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

Use conditional probability formula, it becomes

Expand the above formula in d-dimention

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

Use log at both side

Set

and

s.t.

In this condition, the predict result should be class 1 and the classifier is a linear classifier. We set the . It is a d+1-dimension linear classifier.

(2) For LR,

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:

Use ln at both side:

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)).

Since y+(1-y)=1,

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

The log-likelihood function should be maximized

Define:

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

s.t.

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

which is the loss function.

(2)

Define

s.t.

Use ln at both side.

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