COMP9318 Assignment 1

Zid: 5143329

Name: Xinyu Wang

1(1)

Location	Time	Item	SVM(Quantity)
Sydney	2005	PS2	1400
Sydney	2005	ALL	1400
Sydney	2005	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

1(2)

SELECT LOCATION,TIME, ITEM, SUM(QUANTITY)

FROM SALE S

GROUP BY LOCATION,TIME, ITEM

UNION

SELECT LOCATION,NULL,NULL, SUM(QUANTITY)

FROM SALE S

GROUP BY LOCATION

UNION

SELECT NULL,NULL, ITEM, SUM(QUANTITY)

FROM SALE S

GROUP BY ITEM

UNION

SELECT NULL,TIME, NULL, SUM(QUANTITY)

FROM SALE S

```
GROUP BY TIME
UNION

SELECT NULL,TIME, ITEM,   SUM(QUANTITY)
FROM SALE S
GROUP BY TIME, ITEM
UNION

SELECT LOCATION,NULL, ITEM,   SUM(QUANTITY)
FROM SALE S
GROUP BY LOCATION, ITEM
UNION

SELECT LOCATION,TIME, NULL,   SUM(QUANTITY)
FROM SALE S
GROUP BY LOCATION,TIME
UNION

SELECT NULL,NULL, NULL,   SUM(QUANTITY)
FROM SALE S
```

1(3):

Location	Time	Item	SVM(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

1(4):

F(Location, Time, Item) = 15*Location + 10*Time + 7*Item

Offset	SUM(Quantity)
32	1400
25	1400
42	1500
56	500
35	2000
22	2900

36	500
15	3400
54	1700
40	1700
44	1700
30	1700
17	1400
24	1700
10	3100
27	1500
41	500
20	2000
7	2900
21	500
14	1700
0	5100

2.

(1)

Since x is classified to label **y** = 1 according to Naïve Bayes classifier,

$$\mathrm{P}(\mathrm{y}=1)\cdot\prod_{i=1}^{d}\mathrm{P}(\mathrm{x}_i\mid \mathrm{y}=1)-\mathrm{P}(\mathrm{y}=0)\cdot\prod_{i=1}^{d}\mathrm{P}(\mathrm{x}_i\mid \mathrm{y}=0)>0$$

Use presentation as this:

$\mathrm{P}(\mathrm{y}=1)=p$

$\mathrm{P}(\mathrm{x}_i=1\mid \mathrm{y}=1)=a_i$

$\mathrm{P}(\mathrm{x}_i=1\mid \mathrm{y}=0)=b_i$

$\mathrm{P}(\mathrm{y}=0)=1-p$

$\mathrm{P}(\mathrm{x}_i\mid \mathrm{y}=1)= {a_i}^{x_i}\cdot(1-a_i)^{(1-x_i)}$

$\mathrm{P}(\mathrm{x}_i\mid \mathrm{y}=0)= {b_i}^{x_i}\cdot(1-b_i)^{(1-x_i)}$

According to functions above:

$$p\cdot\prod_{i=1}^{d}{a_i}^{x_i}\cdot(1-a_i)^{(1-x_i)}-(1-p)\cdot\prod_{i=1}^{d}{b_i}^{x_i}\cdot(1-b_i)^{(1-x_i)}>0$$

We can know that:

$$p\cdot\prod_{i=1}^{d}\cdot\frac{{a_i}^{x_i}\cdot(1-a_i)^1}{(1-a_i)^{x_i}}-(1-p)\cdot\prod_{i=1}^{d}\cdot\frac{{b_i}^{x_i}\cdot(1-b_i)^1}{(1-b_i)^{x_i}}>0$$

$$=p\cdot\prod_{i=1}^{d}\cdot\left(\frac{a_i}{1-a_i}\right)^{x_i}\cdot(1-a_i)-(1-p)\cdot\prod_{i=1}^{d}\cdot\left(\frac{b_i}{b_i}\right)^{x_i}\cdot(1-b_i)>0$$

$$\to\qquad \ln p+\sum_{i=1}^{d}\ln(1-a_i)+x_i\cdot\sum_{i=1}^{d}\ln\frac{a_i}{(1-a_i)}-\ln(1-p)-\sum_{i=1}^{d}\ln(1-b_i)-x_i\cdot\sum_{i=1}^{d}\ln\frac{b_i}{(1-b_i)}>0$$

So, we have:

$$\ln\frac{p}{1-p}+\sum_{i=1}^{d}\ln\frac{(1-a_i)}{(1-b_i)}+\sum_{i=1}^{d}x_i\cdot\ln\left(\frac{a_i}{(1-a_i)}\cdot\frac{(1-b_i)}{b_i}\right)>0\quad\text{---(1)}$$

We know that p, a_i, b_i are constants.

So, function (1) is like $w_{d+1}+\sum_{i=1}^{d}w_i\cdot x_i>0$

So we reached the conclusion.

2(2):

For logistic regression classifier, when X happen, it means

$$\mathbf{L}(\mathbf{w}_{LR}) = \prod_{i=1}^{d} \boldsymbol{p}(x_i; \boldsymbol{w}_i)$$

So, $\ln \mathbf{L}(\mathbf{w}_{LR})) = \sum_{i=1}^{d} \boldsymbol{p}(x_i; \boldsymbol{w}_i)$

In order to maximize $\ln \mathbf{L}(\mathbf{w}_{LR}))$**,** we have to do d times of partial deviratives to take each $\boldsymbol{w}_i\,'\boldsymbol{s}$ partial deviratives.
So, it is obviously much more difficult than getting $\mathbf{w}_{NB}$

3.

