COMP9318 (17S1) ASSIGNMENT 1

Yu Feng z5094935

May 2017

1 Q1. (40 marks)

1.1

	Location	Time	Item	Quantity
0	Melbourne	2005	XBox 360	1700
1	Melbourne	2005	ALL	1700
2	Melbourne	ALL	XBox 360	1700
3	Melbourne	ALL	ALL	1700
4	Sydney	2005	PS2	1400
5	Sydney	2005	ALL	1400
6	Sydney	2006	PS2	1500
7	Sydney	2006	Wii	500
8	Sydney	2006	ALL	2000
9	Sydney	ALL	PS2	2900
10	Sydney	ALL	Wii	500
11	Sydney	ALL	ALL	3400
12	ALL	2005	PS2	1400
13	ALL	2005	XBox 360	1700
14	ALL	2005	ALL	3100
15	ALL	2006	PS2	1500
16	ALL	2006	Wii	500
17	ALL	2006	ALL	2000
18	ALL	ALL	PS2	2900
19	ALL	ALL	Wii	500
20	ALL	ALL	XBox 360	1700
21	ALL	ALL	ALL	5100

1.2

SELECT *
FROM [Location] CROSS JOIN [Time] CROSS JOIN [ITEM]

1.3

	Location	Time	Item	Quantity
0	Sydney	2005	PS2	1400.0
1	Sydney	2005	ALL	1400.0
2	Sydney	2006	PS2	1500.0
3	Sydney	2006	ALL	1500.0
4	Sydney	ALL	PS2	2900.0
5	Sydney	ALL	ALL	2900.0
6	ALL	2005	PS2	1400.0
7	ALL	2005	ALL	1400.0
8	ALL	2006	PS2	1500.0
9	ALL	2006	ALL	1500.0
10	ALL	ALL	PS2	2900.0
11	ALL	ALL	ALL	2900.0

1.4

$$f(Location, Time, Item) = 4^2 \cdot f(Location) + 4 \cdot f(Time) + f(Item)$$

offset	Quantity
21	1400
25	1500
27	500
38	1700

2 Q2. (30 marks)

2.1

naive bayes classifier:

$$f(x) = argmax_{x \in \{C_j\}} \prod_{i=1}^{n} P(a_i|C_j) \cdot P(C_j)$$

because it only have two class – 0,1 and each feature only have two value. so we can change the function to

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

if $f(x) > 0$, $f(x)$ will be classified to 0, otherwise it is 1. because $a_i \in \{0,1\}$, $a_1 = 1 - a_0$. the function could be:

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

let

$$x_i = \log \frac{P(a_i|C_0)}{(1 - P(a_i|C_0))},\ x_0 = 2 \cdot \log(P(C_0)) - 1$$

It will equal the vector w_i in binary classication. so they are the same. Because it has x_0 (actually it should be w_0 ..) , the total dimension will be $n+1$. Input x_i should always equals 1 or 0.

2.2

For logical regression, in order to maximize log-likelihood function, it need to take partial derivatives. And we do a lot gradient ascent to get properly w_i. It should do a lot calculation.
For naive bayes classifier, we only need to calculate each $P(x_i|C)$ once, which means we have already 'know' what should w_i be.
So naive bayes classifier is much faster than logical regression.

3 Q3. (30 marks)

3.1

Add this function after line 8.
canStop ← IsCenterChange(C,G)

Algorithm 1

Require: C set of k centers, G set of clusters

```
function IsCenterChange(C, G)
    for all g ∈ G do
        temp_i ← ComputeCenter(g)
        if c_i ≠ temp_i then
            return false
        end if
    end for
    return true
end function
```

3.2

After calculate center, the new center is the minimum point of $\sum_{i=0}^{n} Cost(g_i)$ (by definition of center point).
After find nearest center, if the distance from a point to current cluster center is larger than to another center, this point will move to the other center. So at this step, the total distance will decrease too(otherwise the point will stay in current cluster set).
Combined with these two steps, the total distance will decrease too.

3.3

Using conclusion of 3.2, we know the total cost(distance) of clustering will never increased.

1. if the old clustering is the same as the new, then the next clustering will again be the same.
2. If the new clustering is different from the old then the newer one has a lower cost.

Total possible cluster is k^N. k is the number of clusters, N is number of entries. So the loop is finite. Until all cluster get their local minimum, the loop will be end.
So it always converges to a local minimum.

References