal, False, yes)	0.285714
3	1.0	(False, yes, rainy)	0.214286
4	1.0	(normal, cool, yes)	0.214286
5	1.0	(high, no, sunny)	0.214286