9318 Assignment

Name: FUDI ZHENG

ID:5126586

Question 1:

1.

0	location	time	item	quantity
1	Sydney	2005	PS2	1400
2	Sydney	2006	PS2	1500
3	Sydney	2006	Wii	500
4	Sydney	2005	ALL	1400
5	Sydney	2006	ALL	2000
6	Sydney	ALL	PS2	2900
7	Sydney	ALL	Wii	500
8	Sydney	ALL	ALL	3400
9	Melbourne	2005	XBox 360	1700
10	Melbourne	2005	ALL	1700
11	Melbourne	ALL	XBox 360	1700
12	Melbourne	ALL	ALL	1700
13	ALL	2005	PS2	1400
14	ALL	2006	PS2	1500
15	ALL	2006	Wii	500
16	ALL	2005	XBox 360	1700
17	ALL	ALL	PS2	2900
18	ALL	ALL	Wii	500
19	ALL	ALL	XBox 360	1700
20	ALL	2005	ALL	3100

21	ALL	2006	ALL	2000
22	ALL	ALL	ALL	5100

2.

```
SELECT Location, Time, Item, SUM(Quantity)
From sales
Group By Location, Time, Item with rollup
Union
SELECT Location, Time, Item, SUM(Quantity)
From sales
Group By Item, Location, Time with rollup
Union
SELECT Location, Time, Item, SUM(Quantity)
From sales
Group By Time, Item, Location with rollup
```

3.

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

4.

$f(\text{Location}, \text{Time}, \text{Item}) = (3 * \text{Location} + \text{Time}) * 4 + \text{Item}$

Step 1:

Location	Time	Item	Quantity
1	1	1	1400
1	2	1	1500
1	2	3	500
2	1	2	1700
1	1	0	1400
1	2	0	2000
2	1	0	1700
1	0	1	2900
1	0	3	500
2	0	2	1700
0	1	1	1400
0	2	1	1500
0	2	3	500
0	1	2	1700
0	0	1	2900
0	0	3	500
0	0	2	1700
0	1	0	3100
0	2	0	2000
1	0	0	3400
2	0	0	1700
0	0	0	5100

Step 2:

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

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

MD Array
5100
2900
1700
500
3100
1400
1700
2000
1500
500
3400
2900
500
1400
1400
2000
1500
500
1700
1700
1700
1700

Question 2:

1.

The naive bayes classifier is :

$$f(x) = argmax_{x \epsilon \{C_j\}} \prod_{i=1}^{n} P(\, k_i | C_j) \, \cdot P(C_j)$$

Each feature only have two value(0 and 1). when the feature value is 0:

$$f(x=0) = \prod_{i=1}^{n} P(k_i | C_0) \, \cdot P(C_0)$$

Each feature only have two value(0 and 1). when the feature value is 1:

$$f(x=1) = \prod_{i=1}^{n} P(k_i | C_1) \, \cdot P(C_1)$$

so $f(x) = f(x=0) - f(x=1)$

$$f(x) = \prod_{i=1}^{n} P(k_i \Big| C_0) \, \cdot P(C_0) - \prod_{i=1}^{n} P(k_i | C_1) \, \cdot P(C_1)$$

if $\mathrm{f(x=0) - f(x=1)}$> 1, f (x) will be classified to 1.

if $\mathrm{f(x=0) - f(x=1)}$< 1, f (x) will be classified to 0.

because $\mathrm{k_i}$ ∈ {0, 1}, $\mathrm{k_1} = 1 - \mathrm{k_0}$. the function could be:

$$f(x) = \sum_{i=1}^{n} \log \frac{P(k_i | C_0)}{(1 - P(k_i | C_0))} + 2 \cdot \log\big(P(C_0)\big) - 1$$

Let

$$x_i = \log \frac{P(k_i | C_0)}{\big(1 - P(k_i | C_0)\big)}$$

$$x_0 = 2 \cdot \log\big(P(C_0)\big) - 1$$

It will equal the vector w_i in binary classification. so they are the same. Because it has x_i (actually it should be w_i) , the total dimension will be n + 1. Input x_i should always equals 1 or 0.

2.

For Naïve Bayes classifier, w_{NB} can be learned by calculating the values of $P(y)$ and $P(x|y)$, which can be estimated from frequency counts of training data and this is not that difficult, while for Logistic Regression classifier, it is less restrictive and w_{LR} is chosen arbitrarily and thus requires a full search over the linear space of possible models. The data requirement for learning w_{LR} is $O(n)$, while it is $O(\log n)$ for learning w_{NB}. Therefor Logistic Regression converges slower to its asymptotic accuracy than Naïve Bayes, and learning w_{NB} is much easier than learning w_{LR}.

Question 3:

1.

The loss function of Logistic Regression , the probability P(y=0|x)

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

The loss function of Logistic Regression , the probability P(y=1|x)

$$P(y = 0 \mid x) = 1 - \frac{1}{1 + e^{-w^T x}}$$

the loss function is

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

this can drives(prove process):

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

$$= \sum_{i=1}^{n} y_i * \ln\left(\frac{1}{1+e^{z_i}}\right) + (1 - y_i) * \ln\left(1 - \frac{1}{1+e^{z_i}}\right)$$

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

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

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

because of minimized the loss function equals maximized log-likelihood (loss function=-log likelihood)

$\sum_{i=1}^{n} -y_i * w^T x_i + \ln (1 + \exp (w^T x_i))$ is the loss function for logistic regression that we want to prove.

2.

The loss function of Logistic Regression:

$$\mathrm{loss}(f(w^T x), y) = -\log\left(f(w^T x)\right) \qquad \text{if } y = 1$$

$$\mathrm{loss}(f(w^T x), y) = -\log\left(1 - f(w^T x)\right) \quad \text{if } y = 0$$

the loss function of the total dataset is:

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

$$f(w^T X_i)) = \sum_{i=1}^{n} y_i * \left(\log\frac{1-f(w^T x)}{f(w^T x)}\right) - \log(1 - f(w^T x))$$