



UNITE Distributed Learning
University of Minnesota-Twin Cities

Martin
Last Name (please print)

CS5523
UofMN Course Number#

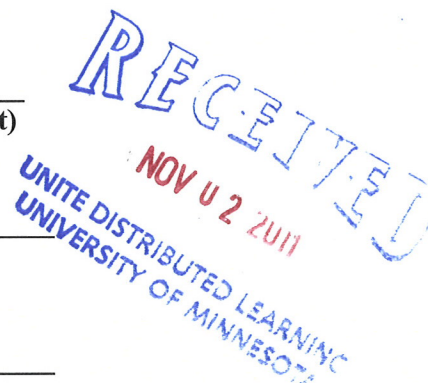
Kumar
Instructor

E-MAIL mart0124

Robert
First Name (please print)

HW 3
Assignment/lab #

11/2/2011
Date Turned in



UNITE course work may be faxed to UNITE Distributed Learning, University of Minnesota-Twin Cities, 612-626-0761. UNITE prefers email submission of homework.

If fax'ed: Confirmed Receipt: _____ (check here)

Total number pages, **INCLUDING** this cover sheet: 5

In case of transmittal problems call: 612.578.6641

Robert F.K. Martin
ID 1505151
Csci 5302
02 Nov 2006
Homework 3

Problem 1a:

The max number of rules can be found using the formula $R = 3^d - 2^{d+1} + 1$, where $d = 7$, the number of items. $R=1932$.

Problem 1b:

Since the max size itemset is 4, then the max size frequent itemset is also 4.

Problem 1c:

There are 6 distinct size-3 itemsets and 2 distinct size-4 itemsets. Then, the number of possible size-3 itemsets, ignoring what the attributes are of each itemset, $\binom{4}{3} + \binom{4}{3} + 6 = 14$.

Problem 1d:

Bread and butter appear 5 times together.

Problem 1e:

Confidence is the support of the consequent and antecedent divided by the support of the antecedent. In order for two size-2 rules to have the same confidence, each of the size-1 itemsets must have the same support. So we just need to find two size-1 itemsets with the same support. This can be achieved with either {bread,butter} or {beer,cookies}.

Problem 2a:

{a,b,c,d}, {c,d,e,f},
{a,b,c,e},
{a,b,c,f},
{a,b,d,e},
{a,b,d,f},
{a,b,e,f},
{a,c,d,e},
{a,c,d,f},
{a,d,e,f}

Problem 2b:

{a,b,c,d},
{a,b,c,e},
{a,b,d,e}

Problem 2c:

∅. Each of the above size-4 itemsets are composed of at least one size-3 itemset that is not frequent.

Problem 2d:

There are no size-5 itemsets that will survive pruning since there are no size-4 itemsets that survived pruning.

Problem 3a:

Min: 0, achieved when $P(B|A)=P(B)$.

Max: ∞ , achieved when $P(B)=1$.

Problem 3b:

It increases.

Problem 3c:

It decreases.

Problem 3d:

It decreases.

Problem 3e:

no.

Problem 3f:

0

Problem 3g:

no

Problem 3h:

no

Problem 4a:

$\alpha_1 = 20, \alpha_2 = 29$ ✓

Problem 4b:

$\alpha_1 = 18, \alpha_2 = 22$ -2

Problem 4c:

$\alpha_1 = 5, \alpha_2 = 7$ ✓

Problem 4d:

The three different discretization choices produced three different results with very little overlap between the first two. The point being that the choice of intervals can have significant impact on the final results. ✓ 1

Problem 5a:

We would like symmetry, so that if user 1 is like user 2, we would also like user 2 to be like user 1. Additionally, we would like inversion to hold since if two users are similar in their likes, we would

also like them to be similar in their dislikes. But since we have designated 0 to mean dislike, we do not want null addition to hold since we do care about similar dislikes between two users.

Problem 5b:

We should choose correlation because it is symmetric and inversion invariant. Confidence is not symmetric and cosine is not inversion invariant.

Problem 5c:

In this case, we do want null addition to hold but not inversion. So we should choose the cosine measure since it is still symmetric but is null addition invariant but not invariant to inversion.

Problem 6a:

Set S and T may be equal under the condition that h_c is no more than the minimum confidence of any itemset in S. If h_c is higher, then T is a subset of S.

Problem 6b:

T is a subset of S since S will include all itemsets with support $\geq b$ plus those that are also $\geq a$.

Problem 6c:

S will be a subset of T if S contains any itemset with confidence $\geq h_c$. If all the itemsets in S have confidence $< h_c$, then T and S are disjoint.

Problem 7a:

{a,b,c},
{a,c,d}

Problem 7b:

{a,b,c},
{a,b,d},
{a,c,d},
{a,c,e},
{a,d,e},
{b,c,d},
{b,c,e},
{c,d,e}

Problem 7c:

True. This is the definition of maximal frequent itemset: an itemset is maximal frequent if NONE of its itemsets are frequent.

Problem 7d:

True. An itemset can be closed and infrequent. An itemset is closed only if all of its supersets have a lower support than itself, which is unrelated to frequency.

Problem 8a:

ASSUMPTION FOR ALL ANSWERS: All grayed areas are frequent. (c) will produce the highest number of frequent itemsets. Since (c) has the longest single frequent itemset, every subset of that largest frequent itemset is also frequent. So it has something like $\binom{600}{1} + \binom{600}{2} + \dots + \binom{600}{600}$ which is an extremely large number.

Problem 8b:

(a) will produce 5 closed frequent itemsets since none of the blocks overlaps with any other block.

Problem 8c:

(c) has the longest size frequent itemset of roughly size-800.

Problem 8d:

(b) will produce a frequent itemset with support of 8000, which is larger than any other itemset in the choices.

Problem 8e:

(c) will also produce the longest sized closed itemset which is the same as the longest sized frequent itemset.