

Problem 1

Data_A

A	B	C	D	E	F	G	H	I	J
1	1	1	1	1					
1	1	1	1	1					
1	1	1	1	1					
1	1	1	1	1					
1	1	1	1	1					
					1	1	1	1	1
					1	1	1	1	1
					1	1	1	1	1
					1	1	1	1	1
					1	1	1	1	1

Frequent itemsets of A		frequency
1 item	A\B\C\D\E\F\G\H\I\J	0.5
2 items	AB\AC\AD\AE\BC\BD\BE\CD\CE\DE FG\FH\FI\FJ\GH\GI\GJ\HI\HJ\IJ	0.5
3 items	ABC\ABD\ABE \ACD \ACE \ADE \BCD \BCE \BDE \CDE \FGH \FGI \FGJ \FHI \FHJ \FIJ \GHI \GHJ \GIJ \HIJ	0.5
4 items	ABCD\ABCE\ABDE\ACDE\BCDE\FGHI\FGHJ\FHIJ\FGIJ\GHIJ	0.5
5 items	ABCDE\FGHIJ	0.5

Data_B

A	B	C	D	E	F	G	H	I	J
1	1								
1	1								
1	1								
1	1		1	1					
1	1								
1	1					1			
1	1					1			
1	1								
1	1								
1	1								

Frequent itemsets of B		frequency
1 item	A(1)\B(1)\G(0.2)	
2 items	AB(1)\AG(0.2)\BG(0.2)	

3 items	ABG(0.2)	
4 items		
5 items		

Data_C

A	B	C	D	E	F	G	H	I	J
1	1	1							
1	1	1							
	1	1	1						
		1	1						
			1	1	1	1	1	1	1

Frequent itemsets of C		frequency
1 item	A(0.2)\B(0.3)\C(0.4)\D(0.3)	
2 items	AB(0.2)\AC(0.2)\BC(0.3)\CD(0.2)	
3 items	ABC(0.2)	
4 items		
5 items		

a:

the number of frequent itemsets for A: 62

the number of frequent itemsets for B: 7

the number of frequent itemsets for C: 9

The most is dataset A.

b:

Dataset A has the longest. It is 5.

c:

Dataset A has the highest. It is 0.5.

d:

Dataset B is the most different, whose items has the difference of $0.8=1-0.2$.

e:

the number of maximal frequent itemset for A: 2. (ABCDE\FGHIJ)

the number of maximal frequent itemset for B: 1. (ABG)

the number of maximal frequent itemset for C: 2. (ABC\CD)
 Dataset A and C has the most number of maximal frequent itemset.

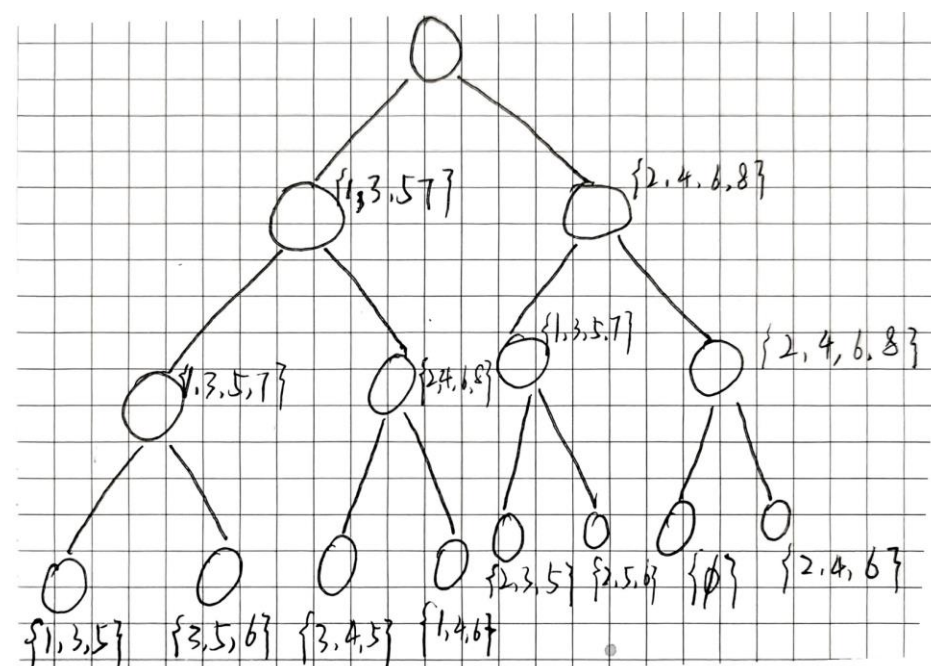
f:

the number of closed frequent itemset for A: 2. (ABCDE\FGHIJ)
 the number of closed frequent itemset for B: 1. (ABG)
 the number of closed frequent itemset for C: 5. (C\D\BC\CD\ABC)
 Dataset C has the most number of closed frequent itemset.

Problem 2

a: create the hash tree:

Inserting Itemset	Path Taken (Root → Nodes)	Action Taken
{1, 2, 4}	Left → Right → Right	Inserted into leaf node
{1, 3, 5}	Left → Left → Left	Inserted into leaf node
{1, 4, 6}	Left → Right → Right	New leaf node created and inserted
{2, 3, 5}	Right → Left → Left	Inserted into leaf node
{2, 5, 6}	Right → Left → Right	Inserted into leaf node
{3, 4, 5}	Left → Right → Left	Inserted into leaf node
{3, 5, 6}	Left → Left → Right	Inserted into leaf node
{2, 4, 6}	Right → Right → Right	Inserted into leaf node



b:

Leaf nodes: 8

Internal nodes: 7

c:

Transaction {1, 2, 3, 5, 6} contains the following 3-itemsets:

{1, 2, 3}

{1, 2, 5}

{1, 2, 6}

{1, 3, 5}

{1, 3, 6}

{1, 5, 6}

{2, 3, 5}

{2, 3, 6}

{2, 5, 6}

{3, 5, 6}

According to the hash tree path, check the leaf nodes for each 3-itemset in the hash tree.

{1, 3, 5}: Path is left → left → left, reaching a leaf node, match found.

{2, 3, 5}: Path is right → left → left, reaching a leaf node, match found.

{2, 5, 6}: Path is right → left → right, reaching a leaf node, match found.

{3, 5, 6}: Path is left → left → right, reaching a leaf node, match found.

The candidate 3-itemsets included in transaction {1, 2, 3, 5, 6} are:

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

Problem 3

a:(support 0.3)

a	0.6	F	abz	0.1	I
b	0.7	F	axy	0.3	F
x	0.6	F	axz	0.2	I
y	0.7	F	ayz	0.2	I
z	0.5	F	bxy	0.3	F
ab	0.4	F	byz	0.1	I
ax	0.4	F	xyz	0.1	I
ay	0.4	F	bxz	0.1	I
az	0.3	F	abxy	0.2	I
xy	0.5	F	abxz	0.2	I
xz	0.2	I	abyz	0	N
yz	0.3	F	axyz	0.1	I
bx	0.4	F	bxyz	0	N
by	0.4	F	abxyz	0	N
bz	0.3	F			
abx	0.3	F			
aby	0.2	I			

b:

the percentage of frequent itemset: $17/34=0.5$.

c:

the pruning ratio: $17/34=0.5$.

d:

the false alarm rate: $11/31=0.3548$ (approximate)