Niveditha Chandrashekar Asha
nc765@scarletmail.rutgers.edu

Homework 2

(5 pts) The following contingency table summarizes supermarket transaction data, where hotdogs refer to the transactions containing hot dogs, \bar{hotdogs} refers to the transactions that do not contain hot dogs, hamburgers refer to the transactions containing hamburgers, and \bar{hamburgers} refers to the transactions that do not contain hamburgers.

	$hotdogs$	$\overline{hotdogs}$	$\sum_{row}$
$hamburgers$	2000	500	2500
$\overline{hamburgers}$	1000	1500	2500
$\sum_{col}$	3000	2000	5000

Suppose the association rule (*hotdogs => hamburgers*) is mined. Given a minimum support threshold of 25% and a minimum confidence threshold of 50%, is this association rule strong? Why or why not.

Support Threshold = 25%
Confidence Threshold = 50%

Support = together / total = (2000/5000) *100 = 40%
Confidence(X->Y) = together/X = (2000/3000) *100 = 66.66%

Yes, this Association rule is strong.
An **association rule** having support and confidence greater than or equal to a user-specified minimum support threshold and respectively a minimum confidence threshold

Niveditha Chandrashekar Asha
nc765@scarletmail.rutgers.edu

Q2) (5 pts) Satellite radio customers. An analyst at a subscription-based satellite radio company has been given a sample of data from their customer database with the goal of finding groups of customers that are associated with one another. The data consist of company data, together with purchased demographic data that are mapped to the company data – see the figure below. The analyst decides to apply Association Rules to learn more about the associations between customers. Comment on this approach.

When you analyze the Satellite radio customer data set, the analyst's decision to apply association rules may not be the best approach to identify and understand the groups of customers who are associated with one another. Association rules do not consider sequential information; in this customer data set, the association between the items in the columns can be determined, but not among the rows. Therefore, the better way to go about this would be to use Cluster analysis

The analyst should use clustering analysis as it shares similar characteristics are grouped into one cluster described maximum relation between them.

QUESTION 3

1) **(2 pts) Select two values in the matrix above and explain their meaning.**

In all the 12 transactions shown above **bags** are purchased only in the 10th transaction

Blush is purchased in transaction number 1, 3 5 ,7 ,10

Nail polish is purchased in transaction number 1, 2, 4 ,7 ,8,10,11,12

Niveditha Chandrashekar Asha
nc765@scarletmail.rutgers.edu

Rule #	Conf. %	Antecedent (a)	Consequent (c)	Support(a)	Support (c)	Support (a U c)	Lift Ratio
2	60.19	Bronzer, Nail Polish=>	Brushes, Concealer	103	77	62	3.909
1	80.52	Brushes, Concealer=>	Bronzer, Nail Polish	77	103	62	3.909
4	56.36	Brushes=>	Bronzer, Concealer, Nail Polish	110	76	62	3.708
3	81.58	Bronzer, Concealer, Nail P	Brushes	76	110	62	3.708
6	76.36	Brushes=>	Bronzer, Nail Polish	110	103	84	3.707
5	81.55	Bronzer, Nail Polish=>	Brushes	103	110	84	3.707
8	56.88	Concealer, Nail Polish=>	Bronzer, Brushes	109	84	62	3.386
7	73.81	Bronzer, Brushes=>	Concealer, Nail Polish	84	109	62	3.386
10	70	Brushes=>	Concealer, Nail Polish	110	109	77	3.211
9	70.64	Concealer, Nail Polish=>	Brushes	109	110	77	3.211
12	50	Brushes=>	Blush, Nail Polish	110	82	55	3.049
11	67.07	Blush, Nail Polish=>	Brushes	82	110	55	3.049

2) (2 pts) For the first row, explain the "Conf. %" output and how it is calculated.

The first row has Rule#2, the confidence is 60.19%, The confidence of a rule says that % of antecedent (IF) transactions that also have the consequent (THEN) item set.
Therefore by interpreting this value we can say that 60.19% times we say that Brushes and Concealers appear along with Bronzer and Nail polish whenever Bronzer and Nail polish is purchased.

Conf.% = {Brushes, concealer, Bronzer, Nail polish} / {Bronzer, Nail polish} = 62/103 = > 0.6019

60.19

3) (3 pts) For the first row, explain the rule that is represented there in words.

Rule =>
{Bronzer, Nail Polish} = > {Brushes, Concealer}
This rule is about the purchase of Brushes and concealer with Bronzer and Nail polish

Niveditha Chandrashekar Asha
nc765@scarletmail.rutgers.edu

Support (a) = > Support {Bronzer, Nail polish} = Frequency of purchase of **{Bronzer, Nail Polish}**

Support(c) => Support **{Brushes, Concealer} =>** Frequency of purchase of Brushes and Concealer

Support(aUc) => **{Bronzer, Nail Polish, Brushes, Concealer} =>** Frequency of purchase of Bronzer, Nail Polish, Brushes, Concealer

The confidence is 60.19%, The confidence of a rule says that % of antecedent (IF) transactions that also have the consequent (THEN) item set.
Therefore by interpreting this value we can say that 60.19% times we say that Brushes and Concealers appear along with Bronzer and Nail polish whenever Bronzer and Nail polish is purchased.

4) (3 pts) Out of all the rules shown, which would you choose to act on first, and why?

I would choose to act on Rule 1 first for the following reasons: -

1. 1) High lift ratio of 3.909
2. 2) Confidence % = Support (a U c) / Support (a)

   = 62 / 77 = 80.52% Therefore, High confidence %

the highest lift ratio indicates that if a person buys Brushes and Concealer, he is more likely to also buy Bronzer and Nail Polish; Rule 1.