Student number: z5192086
Student Name: Pan Luo
File description: 19 T1, COMP9318 Assignment

Q1:
(1)

Location	Time	Item	Sum(Quantity)
Sydney	2005	PS2	1400
Sydney	2006	PS2	1500
Sydney	2006	Wii	500
Melbourne	2005	Xbox360	1700
Sydney	2005	All	1400
Sydney	2006	All	2000
Melbourne	2005	All	1700
Sydney	All	PS2	2900
Sydney	All	Wii	500
Melbourne	All	Xbox360	1700
Sydney	All	All	3400
Melbourne	All	All	1700
All	2005	PS2	1400
All	2005	Xbox360	1700
All	2006	PS2	1500
All	2006	Wii	500
All	All	PS2	2900
All	All	Wii	500
All	All	Xbox360	1700
All	2005	All	3100
All	2006	All	2000
All	All	All	5100

As above shows, there are 22 tuples in the complete data cube of R.

(2)The equal SQL language are following:
Select Location, Time, Item, Sum(Quantity)
From Sales
Group by Location, Time, Item
Union All
Select Location, Time, Sum(Quantity)
From Sales
Group by Location, Time
Union All
Select Location, Item, Sum(Quantity)
From Sales
Group by Location, Item
Union All

```
Select Time, Item, Sum(Quantity)
From Sales
Group by Time, Item
Union All
Select Location, Sum(Quantity)
From Sales
Group by Location
Union All
Select Time, Sum(Quantity)
From Sales
Group by Time
Union All
Select Item, Sum(Quantity)
From Sales
Group by Item
Union All
Select Sum(Quantity)
From Sales
```

(3)

Location	Time	Item	Quantity
Sydney	2006	ALL	2000
Sydney	ALL	PS2	2900
Sydney	ALL	ALL	3400
ALL	2005	ALL	3100
ALL	2006	ALL	2000
ALL	ALL	PS2	2900
ALL	ALL	ALL	5100

(4)

Offset	Location	Time	Item	Sum(Quantity)
18	1	1	1	1400
22	1	2	1	1500
24	1	2	3	500
32	2	1	2	1700
17	1	1	0	1400
21	1	2	0	2000
30	2	1	0	1700
14	1	0	1	2900
15	1	0	3	500
28	2	0	2	1700
13	1	0	0	3400
26	2	0	0	1700

5	0	1	1	1400
6	0	1	2	1700
9	0	2	1	1500
11	0	2	3	500
1	0	0	1	2900
3	0	0	3	500
2	0	0	2	1700
4	0	1	0	3100
8	0	2	0	2000
0	0	0	0	5100

The mapping offset function I use is f(n) = 13*Location+4*Time+ 1*Item which ensure each combination of two sets do not equal to the third set. And that ensure the uniqueness of the offset value.

Q2:

(1)For Naïve Bayes classifier, xi belong to class yi if P(xi|yi) is the maximum probability among all the i numbers. So according to the assumption, there are two classes 0 and 1, and x is a binary vector. In order to make Naïve Bayes to be a binary classifier:

NB(X)=1 if P(y=1|x)/ P(y=0|x)>1,

NB(X)=0 if P(y=1|x)/ P(y=0|x)<1

So we should find how to represent P(y=1|x)/ P(y=0|x), then

$$NB(x) = \begin{cases} 1 & \text{if } \frac{P(y=1|x)}{P(y=0|x)} > 1 \\ 0 & \text{if } \frac{P(y=1|x)}{P(y=0|x)} < 1 \end{cases}$$

According to Bayes Rules,

$$\frac{P(y=1|x)}{P(y=0|x)} = \frac{P(x|y=1) \cdot P(y=1)}{P(x)} \cdot \frac{P(x)}{P(x|y=0) \cdot P(y=0)}$$

$$= \frac{P(x|y=1) \cdot P(y=1)}{P(x|y=0) \cdot P(y=0)}$$

$$= \frac{P(y=1)}{P(y=0)} \cdot \prod_{i=1}^{n} \frac{P(X_i|y=1)}{P(X_i|y=0)}$$

where n represents the different feature vectors in X

Set $\alpha_i = P(x_i=1|y=1)$, then

$1-\alpha_i = P(X_i=0|y=1)$, so

$$P(X_i|y=1) = \alpha_i^{x_i} \cdot (1-\alpha_i)^{1-x_i}$$

Similarly, set $\beta_i = P(X_i=1 \mid y=0)$, so

$1-\beta_i = P(X_i=0 \mid y=0)$, then

$$P(X_i \mid y=0) = \beta_i^{x_i} \cdot (1-\beta_i)^{1-x_i}$$

So $\dfrac{P(y=1)}{P(y=0)} \cdot \prod_{i=1}^{n} \dfrac{P(X_i \mid y=1)}{P(X_i \mid y=0)}$

$$= \frac{P(y=1)}{P(y=0)} \cdot \prod_{i=1}^{n} \frac{\alpha_i^{x_i} \cdot (1-\alpha_i)^{1-x_i}}{\beta_i^{x_i} \cdot (1-\beta_i)^{1-x_i}}$$

