COMP9318 Assignment 1

Chunnan Sheng
z5100764

Question 1 (1)

	Location	Time	Item	Quantity
0	Melbourne	2005	XBox360	1700
1	Melbourne	2005	ALL	1700
2	Melbourne	ALL	XBox360	1700
3	Melbourne	ALL	ALL	1700
4	Sydney	2005	PS2	1400
5	Sydney	2005	ALL	1400
6	Sydney	2006	PS2	1500
7	Sydney	2006	Wii	500
8	Sydney	2006	ALL	2000
9	Sydney	ALL	PS2	2900
10	Sydney	ALL	Wii	500
11	Sydney	ALL	ALL	3400
12	ALL	2005	PS2	1400
13	ALL	2005	XBox360	1700
14	ALL	2005	ALL	3100
15	ALL	2006	PS2	1500
16	ALL	2006	Wii	500
17	ALL	2006	ALL	2000
18	ALL	ALL	PS2	2900
19	ALL	ALL	Wii	500
20	ALL	ALL	XBox360	1700
21	ALL	ALL	ALL	5100

Question 1 (2)

```
SELECT Location, Time, Item, SUM(Quantity)
FROM Sales
GROUP BY Location, Time, Item

UNION ALL

SELECT Location, Time, NULL, SUM(Quantity)
FROM Sales
GROUP BY Location, Time

UNION ALL

SELECT Location, NULL, Item, SUM(Quantity)
FROM Sales
GROUP BY Location, Item

UNION ALL

SELECT Location, NULL, NULL, SUM(Quantity)
FROM Sales
GROUP BY Location

UNION ALL

SELECT NULL, Time, Item, SUM(Quantity)
FROM Sales
GROUP BY Time, Item

UNION ALL
```

```
SELECT NULL, Time, NULL, SUM(Quantity)
FROM Sales
GROUP BY Time

UNION ALL

SELECT NULL, NULL, Item, SUM(Quantity)
FROM Sales
GROUP BY Item

UNION ALL

SELECT NULL, NULL, NULL, SUM(Quantity)
FROM Sales

ORDER by Location, Time, Item
```

Question 1 (3)

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

Question 1 (4)

The original output:

	Location	Time	Item	Quantity
0	0	0	0	5100
1	0	0	1	2900
2	0	0	2	1700
3	0	0	3	500
4	0	1	0	3100
5	0	1	1	1400
6	0	1	2	1700
7	0	2	0	2000
8	0	2	1	1500
9	0	2	3	500
10	1	0	0	3400
11	1	0	1	2900
12	1	0	3	500
13	1	1	0	1400
14	1	1	1	1400
15	1	2	0	2000
16	1	2	1	1500
17	1	2	3	500
18	2	0	0	1700
19	2	0	2	1700
20	2	1	0	1700
21	2	1	2	1700

The offset function = 3 * 4 * location + 4 * time + item
Then we get the MOLAP cube in tabular form:

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

The multi-dimensional array of 3D mode:

Slice of Time == 0

		Item			
		0	1	2	3
Location	0	5100	2900	1700	500
	1	3400	2900	-	500
	2	1700	-	1700	-

Slice of Time == 1

		Item			
		0	1	2	3
Location	0	3100	1400	1700	-
	1	1400	1400	-	-
	2	1700	-	1700	-

Slice of Time == 2

		Item			
		0	1	2	3
Location	0	2000	1500	-	500
	1	2000	1500	-	500
	2	-	-	-	-

Question 2 (a)

According to Bayes Rule, we have

$$P(y=1|\mathbf{x})=\frac{P(\mathbf{x}|y=1)P(y=1)}{P(\mathbf{x})} \quad \text{and} \quad P(y=0|\mathbf{x})=\frac{P(\mathbf{x}|y=0)P(y=0)}{P(\mathbf{x})}$$

where $\mathbf{x}$ represents $\langle x_{1,} x_{2,} \ldots, x_d \rangle$.

