Chapter 5

7)

a) $P(A=1|+) = \frac{P(A,+)}{P(+)} = \frac{0.3}{0.5} = \boxed{\frac{3}{5} = 0.6}$ $\quad P(A=0|+) = \boxed{\frac{2}{5} = 0.4}$

$P(B=1|+) = \frac{P(B,+)}{P(+)} = \frac{0.1}{0.5} = \boxed{\frac{1}{5} = 0.2}$ $\quad P(B=0|+) = \boxed{\frac{4}{5} = 0.8}$

$P(C=1|+) = \frac{P(C|+)}{P(+)} = \frac{0.4}{0.5} = \boxed{\frac{4}{5} = 0.8}$ $\quad P(C=0|+) = \boxed{0.2}$

$P(A=1|-) = \frac{P(A,-)}{P(-)} = \frac{0.2}{0.5} = \boxed{\frac{2}{5} = 0.4}$ $\quad \boxed{P(A=0|-) = 0.6}$

$P(B=1|-) = \frac{P(B,-)}{P(-)} = \frac{0.2}{0.5} = \boxed{\frac{2}{5} = 0.4}$ $\quad \boxed{P(B=0|-) = 0.6}$

$P(C=1|-) = \frac{P(C,-)}{P(-)} = \frac{0.5}{0.5} = \boxed{\frac{1}{1} = 1.0}$ $\quad \boxed{P(C=0|-) = 0}$

b) $P(+|A=0,B=1,C=0) = \frac{P(A=0,B=1,C=0|+)\ P(+)}{P(A=0,B=1,C=0)}$

$= \frac{P(A=0|+) \cdot P(B=1|+)\ P(C=0|+)\ P(+)}{P(A=0,B=1,C=0)}$

$= \frac{0.4 \times 0.2 \times 0.2 \times 0.5}{P(A=0,B=1,C=0)} = \frac{0.008}{K}$

$P(-|A=0,B=1,C=0) = \frac{P(A=0,B=1,C=0|-)\ P(-)}{K}$

$= \frac{P(A=0|-)\ P(B=1|-)\ P(C=0|-)\ P(-)}{K}$

$= \frac{0.6 \times 0.2 \times 0 \times 0.5}{K} = \frac{0}{K}$

The conditional probability of K given a label of '+' is greater than the probability of K given a label of '-'.

$\boxed{\therefore \text{ The class label would by } +}$

8) a) $P(A=1|+) = \frac{0.3}{0.5} = \boxed{0.6}$ $P(A=1|-) = \frac{0.2}{0.5} = \boxed{0.4}$

$P(B=1|+) = \frac{0.2}{0.5} = \boxed{0.4}$ $P(B=1|-) = \frac{0.2}{0.5} = \boxed{0.4}$

$P(C=1|+) = \frac{0.4}{0.5} = \boxed{0.8}$ $P(C=1|-) = \frac{0.1}{0.5} = \boxed{0.2}$

b)

$$P(+|A=1,B=1,C=1) = \frac{P(A=1,B=1,C=1|+)\,P(+)}{K}$$

$$= \frac{P(A=1|+)\,P(B=1|+)\,P(C=1|+)\,P(+)}{K}$$

$$= \frac{0.6 \times 0.4 \times 0.8 \times 0.5}{K} = \frac{0.096}{K}$$

$$P(-|A=1,B=1,C=1) = \frac{P(A=1|-)\,P(B=1|-)\,P(C=1|-)\,P(-)}{K}$$

$$= \frac{0.4 \times 0.4 \times 0.2 \times 0.5}{K} = \frac{0.016}{K}$$

Because $P(+|K) > P(-|K)$ K is labled +

c) $P(A=1) = 0.5$
$P(B=1) = 0.4$
$P(A=1, B=1) = 0.2$

A + B are independent of eachother

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

A + B are independent of eachother

e) $P(A=1, B=1|+) = \frac{0.1}{0.5} = 0.2$

$P(A=1|+) = 0.6$
$P(B=1|+) = 0.4$

$0.6 \times 0.4 = 0.24$

A + B are NOT conditionally independent of one another chapter 6

2) a) $\sigma\{e\}=8$ $\sigma\{b,d\}=2$ $\sigma\{b,d,e\}=2$

$s(e)=\frac{\sigma e}{10}=\boxed{0.8}$ $s(b,d,e)=\frac{2}{10}=\boxed{0.2}$

$s(b,d)=\frac{2}{10}=\boxed{0.2}$

b) $c(\{b,d\}\rightarrow\{e\})=\frac{\sigma\{b,d,e\}}{\sigma\{b,d\}}=\frac{2}{2}=\boxed{1=100\%}$

