Cong Cong

Z3414050

Assignment1

COMP9318

Q1.

1):

|  |  |  |  |
| --- | --- | --- | --- |
| Location | Time | Item | Quantity |
| 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 ALL

SELECT Location, Time, ALL, SUM(Quantity)

FROM Sales

GROUP BY Location, Time

UNION ALL

SELECT Location, ALL, Item, SUM(Quantity)

FROM Sales

GROUP BY Location, Item

UNION ALL

SELECT ALL, Time, Item, SUM(Quantity)

FROM Sales

GROUP BY Time, Item

UNION ALL

SELECT Location, ALL, ALL, SUM(Quantity)

FROM Sales

GROUP BY Location

UNION ALL

SELECT ALL, Time, ALL, SUM(Quantity)

FROM Sales

GROUP BY Time

UNION ALL

SELECT ALL, ALL, Item, SUM(Quantity)

FROM Sales

GROUP BY Item

UNION ALL

SELECT ALL, ALL, ALL, SUM(Quantity)

FROM Sales

3):

|  |  |  |  |
| --- | --- | --- | --- |
| Location | Time | Item | Quantity |
| Sydney | ALL | ALL | 3400 |
| 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):

In order to find an injective mapping function:

I tried but it doesn’t provide an one to one mapping, then I take and this gives me an one to one mapping.

|  |  |  |  |
| --- | --- | --- | --- |
| Location | Time | Item | offset |
| 1 | 1 | 1 | 17 |
| 1 | 2 | 1 | 21 |
| 1 | 2 | 3 | 23 |
| 2 | 1 | 2 | 30 |
| 1 | 1 | 0 | 16 |
| 1 | 2 | 0 | 20 |
| 2 | 1 | 0 | 28 |
| 1 | 0 | 1 | 13 |
| 1 | 0 | 3 | 15 |
| 2 | 0 | 2 | 26 |
| 0 | 1 | 1 | 5 |
| 0 | 2 | 1 | 9 |
| 0 | 2 | 3 | 11 |
| 0 | 1 | 2 | 6 |
| 1 | 0 | 0 | 12 |
| 2 | 0 | 0 | 24 |
| 0 | 1 | 0 | 4 |
| 0 | 2 | 0 | 8 |
| 0 | 0 | 1 | 1 |
| 0 | 0 | 3 | 3 |
| 0 | 0 | 2 | 2 |
| 0 | 0 | 0 | 0 |

And the sparse multi-dimensional array is (sorted by offset):

|  |  |
| --- | --- |
| 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 |

Q2

1):

I used , so if > 1, I can classify input features as class, otherwise, the input features can be classified as

class.

Next:

**And based on Bayesian Theorem:**

**The above equation equals to:**

Now, we can use **Bernoulli Naïve Bayes** to further simplify :

And the above equation can be simplified to:

I assume:

And equation is:

Now, we can substitute equation [2] back to equation [1], and we obtain:

If I assume:

Finally, the equation can be simplified to:

This is obviously a linear classifier in d+1 dimension space, with the vector

2):

From part 1), I know that

Every element in can be obtained directly

However, if I use Logistic Regression:

Here, I need to find which maximize the likelihood:

And Log-likelihood is:

So, if we take the derivative of :

As, is a function of w, our aim is to try to make our estimation as close to the observed data as possible.

To achieve that goal, there are several ways to do, one of them is to use **Gradient Ascent,** but obviously this method is more complicated than calculating directly.

Q3:

1):

is the percentages sample in the mixture, and is the percentage of Object in the sample , now after the measurements, we are given the percentage of Object in the whole mixture and noted as .

The likelihood function can be written as:

Take the log of the above equation:

And the above equation is the log likelihood function.

2):

To make simplification easier, it is better to use ln instead of using log, now the values of are given and the values of are provided in the table, thus the only unknown variable in the above equation is , but I know and I can use to represent . First substituting the known values into the equation:

To find the MLE, I first take the derivative of the above equation and set it equals to zero, what I get is:

Solve the above equation, I got and

And the expected percentage of each component is: