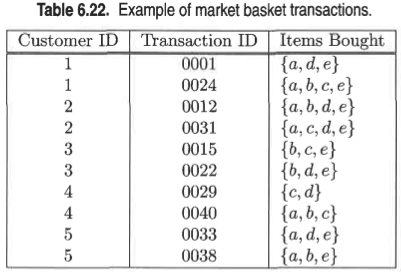
1) Read Chapter 6 (only sections 6.1 and 6.7).

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

2. Consider the data set shown in Table 6.22.



1. Compute the support for itemsets {e}, {b,d}, and {b,d,e} by treating each transaction ID as a market basket.

Ans.

s({e}) = 8/10 = 0.8

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

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

1. 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.

c(b,d → e) = 0.2/0.2

c(e → b,d) = 0.2/0.8

No, confidence is not a symmetric measure.

1. 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.)

Ans.

s({e}) = 4/5 = 0.8

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

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

1. Use the results in part (c) to compute the confidence for the association rules {b, d} → {e} and {e} → {b,d}.

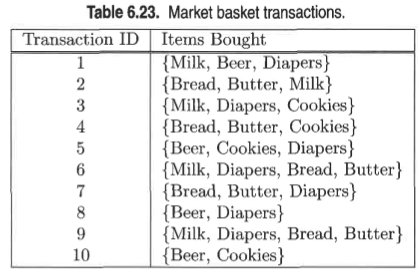
Ans.

c(b,d → e) = 0.8/1.0

c(e → b,d) = 0.8/0.8

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

6. Consider the market basket transactions shown in Table 6.23.



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

Ans. {bread, butter} with support 5/10.

(e) Find a pair of items, a and b, such that the rules, {a} → {b} and {b} → {a} have the same confidence.

Ans.

Beer → cookies, confidence = 2/4

Or, bread → butter, confidence = 5/5

4) Using the data at [www.stats202.com/more\_stats202\_logs.txt](http://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 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({65.57.245.11}) = 5021 / 14803 = 0.33

Support({“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 (rule) / support ({65.57.245.11})

= 1619/14803 \* 14803/5021

= 1619 / 5021

= 0.322