$c(e\rightarrow b,d)=\frac{\sigma(e,b,d)}{\sigma(e)}=\frac{2}{8}=\boxed{\frac{1}{4}=25\%}$

c) $\sigma\{e\}=4$ $\sigma\{b,d\}=5$ $\sigma\{b,d,e\}=4$

$s(e)=\frac{4}{5}=\boxed{0.8}$ $s(b,d,e)=\frac{4}{5}=\boxed{0.8}$

$s(b,d)=\frac{5}{5}=\boxed{1}$

d) $c(b,d\rightarrow e)=\boxed{\frac{4}{5}=0.8=80\%}$

$c(e\longrightarrow b,d)=\boxed{\frac{4}{4}=1=100\%}$

e) There isn't any obvious relationships amongst S_1 and S_2 + C_1+C_2

6) a) $\text{max\_rules}=3^d-2^{d+1}+1=3^6-2^7+1=\boxed{602 \text{ rules}}$

b) max_fis = size of largest transaction
$=\boxed{4}$

c) $\binom{\#\text{items}}{\text{size of }k}=\binom{6}{3}=\frac{6\times5\times4\times\cancel{3\times2\times1}}{3\times2\times1\times\cancel{3\times2\times1}}=\frac{120}{6}=20$

$\boxed{\binom{6}{3}=20}$

d) Support = $\frac{\sigma(X \cup Y)}{\# t}$

{Bread, Butter}

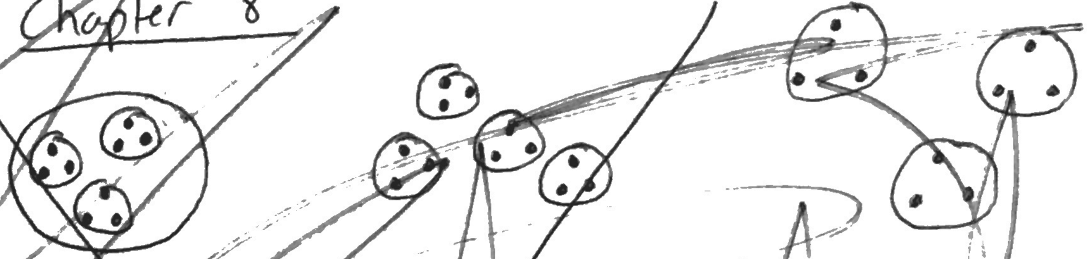
e) conf(Bread → Butter) = $\frac{5}{5} = 1$

conf(Butter → Bread) = $\frac{5}{5} = 1$

$\boxed{\{Bread, Butter\}}$

Chapter 8

2)



low for just one cluster:

This is a good attribute that is representative of the cluster

SSE for 1 variable low for all clusters:

the variable is not good for dividing the data point into separate groups, we cannot use it to figure out shit

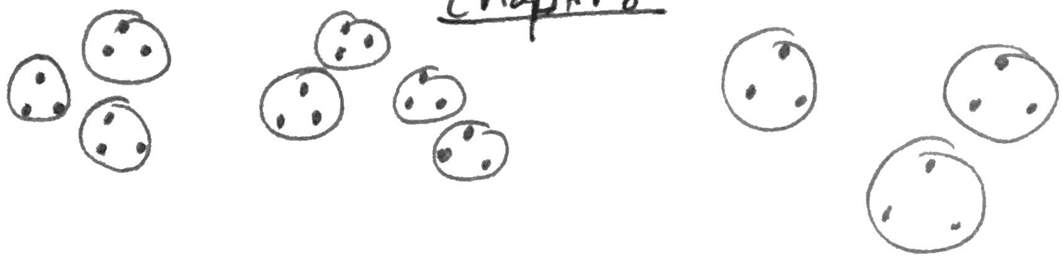
High for all clusters:

The attribute could potentially be noise

High for one cluster:

This attribute is not good to use to define the cluster. It could be that the cluster was created from another attribute that was more representative of the objects

Chapter 8

2) 

11)

(a) SSE low for all clusters

The variable is not useful for dividing observations into groups. We can essentially treat the variable as a constant

(b) SSE low for 1 cluster

This is a good attribute that represents the observations in the cluster

(c) high for all clusters

Attribute could potentially be noise

(d) high for one cluster

this attribute has an SSE that contradicts the observations of that cluster.
ATTRIBUTE NOT GOOD FOR DEFINING THE CLUSTER

(e) per variable SSE helps us to identify & remove attributes that do not help to form unique clusters. It basically helps us to optimize our clustering by using ~~the~~ attributes that highlight the characteristics of each cluster.