

1) Read Chapter 6 (only sections 6.1 and 6.7).

2) Do Chapter 6 textbook problem #2 (parts a,b,c,d only) on page 404.

$$\textcircled{a}. \quad s(\{e\}) = 8/10 = 0.8$$

$$s(\{b,d\}) = 2/10 = 0.2$$

$$s(\{b,d,e\}) = 2/10 = 0.2$$

$$\textcircled{b}. \quad c(bd \rightarrow e) = \frac{s(\{b,d,e\})}{s(\{b,d\})} = \frac{0.2}{0.2} = 100\%$$

$$c(e \rightarrow bd) = \frac{s(\{b,d,e\})}{s(\{e\})} = \frac{0.2}{0.8} = 25\%$$

$$\textcircled{c}. \quad s(\{e\}) = 4/5 = 0.8$$

$$s(\{b,d\}) = 5/5 = 1$$

$$s(\{b,d,e\}) = 4/5 = 0.8$$

$$\textcircled{d}. \quad c(bd \rightarrow e) = \frac{s(\{b,d,e\})}{s(\{b,d\})} = \frac{0.8}{1} = 80\%$$

$$c(e \rightarrow bd) = \frac{s(\{b,d,e\})}{s(\{e\})} = \frac{0.8}{0.8} = 100\%$$

3) Do Chapter 6 textbook problem #6 (parts d,e only) on page 406.

find an itemset (of size 2 or larger) that has the largest support.

sets with one item

Itemset	support
Diapers	7
Milk	5
Bread	5
Butter	5
Beer	4
Cookies	4

sets with two items

Itemset	support
{Diapers, Milk}	4
{Diapers, Bread}	3
{Diapers, Butter}	3
{Diapers, Beer}	3
{Diapers, Cookies}	2
{Milk, Bread}	3
{Milk, Butter}	2
{Milk, Beer}	1
{Milk, Cookies}	1
{Bread, Butter}	5
{Bread, Beer}	0
{Bread, Cookies}	1
{Butter, Bread}	0
{Butter, Cookies}	1
{Beer, Cookies}	2

* We can consider the support of 2-items itemsets, since adding items will never grow the support.
 $\therefore \{ \text{Bread, Butter} \}$ is largest support

* Find a pair of items a, b such that rules $a \rightarrow b$ and $b \rightarrow a$ have same Confidence

$$\text{Confidence}(a \rightarrow b) = \frac{\text{support}(\{a, b\})}{\text{support}(\{a\})}$$

$$\text{Confidence}(\{ \text{Bread, Butter} \}) = \frac{\text{support}(\{ \text{Bread, Butter} \})}{\text{support}(\{ \text{Bread} \})} = 5/5 = 1$$

$$c(\text{Butter} \rightarrow \text{Bread}) = 5/5 = 1$$

||| $\{ \text{milk, Bread} \}, \{ \text{milk, Butter} \}, \{ \text{Beer, Cookies} \}$ have same Confidence when reversed.

4) Using the data at www.stats202.com/more_stats202_logs.txt and treating each row as a "market basket" compute the support and confidence for the rule $ip=65.57.245.11 \rightarrow \text{"Mozilla/5.0 (X11; U; Linux i686 (x86_64); en-US; rv:1.8.1.3) Gecko/20070309 Firefox/2.0.0.3"}$.

State what the support and confidence values mean in plain English in this context.

Support for the above rule : Transactions containing all the items in the above rule
/Total number of logs

Confidence for the above rule: Support (entire rule) / Support (IP address)

```
$ python stats.py
Count of A: 3636
Count of B: 234
Count of AB: 1385
Total Transactions: 14809
-----
Support of A: 0.2455263690998717
Support of B: 0.015801201971773923
Support of AB: 0.09352420825173881
Confidence of AB: 0.3809130913091309
```

Support : The number of transactions that include the items in the (X) and (Y) part of the rule as a percentage of the total number of transactions. It is a measure of how frequently the collection of items occur together as a percentage of all transactions.

$$X \rightarrow Y$$

Support =
$$\frac{\text{Number of transactions containing all the items in X and Y}}{\text{Total number of transactions}}$$

Here X refers to the IP address and Y refers to browser information

Confidence : It is the ratio of the no of transactions that includes all items in {B} as well as the no of transactions that includes all items in {A} to the no of transactions that includes all items in {A}.

$$A \rightarrow B$$

Confidence =
$$\text{Support} (\{A, B\}) / \text{Support} (\{A\})$$

Here A refers to the IP address and B refers to browser information