9318----Assignment

Jia SUN z5145482

Question 1

1)

Location	Time	Item	SUM(Quantity)
Sydney	2005	PS2	1400
Sydney	2005	ALL	1400
Sydney	2006	PS2	1500
Sydney	2006	Wii	500
Sydney	2006	ALL	2000
Sydney	ALL	PS2	2900
Sydney	ALL	Wii	500
Sydney	ALL	ALL	3400
Melbourne	2005	XBox360	1700
Melbourne	2005	ALL	1700
Melbourne	ALL	XBox360	1700
Melbourne	ALL	ALL	1700
ALL	2005	PS2	1400
ALL	2005	XBox360	1700
ALL	2005	ALL	3100
ALL	2006	PS2	1500
ALL	2006	Wii	500
ALL	2006	ALL	2000
ALL	ALL	PS2	2900
ALL	ALL	Wii	500
ALL	ALL	XBox360	1700
ALL	ALL	ALL	5100

2.)

```
(SELECT Location, Year, Item, Quantity
    FROM Sales)
UNION
(SELECT Location, Year, 'ALL', SUM(Quantity)
    FROM Sales
    GROUP BY(Location, Year))
UNION
(SELECT Location, 'ALL', Item, SUM(Quantity)
    FROM Sales
    GROUP BY(Location, Item))
UNION
(SELECT 'ALL', Year, Item, SUM(Quantity)
    FROM Sales
    GROUP BY(Year, Year))
UNION
(SELECT Location, 'ALL', 'ALL', SUM(Quantity)
    FROM Sales
    GROUP BY(Location))
UNION
(SELECT 'ALL', Year, 'ALL', SUM(Quantity)
    FROM Sales
    GROUP BY(Year))
UNION
(SELECT 'ALL', 'ALL', Item, SUM(Quantity)
    FROM Sales
    GROUP BY(Item))
UNION
(SELECT 'ALL', 'ALL', 'ALL', SUM(Quantity)
    FROM Sales)
```

3)

Location	Time	Item	SUM(Quantity)
Sydney	2006	ALL	2000
Sydney	ALL	PS2	2900
ALL	2005	ALL	3100
ALL	2006	ALL	2000
ALL	ALL	PS2	2900
ALL	ALL	ALL	5100

4)

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

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

Question 2

1)

By multinomial Naïve Bayes functions:

①P($X_1 = x_1,\ X_2 = x_2,\ X_3 = x_3, \ldots\ldots,\ X_n = x_n) = \frac{k!}{X_1!X_2!X_3!\ldots\ldots X_n!} P1^{x_1}P2^{x_2} \ldots\ldots Pn^{x_n}$

y∈{0, 1}

Log likelihood and we can get the decision boundary:

②$\log\left(\frac{P(Y=1|X=x)}{P(Y=0|X=x)}\right)$=0

By combining ① and ② $\sum_{i=1}^{d} \log\left(\frac{P(X_i|y=1)}{P(X_i|y=0)}\right) + \log\left(\frac{P(y=1)}{P(y=0)}\right) = 0$

Therefore, we can get:

$$\sum_{i=1}^{d} \log\left(\frac{P_i(y=1|X_i)}{P_i(y=0|X_i)}\right) * X_i + \log\left(\frac{P(y=1)}{P(y=0)}\right) = 0$$

We can get $\vec{W}$ easily

$$\overrightarrow{Wnb} = (\log\left(\frac{P_1(y=1|X_1)}{P_1(y=0|X_1)}\right), \log\left(\frac{P_2(y=1|X_2)}{P_2(y=0|X_2)}\right), \ldots\ldots, \log\left(\frac{P_n(y=1|X_d)}{P_n(y=0|X_d)}\right))$$

Then, we can get a linear form($W^T X + \mathrm{b} = 0$) for multinomial Naïve Bayes Classifier.

And $\overrightarrow{Wnb}$ can be got by the equation directly.