Also, the Naive Bayes classifier $NB(\mathbf{x})$ provides that:

$$NB(\mathbf{x})=\begin{cases} 1, & \frac{P(y=1|\mathbf{x})}{P(y=0|\mathbf{x})}\geq 1 \\ 0, & \frac{P(y=1|\mathbf{x})}{P(y=0|\mathbf{x})}<1 \end{cases}$$

So, we have to determine value of $\frac{P(y=1|\mathbf{x})}{P(y=0|\mathbf{x})}$.

According to Naive Bayes, we can find out:

$$\begin{aligned}
&\frac{P(y=1|\mathbf{x})}{P(y=0|\mathbf{x})} \\
&=\frac{P(\mathbf{x}|y=1)P(y=1)}{P(\mathbf{x}|y=0)P(y=0)} \\
&=\frac{P(y=1)\prod_{i=1}^{d} P(x_i|y=1)}{P(y=0)\prod_{i=1}^{d} P(x_i|y=0)} \\
&=\frac{P(y=1)}{P(y=0)}\prod_{i=1}^{d}\frac{P(x_i|y=1)}{P(x_i|y=0)}
\end{aligned}$$

Let $p=P(y=1)$, then $1-p=P(y=0)$;

Let $a_i=P(x_i=1|y=1)$, then $1-a_i=P(x_i=0|y=1)$, then $P(x_i|y=1)=a_i^{x_i}(1-a_i)^{1-x_i}$;

Let $b_i=P(x_i=1|y=0)$, then $1-b_i=P(x_i=0|y=0)$, then $P(x_i|y=0)=b_i^{x_i}(1-b_i)^{1-x_i}$.

Then, $\frac{P(y=1|\mathbf{x})}{P(y=0|\mathbf{x})}=\frac{p}{1-p}\prod_{i=1}^{d}\frac{a_i^{x_i}(1-a_i)^{1-x_i}}{b_i^{x_i}(1-b_i)^{1-x_i}}$.

Then,

$$\begin{aligned}
&\log\frac{P(y=1|\mathbf{x})}{P(y=0|\mathbf{x})} \\
&=\log\frac{p}{1-p}+\sum_{i=1}^{d}\log\frac{a_i^{x_i}(1-a_i)^{1-x_i}}{b_i^{x_i}(1-b_i)^{1-x_i}} \\
&=\log\frac{p}{1-p}+\sum_{i=1}^{d}\log\frac{1-a_i}{1-b_i}+\sum_{i=1}^{d}x_i\log\frac{a_i(1-b_i)}{b_i(1-a_i)}
\end{aligned}$$

.

$$\text{Let } b=\log\frac{p}{1-p}+\sum_{i=1}^{d}\log\frac{1-a_i}{1-b_i} \;;$$

$$\text{Let } \boldsymbol{w}=\begin{pmatrix} w_1 \\ w_2 \\ \ldots \\ w_d \end{pmatrix} \text{ where } w_i=\log\frac{a_i(1-b_i)}{b_i(1-a_i)} \text{ , then}$$

$$\log\frac{P(y=1|\boldsymbol{x})}{P(y=0|\boldsymbol{x})}=b+\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x}$$

So, the Naive Bayes classifier can be transformed into a Linear classifier:

$$NB(\boldsymbol{x})=\begin{cases} 1, & b+\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x}\geq 0 \\ 0, & b+\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x}<0 \end{cases}$$

Question 2 (b)

Naive Bayes is more straightforward than Logistic Regression because Naive Bayes can finish training in linear time on behalf of size of its data-set. Furthermore, Naive Bayes is also simple to do predictions since it only needs to pick up trained values directly for its prediction equation like

$$\frac{P(y=1|\boldsymbol{x})}{P(y=0|\boldsymbol{x})}=\frac{P(y=1)}{P(y=0)}\prod_{i=1}^{d}\frac{P(x_i|y=1)}{P(x_i|y=0)}$$

In contrast, Logistic Regression needs more sophisticated training process like Gradient Ascent. The time complexity of Gradient Ascent $\boldsymbol{w}\leftarrow\boldsymbol{w}+\eta\frac{\partial\ln L(\boldsymbol{w})}{\partial w_j}$ is sometimes unpredictable.

