**COMP9318 (18S1) ASSIGNMENT 1**

**Z5040672**

**Yuyao Guo**

**Question 1**

(1)

|  |  |  |  |
| --- | --- | --- | --- |
| Location | Time | Item | SUM(Quantity) |
| Sydney | 2005 | PS2 | 1400 |
| Sydney | 2005 | ALL | 1400 |
| Sydney | 2006 | PS2 | 1500 |
| Sydney | 2006 | Wii | 500 |
| Sydney | 2006 | ALL | 2000 |
| Sydney | ALL | PS2 | 2900 |
| Sydney | ALL | Wii | 500 |
| Sydney | ALL | ALL | 3400 |
| Melbourne | 2005 | Xbox 360 | 1700 |
| Melbourne | 2005 | ALL | 1700 |
| Melbourne | ALL | Xbox 360 | 1700 |
| Melbourne | ALL | ALL | 1700 |
| ALL | 2005 | PS2 | 1400 |
| ALL | 2005 | Xbox 360 | 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 | Xbox 360 | 1700 |
| ALL | ALL | ALL | 5100 |

(2)

图片包含 文字

已生成高可信度的说明

图片包含 墙壁

已生成极高可信度的说明

(3)

|  |  |  |  |
| --- | --- | --- | --- |
| Location | Time | Item | SUM(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 |

(4)  **F(Location, Time, Item) = 2\*Location + 5\*Time + 11\*Item**

|  |  |
| --- | --- |
| Offset | SUM(Quantity) |
| 0 | 5100 |
| 2 | 3400 |
| 4 | 1700 |
| 5 | 3100 |
| 7 | 1400 |
| 9 | 1700 |
| 10 | 2000 |
| 11 | 2900 |
| 12 | 2000 |
| 13 | 2900 |
| 16 | 1400 |
| 18 | 1400 |
| 21 | 1500 |
| 22 | 1700 |
| 23 | 1500 |
| 26 | 1700 |
| 27 | 1700 |
| 31 | 1700 |
| 33 | 500 |
| 35 | 500 |
| 43 | 500 |
| 45 | 500 |

**Question 2**

**(1) Prove that if the feature vectors are d-dimension, then a Naïve Bayes classifier is a linear classifier in a d + 1-dimension space.**

When **x** is classified to label **y** = 1 by Naïve Bayes classifier, it means:

Because **x** is a binary vector, each dimension of **x** takes only 0 or 1, let us use following presentation:

We can know that:

According to (a):

Then:

Because are constants, so the form of (b) is like to , which is the form of linear.

Therefore, if the feature vectors are d-dimension, then a Naïve Bayes classifier is a linear classifier in a d+1-dimension space.

**(2) Briefly explain why learning is much easier than learning .**

To learn the parameter , we need to maximize the Log-likehood of Logistic Regression for all samples.

The Log-likehood is *,* and to maximize it, we need to take its partial derivatives for each , then [solve](javascript:;) [a](javascript:;) [system](javascript:;) [of](javascript:;) [equations](javascript:;) or use methods like [gradient](javascript:;) [descent](javascript:;).

To learn the parameter , according to (1), we can know that and for each , .

Therefore, to learn is much easier than learning .

**Question 3**

According to the question, is Sigmoid function, therefore:

Then, we can know:

The like-hood is:

The Loss Function of Logistic Regression is:

(2)

According to the question,

Then, we can know:

The like-hood is:

The Loss Function of Logistic Regression is: