Assignment

Name: Jie Wang
ZID: z5119770

Question 1

(1)The result :

	Location	Time	Item	Quantity
0	Sydney	2005	PS2	1400
1	Sydney	2005	ALL	1400
2	Sydney	2006	Wii	500
3	Sydney	2006	PS2	1500
4	Sydney	2006	ALL	2000
5	Sydney	ALL	Wii	500
6	Sydney	ALL	PS2	2900
7	Sydney	ALL	ALL	3400
8	Melbourne	2005	Xbox 360	1700
9	Melbourne	2005	ALL	1700
10	Melbourne	ALL	Xbox 360	1700
11	Melbourne	ALL	ALL	1700
12	ALL	2005	Xbox 360	1700
13	ALL	2005	PS2	1400
14	ALL	2005	ALL	3100
15	ALL	2006	Wii	500
16	ALL	2006	PS2	1500
17	ALL	2006	ALL	2000
18	ALL	ALL	Wii	500
19	ALL	ALL	Xbox 360	1700
20	ALL	ALL	PS2	2900
21	ALL	ALL	ALL	5100

(2)The code :

```
SELECT   Location, Time, Item, SUM(Quantity)  FROM   R
GROUP BY    Location, Time, Item
UNION ALL
SELECT   Location, Time, ALL, SUM(Quantity) FROM   R
GROUP BY   Location, Time
UNION ALL
SELECT   Location, ALL, Item, SUM(Quantity) FROM   R
GROUP BY   Location, Item
UNION ALL
SELECT   ALL, Time, Item, SUM(Quantity) FROM  R
GROUP BY Time, Item
UNION ALL
SELECT   Location, ALL, ALL, SUM(Quantity)  FROM  R
```

```
GROUP BY Location
UNION ALL
SELECT ALL, Time, ALL, SUM(Quantity)  FROM   R
GROUP BY Time
UNION ALL
SELECT   ALL, ALL, Item, SUM(Quantity)  FROM   R
GROUP BY Item
UNION ALL
SELECT ALL, ALL, ALL, SUM(Quantity)  FROM   R
```

(3)The result:

	Location	Time	Item	Quantity
0	Sydney	2006	ALL	2000
1	Sydney	ALL	PS2	2900
2	Sydney	ALL	ALL	3400
3	ALL	2005	ALL	3100
4	ALL	2006	ALL	2000
5	ALL	ALL	PS2	2900
6	ALL	ALL	ALL	5100

(4)

From the week2 lecture ,we know that how to count the property of each element:

When we know the Location, the conditions of Time ,Item and Quantity are: 3*4

When we know the Time ,the conditions of Item and Quantity are: 4

So the correct function is : h(L,T,I) = 12L + 4T +I

So we change the content to number and count the map index ,the result:

	Location	Time	Item	Quantity	Map index
0	1	1	1	1400	17
1	1	1	0	1400	16
2	1	2	3	500	23
3	1	2	1	1500	21
4	1	2	0	2000	20
5	1	0	3	500	15
6	1	0	1	2900	13
7	1	0	0	3400	12
8	2	1	2	1700	30
9	2	1	0	1700	28
10	2	0	2	1700	26
11	2	0	0	1700	24
12	0	1	2	1700	6
13	0	1	1	1400	5
14	0	1	0	3100	4
15	0	2	3	500	11
16	0	2	1	1500	9

17	0	2	0	2000	8
18	0	0	3	500	3
19	0	0	2	1700	2
20	0	0	1	2900	1
21	0	0	0	5100	0

So we can get the MOLAP architecture:

index	value
17	1400
16	1400
23	500
21	1500
20	2000
15	500
13	2900
12	3400
30	1700
28	1700
26	1700
24	1700
6	1700
5	1400
4	3100
11	500
9	1500
8	2000
3	500
2	1700
1	2900
0	5100

Question 2

(1) We define the logOdds：

$$\text{Odds} = \frac{\Pr[y=1|\mu]}{\Pr[y=0|\mu]}$$

If the Odds larger than 1,that means it is in class(y=1),else in class(y=0),so if we add a log method:

$$\text{logOdds} = \log\left(\frac{\Pr[y=1|\mu]}{\Pr[y=0|\mu]}\right)$$

If it is larger than 0, then the prediction is positive class; otherwise, the classification is the negative class,so:

$$\text{logOdds} = \log\left(\frac{\Pr[y=1|\mu]}{\Pr[y=0|\mu]}\right)$$

$$= \log(\Pr[y=1|\mu]) - \log(\Pr[y=0|\mu])$$

$$= \log(\prod_{i=1}^{d} \Pr[x_i = \mu_i|y = 1] + \log(\Pr[y=1]) - (\log(\prod_{i=1}^{d} \Pr[x_i = \mu_i|y = 0]$$