2)

For Logistic Regression

y = $\frac{1}{1+e^{-w^T X}}$

P($y_i = 1|X_i, w$) = $\frac{e^{w^T Xi}}{1+e^{-w^T Xi}}$

P($y_i = 0|X_i, w$) = 1 - P($y_i = 1|X_i, w$)

Therefore,

P(X_i, y_i) = $\mathrm{P}(y_i = 1|X_i, w)^{y_i}(1 - \mathrm{P}(y_i = 1|X_i, w))^{1-y_i}$

Likelihood : L(w) = $\prod \mathrm{P}(y_i = 1|X_i, w)^{y_i}(1 - \mathrm{P}(y_i = 1|X_i, w))^{1-y_i}$

The log – Likelihood is

l(w) = log($\prod \mathrm{P}(y_i = 1|X_i)^{y_i}(1 - \mathrm{P}(y_i = 1|X_i))^{1-y_i}$)l

= $\sum_{i=1}^{n} y_i \log\mathrm{P}(y_i = 1|X_i) + (1 - y_i)\log(1 - \mathrm{P}(y_i = 1|X_i))$

= $\sum_{i=1}^{n} y_i \log\mathrm{P}(y_i = 1|X_i) - y_i\log(1 - \mathrm{P}(y_i = 1|X_i) + \log(1 - \mathrm{P}(y_i = 1|X_i))$

= $\sum_{i=1}^{n} y_i \frac{\log\mathrm{P}(y_i = 1|X_i)}{\log(1-\mathrm{P}(y_i = 1|X_i)} + \sum_{i=1}^{d} \log(1 - \mathrm{P}(y_i = 1|X_i))$

= $\sum_{i=1}^{n} (y_i - \log(1 + e^{w^T Xi}))$

So, we can get $\frac{\alpha l(w)}{\alpha w} = \sum_{i=1}^{n} \left(y_i - \frac{1}{1+e^{-w^T Xi}}\right) * X_i$

Obviously, we cannot get w^T directly.

To get maximum likelihood, we can use gradient descent and iterate many times to re-weight.

We also need a appropriate learning rate η to make $\overrightarrow{W_{RL}^{++1}}$ = $\overrightarrow{W_{RL}^{+}}$ - $\eta \frac{\alpha l(w)}{\alpha w}$, $\overrightarrow{Wnb}$ is easy to learn.

Question 3

1)

By using logistic regression:

Likelihood : L(w) = $\prod \mathrm{P}(x^i)^{y^i}(1-\mathrm{P}(x^i))^{1-y^i}$

The log – Likelihood is

l(w) = $\sum_{i=1}^{n} y^i \ln\left(p\left(x^i\right)\right) + (1-y^i)\ln\left(1-p(x^i)\right)$

Because the loss function,

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

$= -\sum_{i=1}^{n}[y^i \ln\left(p(x^i)\right) - ln\left(1-p\left(x^i\right)\right) + ln\left(1-p\left(x^i\right)\right)]$

$= -\sum_{i=1}^{n}[y^i \ln\left(\frac{p(x^i)}{1-p(x^i)}\right) + ln\left(1-p\left(x^i\right)\right)]$

As we know $p\left(x^i\right) = \sigma(w^T X)$, So $p\left(x^i\right) = 1 + \frac{1}{e^{-w^T X}}$

l(w) = $-\sum_{i=1}^{n}[y^i ln\left(\frac{1}{e^{-w^T X}}\right) + ln\left(\frac{e^{-w^T X}}{1+e^{-w^T X}}\right)]$

$= -\sum_{i=1}^{n}[y^i(ln(1) - ln\left(e^{-w^T X}\right)) + ln\left(\frac{1}{1+e^{-w^T X}}\right)]$

$= -\sum_{i=1}^{n}[y^i w^T X + \ln(1) - ln\left(e^{w^T X} + 1\right)]$

$=-\sum_{i=1}^{n}[y^i w^T X - ln\left(e^{w^T X} + 1\right)]$

$=\sum_{i=1}^{n}[-y^i w^T X + \ln\left(1 + exp\left(e^{w^T X}\right)\right)]$

2)

By using logistic regression:

Likelihood : L(w) = $\prod \mathrm{P}(x^i)^{y^i}(1-\mathrm{P}(x^i))^{1-y^i}$

The log – Likelihood is

l(w) = $\sum_{i=1}^{n} y^i \ln\left(p\left(x^i\right)\right) + (1-y^i)\ln\left(1-p(x^i)\right)$

Due to the loss function,

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

$= -\sum_{i=1}^{n}[y^i \ln\left(p(x^i)\right) - ln\left(1-p\left(x^i\right)\right) + ln\left(1-p\left(x^i\right)\right)]$

$= -\sum_{i=1}^{n}[y^i \ln\left(\frac{p(x^i)}{1-p(x^i)}\right) + ln\left(1-p\left(x^i\right)\right)]$

Because $p\left(x^i\right) = f(w^T X i)$

So the deduced loss function is -$\sum_{i=1}^{n}[y^i \ln(\frac{f(w^T Xi)}{1-f(w^T Xi)}) + ln(1-f(w^T Xi))]$