COMP9318 Assignment
Fengting YANG
Z5089358

Q1 (1).

Location	Time	Item	SUM
Sydney	2005	PS2	1400
Sydney	2006	PS2	1500
Sydney	2006	Wii	500
Melbourne	2005	XBox 360	1700
Sydney	2005	ALL	1400
Sydney	2006	ALL	2000
Melbourne	2005	ALL	1700
Sydney	ALL	PS2	2900
Sydney	ALL	Wii	500
Melbourne	ALL	XBox 360	1700
ALL	2005	PS2	1400
ALL	2006	PS2	1500
ALL	2006	Wii	500
ALL	2005	XBox 360	1700
Sydney	ALL	ALL	3400
Melbourne	ALL	ALL	1700
ALL	2005	ALL	3100
ALL	2006	ALL	2000
ALL	ALL	PS2	2900
ALL	ALL	Wii	500
ALL	ALL	XBox 360	1700
ALL	ALL	ALL	5100

(2). SELECT location, time, item, sum(quantity)

FROM SALES

GROUP BY location, time, item

UNION

SELECT location, time, "ALL", sum(quantity)

```
FROM SALES
GROUP BY LOCATION, TIME
UNION
SELECT location, "ALL", item, sum(quantity)
FROM SALES
GROUP BY LOCATION, item
UNION
SELECT "ALL", TIME, item, sum(quantity)
FROM SALES
GROUP BY TIME, item
UNION
SELECT location, "ALL", "ALL", sum(quantity)
FROM SALES
GROUP BY LOCATION
UNION
SELECT "ALL", TIME, "ALL", sum(quantity)
FROM SALES
GROUP BY TIME
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
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

(4). The map function is 12*location+4*time+item

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

So the MOLAP cube should be

index	value
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
12	3400
24	1700
4	3100
8	2000
1	2900
3	500
2	1700
0	5100

Q2 (1)

In Naïve Bayes classifier, we have

$$\mathrm{P(y=1|x) > P(y=0|x)} \quad \mathrm{s.t.} \quad \mathrm{y=1}$$

Use conditional probability formula, it becomes

$$\mathrm{P(x|y=1)P(y=1) > P(x|y=0)P(y=0)}$$

Expand the above formula in d-dimention

$$\mathrm{P(y=1)}\prod_{i=0}^{d} P(x_i=1|y=1)^{x_i} P(x_i=0|y=1)^{1-x_i}$$
$$> \mathrm{P(y=0)}\prod_{i=0}^{d} P(x_i=1|y=0)^{x_i} P(x_i=0|y=0)^{1-x_i}$$

The r.h.s. of above formula is positive, so

$$\frac{\mathrm{P(y=1)}\prod_{i=0}^{d} P(x_i=1|y=1)^{x_i} P(x_i=0|y=1)^{1-x_i}}{\mathrm{P(y=0)}\prod_{i=0}^{d} P(x_i=1|y=0)^{x_i} P(x_i=0|y=0)^{1-x_i}} > 1$$

$$\frac{\mathrm{P(y=1)}}{\mathrm{P(y=0)}}\prod_{i=0}^{d}\frac{P(x_i=0|y=1)}{P(x_i=0|y=0)}\cdot\prod_{i=0}^{d}\left(\frac{P(x_i=1|y=1)P(x_i=0|y=0)}{P(x_i=1|y=0)P(x_i=0|y=1)}\right)^{x_i} > 1$$

Use log at both side

$$\log\frac{\mathrm{P(y=1)}}{\mathrm{P(y=0)}} + \sum_{i=0}^{d}\log\frac{P(x_i=0|y=1)}{P(x_i=0|y=0)}$$
$$+\sum_{i=0}^{d} x_i \log\frac{P(x_i=1|y=1)P(x_i=0|y=0)}{P(x_i=1|y=0)P(x_i=0|y=1)} > 0$$

Set

$$\mathrm{b} = \log\frac{\mathrm{P(y=1)}}{\mathrm{P(y=0)}} + \sum_{i=0}^{d}\log\frac{P(x_i=0|y=1)}{P(x_i=0|y=0)}$$

and

$$\mathrm{w} = \log\frac{P(x_i=1|y=1)P(x_i=0|y=0)}{P(x_i=1|y=0)P(x_i=0|y=1)}$$

s.t.

$$b + \sum_{i=0}^{d} w_i \cdot x_i > 0$$

In this condition, the predict result should be class 1 and the classifier is a linear classifier. We set the $x^i = [1\ x^i]$. It is a d+1-dimension linear classifier.

(2) For LR,

$$w_j \leftarrow w_j + \alpha(y^{(i)} - P(x^{(i)}))x^{(i)}{}_j$$

For NB, we can learn by calculating the conditional probability which can be calculated from frequency. The data requirement for LR is O(n), but for NB, it come to O(log n). In consequence, NB has a higher converge speed and it is easier than LR.

Q3(1)

For sigmoid function:

$$y = \frac{1}{1+e^{-z}} \; where \; z = \boldsymbol{w}^T\boldsymbol{x}$$

Use ln at both side:

$$\ln\frac{y}{1-y} = \boldsymbol{w}^T\boldsymbol{x}$$

Obviously, if y is the probability of positive samples(P(y=1|x)), 1-y will be the probability of negative samples(P(y=0|x)).

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

Since y+(1-y)=1,

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

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

We can estimate w by maximum likelihood method. For a dataset

$$\{(x_i, y_i)\}_{i=1}^n$$

The log-likelihood function should be maximized

$$\ell(w) = \sum_{i=1}^{n} \ln P(y_i|x_i; \mathrm{w})$$

Define:

$$\hat{x} = (x; 1)$$

$$P_1(\hat{x}; w) = P(y=1|\hat{x}; \mathrm{w})$$

$$P_0(\hat{x}; w) = P(y=0|\hat{x}'\mathrm{w}) = 1 - P_1(\hat{x}; w)$$

The likelihood part in log-likelihood function can be transformed as:

$$P(y_i|x_i; \mathrm{w}) = y_i P_1(\hat{x}; w) + (1-y_i) P_0(\hat{x}; w)$$

s.t.

$$\ell(w) = \sum_{i=1}^{n} y_i \ln\frac{e^{\boldsymbol{w}^T\boldsymbol{x}}}{1+e^{\boldsymbol{w}^T\boldsymbol{x}}} + (1-y_i)\ln\frac{1}{1+e^{\boldsymbol{w}^T\boldsymbol{x}}}$$

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

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

Thus, to maximize l(w) is to minimize the -l(w).

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

which is the loss function.

(2)

Define

$$P_0 = P(y = 1|x) = f(\boldsymbol{w}^T \boldsymbol{x})$$

$$P_1 = P(y = 0|x) = (1 - f(\boldsymbol{w}^T \boldsymbol{x}))$$

s.t.

$$\mathrm{L} = \prod_{i=1}^{n} P_0^{y_i} (1 - P_0)^{1-y_i}$$

Use ln at both side.

$$\ell(w) = \sum_{i=1}^{n} y_i \ln P_i + (1 - y_i) \ln(1 - P_i)$$

$$= \sum_{i=1}^{n} y_i \ln\, f(\boldsymbol{w}^T \boldsymbol{x}) + (1 - y_i) \ln(1 - f(\boldsymbol{w}^T \boldsymbol{x}))$$

Therefore, the loss function should be -l(w), e.g.

$$\sum_{i=1}^{n} -y_i \ln\, f(\boldsymbol{w}^T \boldsymbol{x}) - (1 - y_i) \ln(1 - f(\boldsymbol{w}^T \boldsymbol{x}))$$