COMP9318 Assignment 1

Name: Keshi Chen

ZID: z5142821

Q1.

(1)

	Location	Time	Item	Quantity
0	Sydney	2005	PS2	1400
1	Sydney	2005	ALL	1400
2	Sydney	2006	PS2	1500
3	Sydney	2006	Wii	500
4	Sydney	2006	ALL	2000
5	Sydney	ALL	PS2	2900
6	Sydney	ALL	Wii	500
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	PS2	1400
13	ALL	2005	Xbox 360	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	Xbox 360	1700

21	ALL	ALL	ALL	5100

(2)

```
select location, time, item, sum(quantity)
from test
group by location, time, item
union
select location, time, 'ALL', sum(quantity)
from test
group by location, time
union
select location, 'ALL', item, sum(quantity)
from test
group by location, item
union
select location, 'ALL', 'ALL', sum(quantity)
from test
group by location
union
select 'ALL', time, item, sum(quantity)
from test
group by  time, item
union
select 'ALL', time, 'ALL', sum(quantity)
from test
group by  time
union
select 'ALL', 'ALL', item, sum(quantity)
from test
group by  item
```

(3)

	Location	Time	Item	Quantity
0	Sydney	2005	PS2	1400

1	Sydney	2005	ALL	1400
2	Sydney	2006	PS2	1500
3	Sydney	2006	ALL	1500
4	Sydney	ALL	PS2	2900
5	Sydney	ALL	ALL	2900
6	ALL	2005	PS2	1400
7	ALL	2005	ALL	1400
8	ALL	2006	PS2	1500
9	ALL	2006	ALL	1500
10	ALL	ALL	PS2	2900
11	ALL	ALL	ALL	2900

(4) f(Location, Time, Item) = 12 * Location + 4 * Time + Item

The process can be represented in 3 steps as below:

Step 1:

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

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

Step 2: f(Location, Time, Item) = 12 * Location + 4 * Time + Item

	Location	Time	Item	Quantity	ArrayIndex
0	1	1	1	1400	17
1	1	1	0	1400	16
2	1	2	1	1500	21
3	1	2	3	500	23
4	1	2	0	2000	20
5	1	0	1	2900	13
6	1	0	3	500	15
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	1	1400	5
13	0	1	2	1700	6
14	0	1	0	3100	4
15	0	2	1	1500	9
16	0	2	3	500	11
17	0	2	0	2000	8
18	0	0	1	2900	1
19	0	0	3	500	3
20	0	0	2	1700	2
21	0	0	0	5100	0

Step 3:

ArrayIndex	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

Q2

(1) The Naïve Bayes Classifier: $f(x) = \text{argmax}_{x \in \{C_j\}} \prod_{i=1} P(\alpha_i|C_j) * P(C_j)$

Since $C_j \in \{0, 1\}$, $\alpha_i \in \{0, 1\}$, f(x) can be represented as:

$$f(x) = \begin{cases} 0, \text{ if } \prod_{i=1} P(\alpha_i| C_0) * P(C_0) - \prod_{i=1} P(\alpha_i| C_1) * P(C_1) > 0 \\ 1, \text{ if otherwise} \end{cases}$$

Then with $\log\prod_i x_i = \sum_i \log x_i$, we have:

$f(x) = \sum_i \log \frac{P(\alpha i|C0)}{P(\alpha i|C1)} + \log\frac{P(C0)}{1-P(C0)}$

Let $w_i = \log \frac{P(\alpha i|C0)}{P(\alpha i|C1)}$, $w_0 = \log\frac{P(C0)}{1-P(C0)}$, $i \in \{1, 2, 3, \dots, d\}$

Then Naïve Bayes Classifier is equal to a binary linear classifier in d+1 dimensional space with input x_i either be 0 or 1.

(2) We have to do partial derivatives to minimize the cost function and gradient descents to get w_i in Logistic Regression Classifier, thus its calculation is complex.

In Naïve Bayes Classifier, we only need to calculate each $P(x_i|C)$ once to get the w_i.

So Naïve Bayes Classifier is much easier than Logistic Regression Classifier.

Q3

(1) Logistic regression model:

$\delta(x) = \frac{1}{1+\exp(-wTx)}$ ①

For one training sample (x_i , y_i), where $y_i \in \{0, 1\}$, we have the probability of y_i:

$P(y_i | x_i , w) = \delta(x)^{y_i} (1 - \delta(x))^{1-y_i}$

Hence when $y_i = 1$, $P(y_i | x_i , w) = \delta(x)$, otherwise $P(y_i | x_i , w) = (1 - \delta(x))$.

Then we have the log-likelihood:

$\text{Log-likelihood} = \ln(P(y_i | x_i , w)) = \ln(\delta(x_i)^{y_i} (1 - \delta(x_i))^{1-y_i})$

$= \ln\delta(x_i)^{y_i} + \ln(1 - \delta(x_i))^{1-y_i}$

$= y_i \ln\delta(x_i) + (1 - y_i) \ln(1 - \delta(x_i))$

For N training samples, we have:

$\text{Log-likelihood} = \sum_{i=1} y_i \ln\delta(x_i) + (1 - y_i) \ln(1 - \delta(x_i))$, $i \in [1, N]$ ②

And since $\delta(x_i) = \frac{1}{1+\exp(-wTxi)} = \frac{\exp(wTxi)}{1+\exp(wTxi)}$ ③

From ② ③we have:

$\text{Log-likelihood} = \sum_{i=1} (y_i w^T x_i - \ln(1 +\exp(w^T x_i)))$, $i \in [1, N]$ ④

Since loss function = - log-likelihood, we have:

$l(w) = \sum_{i=1} (- y_i w^T x_i + \ln(1 + \exp(w^T x_i))) , i \in [1, N]$

Q.E.D

(2) Similarly, $P(y_i | x_i , w) = f(w^T x_i)^{y_i} (1 - f(w^T x_i))^{1-y_i}$

$\ln P(y_i | x_i , w) = y_i \ln f(w^T x_i) + (1 - y_i) (1 - f(w^T x_i))$

For N samples, we have:

$\text{Log-likelihood} = \sum_{i=1} y_i \ln f(w^T x_i) + (1 - y_i) \ln(1 - f(w^T x_i)) , i \in [1, N]$

Hence the loss function:

$l(w) = - \text{log-likelihood} = \sum_{i=1} - y_i \ln f(w^T x_i) - (1 - y_i) \ln(1 - f(w^T x_i)) , i \in [1, N]$