CptS 315: Data Mining

Homework #1 Solutions

Q1:

a) 4 (Four baskets contain both A & B)

b) $4/6$ (Four out of six buckets)

c) $$\text{Conf}(A \Rightarrow B) = \frac{\text{Support}(\{A, B\})}{\text{Support}(\{A\})}$$

$$= \frac{4}{6}$$

$$= 2/3$$

Q2:

a) The index of pair $\{i, j\}$ where $i < j$

$$= (i-1)(n - i/2) + j - i$$

$i = 7$, $j = 8$ $\Rightarrow$ $(7-1)(20 - 7/2) + 8 - 7 = 100$

b) Tabular approach beats triangular matrix only when at most 1/3 of all pairs have a non-zero count

Given, 10 percent of total pairs have a non-zero count.

$$10\% \leq 33.3\%$$

$\Rightarrow$ we prefer tabular method

Q3:

(a) Absolute support

$\{1\} = 4$
$\{2\} = 6$
$\{3\} = 8$
$\{4\} = 8$
$\{5\} = 6$
$\{6\} = 4$

$\{1,2\} = 2$
$\{1,3\} = 3$
$\{1,4\} = 2$
$\{1,5\} = 1$
$\{1,6\} = 0$
$\{2,3\} = 3$
$\{2,4\} = 4$
$\{2,5\} = 2$
$\{2,6\} = 1$

$\{3,4\} = 4$
$\{3,5\} = 4$
$\{3,6\} = 2$
$\{4,5\} = 3$
$\{4,6\} = 3$
$\{5,6\} = 2$

(b)

Pair	Bucket #
{1,2}	2
{1,3}	3
{1,4}	4
{1,5}	5
{1,6}	6
{2,3}	6
{2,4}	8
{2,5}	10
{2,6}	1
{3,4}	1
{3,5}	4
{3,6}	7
{4,5}	9
{4,6}	2
{5,6}	8

(c)

Bucket #	Support	
0	0	
1	5	✓
2	5	✓
3	3	
4	6	✓
5	1	
6	3	
7	2	
8	6	✓
9	3	
10	2	

Frequent Buckets: Those with support above 4

1, 2, 4, 8

d) Pairs mapped to frequent buckets will be counted in second pass of PCY algorithm.

Frequent Buckets = 1, 2, 4, 8

All pairs mapped to frequent Bukets =

= {1,2}, {1,4}, {2,4}, {2,6},
{3,4}, {3,5}, {4,6}, {5,6}