Question 3 (1)

According to some other academic materials, doing Gradient Ascent of $\boldsymbol{w}$ on its likelihood function is equivalent to doing Gradient Descent of $\boldsymbol{w}$ on its loss function.

The likelihood function is $L(\boldsymbol{w})=\prod_{i=1}^{n}\left[p(\boldsymbol{x_i})^{y_i}(1-p(\boldsymbol{x_i}))^{1-y_i}\right]$.

Then log of likelihood function is:

$$\ln L(\boldsymbol{w})=\sum_{i=1}^{n}\left[y_i\ln\left(p(\boldsymbol{x_i})\right)+(1-y_i)\ln\left(1-p(\boldsymbol{x_i})\right)\right] .$$

Provided that $p(\boldsymbol{x_i})=\sigma(\boldsymbol{w^T}\boldsymbol{x_i})=\frac{1}{1+e^{-\boldsymbol{w}^T\boldsymbol{x}_i}}$, then

$$\begin{aligned}
&\ln L(\boldsymbol{w})\\
&=\sum_{i=1}^{n}\left[y_i\ln\frac{1}{1+e^{-\boldsymbol{w}^T\boldsymbol{x}_i}}+(1-y_i)\ln\left(1-\frac{1}{1+e^{-\boldsymbol{w}^T\boldsymbol{x}_i}}\right)\right]\\
&=\sum_{i=1}^{n}\left[y_i\ln\frac{e^{\boldsymbol{w}^T\boldsymbol{x}_i}}{1+e^{\boldsymbol{w}^T\boldsymbol{x}_i}}+(1-y_i)\ln\frac{1}{1+e^{\boldsymbol{w}^T\boldsymbol{x}_i}}\right]\\
&=\sum_{i=1}^{n}\left[y_i\boldsymbol{w^T}\boldsymbol{x_i}-\ln(1+e^{\boldsymbol{w}^T\boldsymbol{x}_i})\right]
\end{aligned}$$

Therefore, to get maximum of $L(\boldsymbol{w})$ is equivalent to applying minimum of

$$-\ln L(\boldsymbol{w})=\sum_{i=1}^{n}\left[-y_i\boldsymbol{w^T}\boldsymbol{x_i}+\ln(1+e^{\boldsymbol{w}^T\boldsymbol{x}_i})\right]$$

By the way, to do Gradient Ascent $\boldsymbol{w}\leftarrow\boldsymbol{w}+\eta\frac{\partial\ln L(\boldsymbol{w})}{\partial w_j}$ is equivalent to do Gradient Descent:

$$\boldsymbol{w}\leftarrow\boldsymbol{w}-\eta\frac{\partial\left(-\ln L(\boldsymbol{w})\right)}{\partial w_j}$$

Question 3 (2)

The likelihood function is $L(\boldsymbol{w})=\prod_{i=1}^{n}\left[p(\boldsymbol{x_i})^{y_i}(1-p(\boldsymbol{x_i}))^{1-y_i}\right]$.

Log of likelihood function is:

$$\ln L(\boldsymbol{w})=\sum_{i=1}^{n}\left[y_i\ln\left(p(\boldsymbol{x_i})\right)+(1-y_i)\ln\left(1-p(\boldsymbol{x_i})\right)\right]$$

Therefore, to get maximum of $L(\boldsymbol{w})$ is equivalent to applying minimum of

$$l(\boldsymbol{w})=-\ln L(\boldsymbol{w})=-\sum_{i=1}^{n}\left[y_i\ln\left(p(\boldsymbol{x_i})\right)+(1-y_i)\ln\left(1-p(\boldsymbol{x_i})\right)\right]$$

where $p(\boldsymbol{x_i})=f(\boldsymbol{w^T}\boldsymbol{x_i})$.