Apply log function on above equation, then

$$\log \frac{P(y=1)}{P(y=0)} + \sum_{i=1}^{n} \log \frac{\alpha_i^{x_i} \cdot (1-\alpha_i)^{1-x_i}}{\beta_i^{x_i} \cdot (1-\beta_i)^{1-x_i}}$$

For the RHS part

$$\sum_{i=1}^{n} \log \frac{\alpha_i^{x_i} (1-\alpha_i)^{1-x_i}}{\beta_i^{x_i} \cdot (1-\beta_i)^{1-x_i}}$$

$$= \sum_{i=1}^{n} \left[\log \alpha_i^{x_i} + \log (1-\alpha_i)^{1-x_i} - \log \beta_i^{x_i} - \log (1-\beta_i)^{1-x_i} \right]$$

$$= \sum_{i=1}^{n} \left[x_i \log \frac{\alpha_i}{\beta_i} + (1-x_i) \log \frac{1-\alpha_i}{1-\beta_i} \right]$$

$$= \sum_{i=1}^{n} \left(x_i \log \frac{\alpha_i}{\beta_i} + \log \frac{1-\alpha_i}{1-\beta_i} - x_i \log \frac{1-\alpha_i}{1-\beta_i} \right)$$

$$= \sum_{i=1}^{n} \left(x_i \underbrace{\boxed{\log \frac{\alpha_i (1-\beta_i)}{\beta_i (1-\alpha_i)}}}_{w_{raw}} + \underbrace{\boxed{\log \frac{1-\alpha_i}{1-\beta_i}}}_{b_{raw}} \right)$$

So the original equation

$$= \underbrace{\log \frac{P(y=1)}{P(y=0)} + \sum_{i=1}^{n} \log \frac{1-\alpha_i}{1-\beta_i}}_{b} + x_i \underbrace{\sum_{i=1}^{n} \log \frac{\alpha_i (1-\beta_i)}{\beta_i (1-\alpha_i)}}_{w}$$

Now I define P(y=1|x)/ P(y=0|x) to the $y=wx_i+b$, so the Naïve Bayes Classifier can be applied on Linear Classification.

The values of w and b are on the above pictures.

(2)For Naïve Bayes Classifier, the value of w can be learned by different probabilities, and these probabilities are depend on the assumption and dataset(various from the number of feature vectors and different classes). So the cost of compute the NB Classifier is cheap.

For Logistic Regression Classifier, the parameter w should be learned by gradient ascent, which may meet the local maximum problem and sometimes hard to learn. So it is unstable.

But for NB Classifier, it uses the assumption that attributes are conditionally independent. So the calculation is easier if we do the smoothing.

What's more, NB Classifier also give another prior probability which may help the confidence of the final result.

Q3:

(1) For the liklihood function

$$L(P_{i,j}|\theta) = P(P_{i,j}|\theta)$$

$$= P(P_{1,1}|\theta) \cdot P(P_{1,2}|\theta) \cdot P(P_{1,3}|\theta) \cdots P(P_{2,3}|\theta)$$

According to the assumption,

$$P(i,j) = P(O_j|S_i)$$

$$P(S_i) = q_i$$

$$P(O_j) = u_j$$

$$\Rightarrow \begin{cases} O_1 = 0.1q_1 + 0.4q_2 \\ O_2 = 0.2q_1 + 0.5q_2 \\ O_3 = 0.7q_1 + 0.1q_2 \end{cases}$$

$$\Downarrow$$

$$O_j = \sum_{i=1}^{2} P_{i,j} q_i$$

So $L(P_{i,j}|\theta) = O_1^{u_1} \cdot O_2^{u_2} \cdot O_3^{u_3}$

The log liklihood function is

$$\log L(P_{i,j},\theta) = u_1 \log O_1 + u_2 \log O_2 + u_3 \log O_3$$

$$= \sum_{j=1}^{3} (u_j \log O_j)$$

$$= \sum_{j=1}^{3} u_j \cdot \log \sum_{i=1}^{2} P_{i,j} q_i$$

(2) $q_2 = 1 - q_1$

According to (1)

the log likelihood function is

$$l(\theta) = 0.3\log[0.1q_1 + 0.4(1-q_1)] + 0.2\log[0.2q_1 + 0.5(1-q_1)] + 0.5\log[0.7q_1 + 0.1(1-q_1)]$$

$$\frac{\partial l}{\partial q} = \frac{-0.09}{0.4-0.3q_1} + \frac{-0.06}{0.5-0.3q_1} + \frac{0.3}{0.6q_1+0.1}$$

Let $\frac{\partial l}{\partial q} = 0 \implies 570q_1^2 - 1179q_1 + 531 = 0$

$q_{11} = 0.635 \quad q_{12} = 1.548$ (invalid)

So $q_2 = 1 - q_1 = 0.365$

Therefore, the expected percentage of each component is:

$O_1 = 0.1q_1 + 0.4q_2 = 0.2095$

$O_2 = 0.2q_1 + 0.5q_2 = 0.3095$

$O_3 = 0.7q_1 + 0.1q_2 = 0.481$