# Data Mining Assignment 3

# Q2)

(a) Compute the support for item sets {e}, {b, d}, and {b, d, e} by treating each transactionID as a market basket. 10 distinct baskets/transactions

Ans:

if {e}: s =

Then 8/10 = 0.8

{b, d}: s = 2/10 = 0.2

{b, d, e}: s = 2/10 = 0.2

(b) Use the results in part (a) to compute the confidence for the association rules {b, d} ---> {e} and {e} ---> {b, d}. Is confidence a symmetric measure?

Ans):Both rules have support 0:2,

(support count is 2):

{b, d} ---> {e}: c = 0.2/0.2 = 1 {e} ---> {b, d}:

c = 0.2/0.8 = 0.25 Support is a symmetric measure, but confidence is not symmetric!

(c) Repeat part (a) by treating each customer ID as a market basket. Each item should be treated as a binary variable (1 if an item appears in at Least one transaction bought by the customer, and 0 otherwise.)

Now we have 5 baskets in total. • {e}: s = 4/5 = 0.8 • {b, d}: s = 5/5 = 1 • {b, d, e}: s = 4/5 = 0.8

(d) Use the results in part (c) to compute the confidence for the association rules {b, d} ---> {e} and {e} ---> {b, d}. • {b, d} ---> {e}: c = 0.8/1 = 0.8 • {e} ---> {b, d}: c = 0.8/0.8 = 1

Q3) d) Find an itemset (of size 2 or larger) that has the largest support.

Put here the table which is in the sol4 pdf Ignoring the 1-itemsets (and ø), the itemset with the largest support is {bread, butter}.

e) Find a pair of items, a and b, such that the rules {a} −-> {b} and {b} −-> {a} have the same confidence. Bread and butter have the same support (s = 5). This means that the rules {bread} ◊ {butter} and {butter} ◊ {bread} have the same confidence (c = 5/5 = 1). The same can be said with beer and cookies (s = 4, c = 2/4 = 0.5)

# 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 → "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.?

# Ans: The rule for which we have to find the support and confidence is {65.57.245.11} -> {“Mozilla/5.0 (X11; U; Linux i686 (x86\_64); en-US; rv:1.8.1.3) Gecko/20070309 Firefox/2.0.0.3"} Support for {65.57.245.11} = 5021 / 14803 = 0.33 Support for {“Mozilla/5.0 (X11; U; Linux i686 (x86\_64); en-US; rv:1.8.1.3) Gecko/20070309 Firefox/2.0.0.3"} = 1619/14803 = 0.109 Confidence for rule {65.57.245.11} -> {“Mozilla/5.0 (X11; U; Linux i686 (x86\_64); en-US; rv:1.8.1.3) Gecko/20070309 Firefox/2.0.0.3"} = support count ({65.57.245.11, “Mozilla/5.0 (X11; U; Linux i686 (x86\_64); en-US; rv:1.8.1.3) Gecko/20070309 Firefox/2.0.0.3"}) / support count ({65.57.245.11}) = 1619 / 5021 = 0.322