$$+ \log(\Pr[y=0])$$

$$= \sum_{i=1}^{d} \log(\Pr[x_i = \mu_i|y = 1]) + \log(\Pr[y=1]) - \sum_{i=1}^{d} \log(\Pr[x_i = \mu_i|y = 0])$$

$$- \log(\Pr[y=0])$$

$$= \sum_{i=1}^{d} \log(Pr[x_i = \mu_i|y = 1]) - \log(\Pr[x_i = \mu_i|y = 0]) + \log(\Pr[y=1])$$

$$- \log(\Pr[y=0])$$

$$= \sum_{i=1}^{d} log \frac{\Pr[x_i=\mu_i|y=1]}{\Pr[x_i=\mu_i|y=0]} + \log\frac{\Pr[y=1]}{\Pr[y=0]}$$

We let the α, β, δ, λ to separate the result:

Define :

$$\alpha(\mathrm{i}, \mu i) = \log\frac{\Pr[x_i=\mu_i|y=1]}{\Pr[x_i=\mu_i|y=0]}$$

$$\beta = \log\frac{\Pr[y=1]}{\Pr[y=0]}$$

Then:

$$\text{logOdds} = \sum_{i=1}^{d} \alpha(\mathrm{i}, \mu_i) + \beta$$

$$= \sum_{i=1}^{d}(\alpha(\mathrm{i},1) * \mu_i + \alpha(\mathrm{i},0) * (1 - \mu_i)) + \beta$$

$$= \sum_{i=1}^{d} \alpha(\mathrm{i}, 0) + \big(\alpha(\mathrm{i}, 1) - \alpha(\mathrm{i}, 0)\big) * \mu_i + \beta$$

$$= \delta + \sum_{i=1}^{d} \lambda * \mu_i$$

So the feature that naive bayes learn is that :

$$\omega^T = [\delta, \lambda_1, \lambda_2 \ldots\ldots, \lambda_d]$$

(2)The reason:

In the logistic regression, it has to learn the whole, but in the naïve bayes , every feature in vector is learned independently.

Question 3

(1)From the question:

$h_w(x) = \frac{1}{1+e^{-w^T x}}$

$= \frac{e^{w^T x}}{e^{w^T x}+1}$

$1- h_w(x) = 1 - \frac{e^{w^T x}}{e^{w^T x}+1}$

$= \frac{1}{e^{w^T x}+1}$

The likelihood Function is that:

$L(w) = \prod_{i=1}^{n} h_w(x_i)^{y_i}(1- h_w(x_i))^{1-y_i}$

From the lecture notes ,we can get the loss function from likelihood function using : -ln(likelihood function), so :

Loss function = -ln(L(w))

$= -\ln(\prod_{i=1}^{n} h_w(x_i)^{y_i}(1- h_w(x_i))^{1-y_i})$

$= -\sum_{i=1}^{n}(y_i * \ln(\frac{e^{w^T x_i}}{e^{w^T x_i+1}}) + (1- y_i) * \ln(\frac{1}{e^{w^T x_i+1}}))$

$= -\sum_{i=1}^{n}(y_i * (w^T x_i - \ln(e^{w^T x_i}+1)) + (1-y_i) * (0 - \ln(e^{w^T x_i}+1))$

$= \sum_{i=1}^{n} -[y_i * w^T x_i - y_i * \ln(e^{w^T x_i}+1) - \ln(e^{w^T x_i}+1) + y_i * \ln(e^{w^T x_i}+1)$

$= \sum_{i=1}^{n}(-y_i * w^T x_i + \ln(e^{w^T x_i}+1))$

$= \sum_{i=1}^{n}(-y_i * w^T x_i + \ln(exp^{w^T x_i}+1))$

(2) When the function is :

$$P[y = 1 \mid \mathbf{x}] = f(\mathbf{w}^\top \mathbf{x})$$

So the likelihood function is :

L(w) = $\prod_{i=1}^{n} f(w^T x_i)^{y_i} \, (1 - f(w^T x_i))^{1-y_i}$

Loss function :

= -log(L(w)

= - log($\prod_{i=1}^{n} f(w^T x_i)^{y_i} \, (1 - f(w^T x_i))^{1-y_i}$)

= - $\sum_{i=1}^{n} y_i * \log f(w^T x_i) \; + \; (1 - y_i) * \log(1 - f(w^T x_i))$

= - $\sum_{i=1}^{n} y_i * \log f(w^T x_i) \; + log(1 - f(w^T x_i)) - y_i * log(1 - f(w^T x_i))$

= -$\sum_{i=1}^{n} y_i * \log \frac{f(w^T x_i)}{1 - f(w^T x_i)}$ + $\log(1 - f(w^T x_i))$