**1 EXERCISE:**

* 1. Chapter 5:

Q7. a) P(A = 1|+) = 3/5=0.6, P(A=0|+) = 2/5=0.4, P(A=1|-)=2/5=0.4, P(A=0|-)=3/5=0.6

P(B=1|+) = 1/5=0.2, P(B=0|+) = 4/5=0.8, P(B=1|-) =2/5=0.4, P(B=0|-) =3/5=0.6

P(C=1|+) = 4/5=0.8, P(C=0|+) = 1/5=0.2, P(C=1|-) =5/5 =1, P(C=0|-) =0/5=0

Q7.b) Predict outcome of P(A=0,B=1,C=0)

1. P(class = +|A=0,B=1,C=0) = P(A=0|+) \* P(B=1|+)\*P(C=0|+)\*P(+)\*a ,

where a Is constant 1/p(x)

Hence P (class = +|A=0, B=1, C=0) = 0.4\*0.2\*0.2\*0.5 = 0.008a

1. P (class = -|A=0, B=1, C=0) = P(A=0|-) \* P(B=1|-) \*P(C=0|-) \*P (-) \*a,

where a Is constant 1/p(x)

Hence P (class = -|A=0, B=1, C=0) = 0.6\*0.6\*0\*0.5\*a = 0

*The class label must be +*

Q8) To estimate probabilities for

Qa)

P(A=1|+) = 3/5 = 0.6

P(B=1|+) = 2/5 = 0.4

P(C=1|+) = 4/5 = 0.8

P(A=1|-) = 2/5 = 0.4

P(B=1|-) = 2/5 = 0.4

P(C=1|-) = 1/5 = 0.2

Qb)

P (+|A=1, B=1, C=1) = 0.6\*0.4\*0.8\*0.5\*a = 0.096a

P (-|A=1, B=1, C=1) = 0.4\*0.4\*0.2\*0.5\*a = 0.016a

*The class label must be +*

Qc)

P(A=1) = 0.5, P(B=1) = 0.4 & P (A=1, B=1) = 1/5 = 0.2

Hence, there is no dependency between A and B

Qd)

P(A=1) = 0.5, P(B=0) = 0.6, P (A=1, B=0) = 0.3

Hence, there is no dependency between A and B.

Qe)

P (A=1, B=1|+) = 0.2, P(A=1|+) = 0.6, P(B=1|+) = 0.4

Given the probabilities with class labels, A&B are not conditionally independent.

**CHAPTER 6:**

Q2)

Qa) Support calculation:

{e} = 8/10 = 0.8

{b,d} = 2/10 = 0.2

{b,d,e} = 2/10 = 0.2

Qb) Confidence Calculation:

{b,d} -> {e} = 0.2/0.2 = 1 or 100%

{e} -> {b,d} = 0.2/0.8 = 0.25 or 25%

**Confidence is not a symmetric measure**, which is proved by confidence calculation in above question.

Qc) Calculate support by considering customer id as transaction id:

{e} = 4/5 = 0.8

{b,d} = 5/5 = 1

{b,d,e} = 4/5 = 0.8

Qd) Calculate confidence:

{b,d} -> {e} = 0.8/1 = 0.8

{e} -> {b,d} = 0.8/0.8 = 1

Qe) Support and confidence of rule r when treating each transaction id as market basket s1 and c1 respectively and support and confidence of rule r when treating each customer id as market basket s2 and c2 respectively **cannot have any relationship**.

Q6)

Qa) Maximum number of Association rules that can be extracted from the market basket transactions are: 3^d - 2^(d+1) + 1, where d=6 in our case

Hence Total Rules = **602.**

Qb) The maximum size of frequent itemset that can be generated from the **transaction**

**data is 4** because the maximum transaction length is 4.

Qc) The expression for maximum number of size 3 frequent itemset that can be generated

is**: 6C3**, where c is the combinations.

Hence the maximum 3-frequent item sets are: **20**

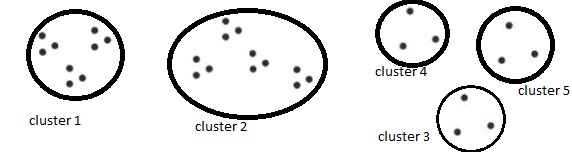
Qd) The itemset that has largest support.

Upon analysis of the transaction data, the itemset **{Bread, Butter}** has **support count** **of 5**, which is maximum of all item sets.

Qe) The itemset which has same confidence in {a} -> {b} & {b} ->{a} is again **{Bread, Butter}**

**CHAPTER 8:**

Q2)



Q11)

1. If the SSE of one variable is less for all cluster then the attribute is constant and will be of no use for clustering.
2. If the SSE of attribute is less for only one cluster or few clusters then this attribute is good as it divides the objects well.
3. If the SSE of attribute is high for all clusters then it could be an outlier or noise.
4. If SSE is high for all clusters then it means the attribute defines different cluster than other attributes and it cannot be used to create the cluster.
5. The SSE per variable can be used to identify which attribute gives better clustering of the data. The attributes which cannot distinguish between variable can be identified with SSE and clustering can be create without it.