Lab 9 Solutions

Problem 1. Consider the following transactional database.

TID	Items
100	1, 2, 3, 5, 6, 8
200	1, 2, 4, 5, 6
300	1, 2, 3, 4, 5, 6, 8
400	1, 2, 3, 5, 7
500	1, 2, 3, 8

(1) Mine all frequent itemsets using Apriori. Show all candidate itemsets and frequent itemsets You should follow the process described in the lecture (i.e., C1 → L1 → C2 → L2 → …). Minimum support = 60% (or 3 or more transactions). Note that you don't need to show the pruning steps. To save your time, L1 and L2 are given below:
L1:

Itemset	1	2	3	5	6	8
Count	5	5	4	4	3	3

L2:

Itemset	{1,2}	{1,3}	{1,5}	{1,6}	{1,8}	{2,3}	{2,5}	{2,6}	{2,8}	{3,5}	{3,8}	{5,6}
Count	5	4	4	3	3	4	4	3	3	3	3	3

Problem 2. Consider the following contingency table.

	C (buys coffee = Yes)	$\overline{C}$ (buys coffee = No)
T (buys tea = Yes)	362	823
$\overline{T}$ (buys tea = No)	49	1527

Compute the *lift*{T->C} and determine whether buying coffee and buying tea are positively correlated, negatively correlated, or not correlated.

Problem 3. Explore the Groceries dataset from the lecture and rules generated from Apriori algorithm below. Create R code for the following questions a-c:

```
rules <- apriori(Groceries,
                 parameter = list(support = 0.006, confidence = 0.25, minlen = 2))
```

a. Create a frequency plot of your interest.
b. Generate a subset of rules of which the right hand side is "tropical fruit".
c. Generate a subset of rules of which the items contain either “berries” or “yogurt” and lift is bigger than 3.

(Hint: look at the R document of the subset method)