CSE 881 HW 3
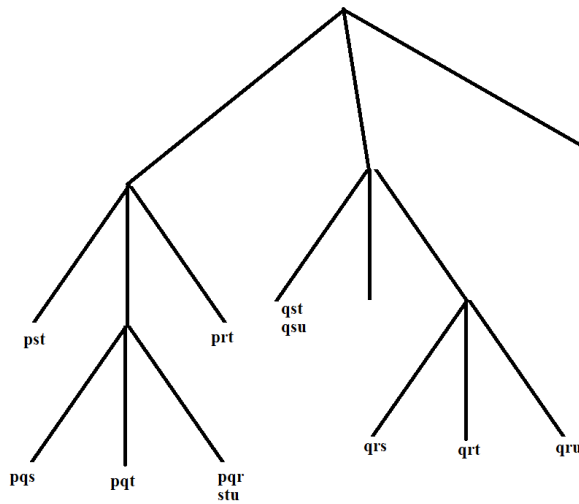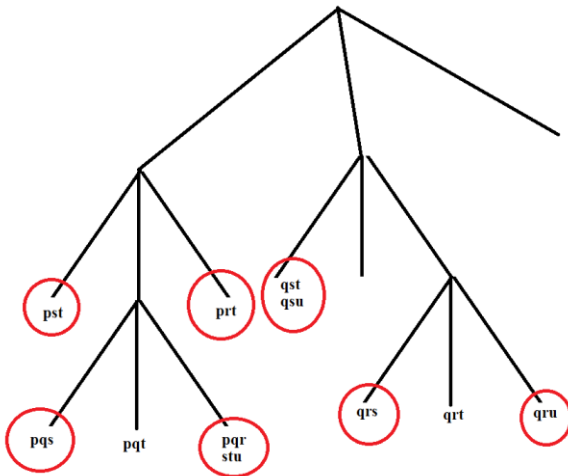
By Abhiram Durgaraju

Problem 1)

a)

b) A total of 7 leaf nodes will be hashed into for the transaction {p,q,r,s,u}. The leaf nodes are circled in red below:

c) All candidate 4-itemsets using Apriori algorithm before pruning:

{p,q,r,s}, {p,q,r,t}, {p,q,s,t}, {q,r,s,t}, {q,r,s,u}, {q,r,t,u}, {q,s,t,u}

d) All candidate 4-itemsets using Apriori algorithm after pruning:

{p,q,r,t}, {p,q,s,t},

We had to prune the other itemsets since at least one of their subsets was not frequent.

Problem 2)

a)

i) False. Suppose the following list of transactions: {A,D}, {A,B,C}, {A,B,C,D}. Clearly, the support of {A,B} is the same as support of {A,B,C} = 2/3. The support of {A} = 3/3 and the support of {A,C} = 2/3. So they are not equal.

ii) **True**. Conf({A,B} -> {C}) = support({A,B,C})/support({A,B}). Since we suppose these are true, confidence is 1, therefore 100%

iii) False. The same logic from part i) applies. Suppose the following list of transactions: {A,D}, {A,B,C}, {A,B,C,D}. Clearly, the support of {A,B} is the same as support of {A,B,C} = 2/3. But note that {A,D} shows that there exists a transaction that contains item A but does not have item C.

iv) False

b)

i) False. Suppose the following list of transactions: {A,B,D}, {A,C,D}, {A,B,C}. Clearly, the support of {A,B} is the same as support of {A,C} = 2/3. But note that transaction {A,B,D} contains item B but does NOT contain item C. Therefore, the assertion is false.

ii) False. Same logic from part i) applies. Suppose the following list of transactions: {A,B,D}, {A,C,D}, {A,B,C}. Conf({A,B} -> {C}) = support({A,B,C})/support({A,B}). In the transaction list we see that support({A,B,C}) = 1/3 whereas support({A,B}) = 2/3. So, . Conf({A,B} -> {C}) = 1/2. Therefore, it is not 100% and assertion is false.

iii) False. Same logic as above, in ii). If the support({A,B,C}) was equal to support({A,B}), then the Conf({A,B} -> {C}) would be 100%. Since we proved that false in ii), this assertion is also false.

iv). **True**. Suppose the following list of transactions: {A,B,C,D},{A,B,C,D,E}. Clearly, the support of {A,B} is the same as the support of {A,C} = 2/2. Also, note that the support of {A,B,D} = 2/2 whereas the support of {A,B,C,D} has a support of 2/2. Since at least one of the supersets has the same support as the original itemset, {A,B,D} is NOT closed.

c)

i) False. Suppose the following list of transactions: {A,B,C}, {B,C,D}, {B,C,E}. Clearly, all transaction that contain {A,B} also contain {B,C}. If Conf({B,C} -> {A}) = 100%, then support ({A,B,C})/support({B,C}) = 1. But we can see from the above list of transactions that support ({A,B,C}) = 1/3 and support({B,C}) = 2/3, and so the assertion is false.

ii)False. Suppose the same list of transactions: {A,B,C}, {B,C,D}, {B,C,E}. Clearly, all transaction that contain {A,B} also contain {B,C}. But, note that support of {A} = 1/3 and support of {C} = 2/3. Therefore the assertion is false.

iii) **True**. Suppose the following list of transactions: {A,B,C,D},{A,B,C,D,E}. Clearly, all transactions that contain of {A,B} also contain {A,C}. Also note that the support of {A,B,D} = 2/2 and the support of {A,B,C,D} has a support of 2/2. Since at least one of the supersets has the same support as the original itemset, {A,B,D} is NOT closed.

d)

i) False

ii)False

iii) **True**

e)

i) {A,B}

ii) {A,B}

3)

a) I downloaded the program

b)

I used the following command:

apriori -s30n1 votes.tab results_3b.txt

I found 33 frequent itemsets

Support(republican) = 38.6207

Support(democrat) = 61.3793

Democratic party has more representatives in the data.

The file results_3b.txt has the list of frequent itemsets.

c)

The file votes.app is attached inside zip folder. I simply wrote the following:

in

democrat out

republican out

d)

I used the following command:

apriori - trs30c70n2 votes.tab results_3d.txt -R votes.app

19 rules were generated.

For each bill, the following party will most likely vote yes:

- education-spending = republican
- adoption-of-the-budget-resolution = democrat
- Physician-fee-freeze = republican
- aid-to-nicaraguan-contras = democrat
- mx-missile = democrat
- el-salvador-aid = republican

The file results_3d.txt is attached to the zip folder

e)

34 rules were generated.

The bills which representative not vote along their party lines are the rule which have confidence between 45% and 55%:

synfuels-corporation-cutback=n

religious-groups-in-schools=y

The results of this part is stored in results_3e.txt