The answer should be started from C2, but for learning purpose I will try to solve the question from the beginning.

MIS (Bread) = 70%, MIS (Milk) = 50% and rest of items 25%

**Transactions:**

Beef, Bread

Bread, Clothes

Bread, Clothes, Milk

Cheese, Boots

Beef, Bread, Cheese, Shoes

Beef, Bread, Cheese, Milk

Bread, Milk, Clothes

The steps for solving the question are as follows:

1. Unique-items are found from transactions, they are sorted based on MIS (minimum item support) in ascending order.

I = {Beef, Bread, Clothes, Milk, Cheese, Boots, Shoes} 7 items

25%

25%

25%

50%

70%

25%

25%

M = {Beef, Boots, Cheese, Clothes, Shoes, Milk, Bread}

1. The support-count or frequency of each item is found from transactions. Following the sorted order the first item i in M that meets MIS (i) is found. Item i is inserted into L. For each subsequent item j in M after i, if, then j is also inserted into L, where is the support-count of j and n is the total number of transactions in T.

14%

43%

43%

43%

43%

14%

86%

{Beef: 3/7, Boots: 1/7, Cheese: 3/7, Clothes: 3/7, Shoes: 1/7, Milk: 3/7, Bread: 6/7}

43%

43%

43%

43%

86%

L = {Beef: 3/7, Cheese: 3/7, Clothes: 3/7, Milk: 3/7, Bread: 6/7}

1. Each item is taken from L and compared its actual support with its MIS. If actual support is greater than or equal to, the item is added to a new set called F1, else, the item is dropped.

F1 = {{Beef}, {Cheese}, {Clothes}, {Bread}}

1. As K = 2, Candidates are generated based on the following from L.

50%

70%

25%

25%

25%

25%

25%

M = {Beef, Boots, Cheese, Clothes, Shoes, Milk, Bread}

86%

43%

43%

43%

43%

L = {Beef: 3/7, Cheese: 3/7, Clothes: 3/7, Milk: 3/7, Bread: 6/7}

C2 = {{Beef, Cheese}, {Beef, Clothes}, {Beef, Milk}, {Beef, Bread},

{Cheese, Clothes}, {Cheese, Milk}, {Cheese, Bread},

{Clothes, Milk}, {Clothes, Bread}

}

1. Each element c is taken from in this case C2 and its frequency is counted.

28% 0% 14% 43%

C2 = {{Beef, Cheese}: 2/7, {Beef, Clothes}: 0/7, {Beef, Milk}: 1/7, {Beef, Bread}: 3/7,

0% 14% 28%

{Cheese, Clothes}: 0/7, {Cheese, Milk}: 1/7, {Cheese, Bread}: 2/7,

28% 43%

{Clothes, Milk}: 2/7, {Clothes, Bread}: 3/7

}

1. The actual-support of each element c in CK in this case C2 is compared with MIS of first element in c. If the actual-support of c is greater or equal to MIS of first element in c, c is added to a new set called Fk in this case F2.

F2 = {{Beef, Cheese}, {Beef, Bread}, {Cheese, Bread}, {Clothes, Milk}, {Clothes, Bread}}

1. For K > 2, function is used.
2. Function basically takes each pair of elements from FK-1  that differ only in the last item and MIS of last item in f1 should be less than or equal to MIS of last item in f2. Then joins f1 and f2 into a new set called c.

1. Inserts c into and generates the k-1 subset of c called s.

S1 =

S2 =

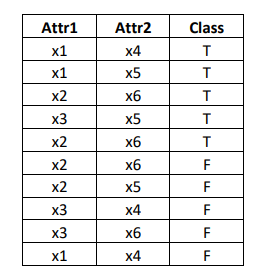
S3 =

1. Now first we check if (Beef) is in s or MIS of first element in c is equal to MIS of second element in, then we need to check if all s are element of FK-1, if the condition doesn’t hold we need to delete c from Ck. In this case subset of c, s are elements of F2, so we keep it.

Step 7 will be repeated for all pairs of F2 and the following C3 is generated.

Step 5 and 6 will be repeated on C3 and the following F3 is generated.

Question 2 (40pts): Suppose that we have the training data set in the Figure below, which has two attributes Attr1 and Attr2, and the Class. Compute all the probability values required to learn a naïve Bayesian classifier. Then predict the class of the following attributes

Attr1=x1, Attr2=x6 Class=?

Attr1=x2, Attr2=x4 Class=?

Attr1=x1, Attr2=x6 Class=?

**Therefore, X belongs to class T**

Attr1=x2, Attr2=x4 Class=?

**Therefore, Y belongs to class F**