RMIT University
School of Science
COSC2110/COSC2111 Data Mining
Tutorial Problems Week 5

1. Consider the following supermarket transactions

TID	Items
T1	Milk, Bread, Coffee
T2	Bread, Tea
T3	Bread, Eggs
T4	Milk ,Bread, Tea
T5	Milk, Eggs
T6	Bread, Eggs
T7	Milk, Eggs
T8	Milk, Bread, Eggs, Coffee
T9	Milk, Bread, Eggs

   (a) Find the frequent item sets for which the minimum support is 2.

   (b) Generate association rules from the 3-item set {Milk,Bread,Eggs}, giving the support and confidence of each rule.

   (c) If the minimum confidence threshold is 70% which rules would be output?

2. Given the transaction database below:

TID	Items bought
100	a , b, c, d, e, f
200	a, c, e, f, g
300	a, d, e
400	b, c, d, f, h
500	a, c, e, f, h

   (a) Suppose that the support threshold is 40% (or count 2), list the frequent 2-itemsets and their support.

   (b) Compute the support and confidence of the following association rules:
   R1: a, b, c ==> d, e
   R2: a, d ==> f

   (c) Given the support threshold of 40%, and the confidence threshold of 50%, use the Apriori algorithm to generate association rules.

3. Consider the following data:

	Age	Prescription	Astigmatic	Tears	Lenses
1	child	myope	no	reduced	NO
2	child	myope	yes	normal	HARD
3	child	hypermetrope	no	normal	SOFT
4	child	hypermetrope	yes	reduced	NO
5	child	myope	no	normal	SOFT
6	adult	myope	yes	reduced	NO
7	adult	hypermetrope	yes	reduced	NO
8	adult	hypermetrope	yes	normal	NO
9	elderly	myope	no	normal	NO
10	elderly	myope	yes	normal	HARD
11	elderly	hypermetrope	no	reduced	NO
12	elderly	hypermetrope	no	normal	SOFT

   (a) Find 2-item sets that have a minimum support of 2.

   (b) Generate association rules from your first 2-item set that have a confidence of 100%.

4. Consider the following rules produced by the Apriori program from the above data.

```
R1:Tears=reduced 5 ==> lenses=NO 5
      <conf:(1)> lift:(1.71) lev:(0.17) [2] conv:(2.08)

R2:Lenses=SOFT 3 ==> Astigmatic=no 3
      <conf:(1)> lift:(2) lev:(0.13) [1] conv:(1.5)

R3:Prescription=hypermetrope Astigmatic=no Tears=reduced 1 ==> lenses=NO 1
      <conf:(1)> lift:(1.71) lev:(0.03) [0] conv:(0.42)
```

   (a) Which rule is the best, and why?

5. Consider the iris data. All the attributes are numeric. What would be result of applying the apriori algorithm directly to this kind of data. Is there a way to get association rules from this kind of data?