(1)

Since σ is Sigmoid function:

$$\mathrm{P}(\mathrm{y} = 1 \mid \mathbf{x}) = \sigma(\mathbf{w}^T\mathbf{x}) = \frac{1}{1+e^{-\mathrm{w}^T\mathrm{x}}}$$

So:

$$\mathrm{P}(\mathbf{y} \mid \mathbf{x}) = \sigma(\mathbf{w}^T\mathbf{x})^{\mathrm{y}} \cdot (1 - \sigma(\mathbf{w}^T\mathbf{x}))^{1-\mathrm{y}}$$

We can know the likelihood is:

$$\mathrm{L}(\mathbf{w}) = \prod_{i=1}^{n} \mathrm{P}\left(\mathrm{y}^{(i)} \,\middle|\, \mathrm{x}^{(i)}; \mathbf{w}\right)$$
$$= \prod_{\mathrm{i}=1}^{\mathrm{n}} \sigma\left(\mathrm{w}^{\mathrm{T}}\, \mathrm{x}^{(\mathrm{i})}\right)^{\mathrm{y}^{(\mathrm{i})}} \cdot (1 - \sigma\left(\mathrm{w}^{\mathrm{T}}\, \mathrm{x}^{(\mathrm{i})}\right))^{1-\mathrm{y}^{(\mathrm{i})}}$$

So the loss function of logistic regression is:

$$l(\mathbf{w}) = -\ln \mathrm{P}(\mathbf{y} \mid \mathbf{x})$$

$$= -ln\prod_{i=1}^{n}\sigma\left(w^T\, x^{(i)}\right)^{y^{(i)}}\cdot(1-\sigma\left(w^T\, x^{(i)}\right))^{1-y^{(i)}}$$

$= (-y_i)\cdot\sum_{i=1}^{n}\cdot ln\,\sigma\left(w^T\, x^{(i)}\right) - (1-y_i)\cdot\ ln(1-$
$\sigma\left(w^T\, x^{(i)}\right))$

$= \ (-y_i)\cdot\sum_{i=1}^{n}\cdot ln\,\sigma\left(w^T\, x^{(i)}\right) - \ \ ln(1-\sigma\left(w^T\, x^{(i)}\right)) +\ y_i\cdot$
$ln(1-\sigma\left(w^T\, x^{(i)}\right))$

$= \ \ (-y_i)\cdot\left(\sum_{i=1}^{n}\cdot ln\,\sigma\left(w^T\, x^{(i)}\right) - \ ln\left(1-\sigma\left(w^T\, x^{(i)}\right)\right)\right) -$
$ln(1-\sigma\left(w^T\, x^{(i)}\right))$

$$= \ \ (-y_i)\sum_{i=1}^{n}\cdot ln\frac{\sigma\left(w^T\, x^{(i)}\right)}{(1-\sigma\left(w^T\, x^{(i)}\right))} - \ ln(1-\sigma\left(w^T\, x^{(i)}\right))$$

According to the sigmoid function:

$\mathrm{P}(\mathrm{y} = 1 \mid \mathbf{x}) = \ \sigma(\mathbf{w}^T\mathbf{x}) = \ \frac{1}{1+e^{-\mathbf{w}^T\mathbf{x}}}$

The function becomes:

$$= (-y_i)\cdot(\sum_{i=1}^{n}\cdot ln\frac{\frac{1}{1+e^{-w^Tx}}}{1-\frac{1}{1+e^{-w^Tx}}} - \ ln(1-\frac{1}{1+e^{-w^Tx}})$$

$$= (-y_i)\cdot\sum_{i=1}^{n}\cdot ln\frac{1}{e^{-w^Tx}+1-1} - \ ln(1-\frac{1}{1+e^{-w^Tx}})$$

$$= (-y_i)\cdot\sum_{i=1}^{n}\cdot ln\frac{1}{e^{-w^Tx}} - \ ln(1-\frac{1}{1+e^{-w^Tx}})$$

$$= (-y_i)\cdot\sum_{i=1}^{n}\cdot lne^{w^Tx} - \ ln(\frac{1+e^{-w^Tx}-1}{1+e^{-w^Tx}})$$

$$= (-y_i)\cdot\sum_{i=1}^{n}\cdot lne^{w^Tx} - \ ln(\frac{1}{1+e^{w^Tx}})$$

$$= (-y_i)\cdot\sum_{i=1}^{n}\cdot lne^{w^Tx} + \ ln(1+e^{w^Tx})$$

$$= \ \sum_{i=1}^{n} -y_i\cdot w^Tx + \ ln(1+e^{w^Tx})$$

(2)

Since

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

So,

$$P(\mathbf{y} \mid \mathbf{x}) = f(\mathbf{w}^T\mathbf{x})^{\mathbf{y}} \cdot (1 - f(\mathbf{w}^T\mathbf{x}))^{1-\mathbf{y}}$$

The likelihood is:

$$L(\mathbf{w}) = \prod_{i=1}^{n} P\left(y^{(i)} \mid x^{(i)}; \mathbf{w}\right)$$

$$= \prod_{i=1}^{n} f\left(w^T x^{(i)}\right)^{y^{(i)}} \cdot (1 - f\left(w^T x^{(i)}\right))^{1-y^{(i)}}$$

So, the loss function of logistic regression is:

$$l(\mathbf{w}) = -\ln P(\mathbf{y} \mid \mathbf{x})$$

$$= -\ln \prod_{i=1}^{n} f\left(w^T x^{(i)}\right)^{y^{(i)}} \cdot (1 - f\left(w^T x^{(i)}\right))^{1-y^{(i)}}$$

$$= \quad (-y_i) \cdot \textstyle\sum_{i=1}^{n} \cdot \ln f\left(w^T x^{(i)}\right) + (1 - y_i) \cdot \ln(1 - f\left(w^T x^{(i)}\right))$$