9318 Assignment 1

Name: YU HAN

SID: Z5219071

1. (1)

	Location	**Time**	**Item**	**Quantity**
1	Sydney	2005	PS2	1400
2	Sydney	2005	ALL	1400
3	Sydney	2006	Wii	500
4	Sydney	2006	PS2	1500
5	Sydney	2006	ALL	2000
6	Sydney	ALL	PS2	2900
7	Sydney	ALL	Wii	500
8	Sydney	ALL	ALL	3400
9	Melbourne	2005	Xbox 360	1700
10	Melbourne	2005	ALL	1700
11	Melbourne	ALL	XBox360	1700
12	Melbourne	ALL	ALL	1700
13	ALL	2005	XBox360	1700
14	ALL	2005	PS2	1400
15	ALL	2005	ALL	3100
16	ALL	2006	Wii	500
17	ALL	2006	PS2	1500
18	ALL	2006	ALL	2000
19	ALL	ALL	XBox360	1700
20	ALL	ALL	Wii	500
21	ALL	ALL	PS2	2900
22	ALL	ALL	ALL	5100

1. (2) SELECT Location, Time, Item, SUM(Quantity) FROM Sales GROUP BY Location, Time, Item

   UNION

   SELECT Location, Time, null, SUM(Quantity) FROM Sales GROUP BY Location, Time UNION

   SELECT Location, null, Item, SUM(Quantity) FROM Sales GROUP BY Location, Item UNION

   SELECT null, Time, Item, SUM(Quantity) FROM Sales, Item UNION

   SELECT Location, null, null, SUM(Quantity) FROM Sales GROUP BY Location UNION

   SELECT null, Time, null, SUM(Quantity) FROM Sales GROUP BY Time UNION

   SELECT null, null, Item, SUM(Quantity) FROM Sales GROUP BY Item UNION

   SELECT null, null, null, SUM(Quantity) FROM Sales

1. (3)

	Location	Time	Item	Quantity
1	Sydney	2006	ALL	2000
2	Sydney	ALL	PS2	2900
3	ALL	2005	ALL	3100
4	ALL	2006	ALL	2000
5	ALL	ALL	PS2	2900
6	Sydney	ALL	ALL	3400
7	ALL	ALL	ALL	5100

1. (4) f(Location, Time, Item) = f(Location) + 17*f(Time) + 23*f(Item)

	Index	Quality
1	41	1400
2	18	1400
3	104	500
4	58	1500
5	35	2000
6	24	2900
7	70	500
8	1	3400
9	65	1700
10	19	1700
11	48	1700
12	2	1700
13	63	1700
14	40	1400
15	17	3100
16	103	500
17	57	1500
18	34	2000
19	46	1700
20	69	500
21	23	2900
22	0	5100

2. (1) If the Naive Bayes classifier is a linear classifier, then $\frac{P(x|y=1)P(y=1)}{P(x|y=0)P(y=0)} \geqslant 1$

When $P(y=1|x) \geqslant P(y=0|x)$

Based on Naive definition, $P(x|y) = \prod_{j=0}^{d} P(x_j|y)$

$\therefore \frac{P(x|y)P(y=1)}{P(x|y=0)P(y=0)} = \prod_{j=0}^{d} \frac{P(x_j|y=1)}{P(x_j|y=0)} * \frac{P(y=1)}{P(y=0)}$

Let $P = P(y=1), M_j = P(x_j = 1|y=1), N_j = P(x_j = 1|y=0)$

$\therefore P(x_j|y=1) = M_j^{x_j} * (1-M_j)^{1-x_j}$

Because x is a binary factor, $\therefore P(x_j|y=0) = N_j^{x_j} * (1-N_j)^{1-x_j}$

$\because y=1, \frac{P}{1-P} * \prod_{j=0}^{d} \frac{M_j^{x_j}(1-M_j^{1-x_j})}{N_j^{x_j}(1-N_j^{1-x_j})} \geq 1$

$\therefore log(\frac{P}{1-P} \prod_{j=0}^{d} \frac{1-M_j}{1-N_j}) + \sum_{j=0}^{d} x_j log(\frac{M_j}{N_j} * \frac{1-N_j}{1-M_j}) \geq 0$

$\because log(\frac{P}{1-P} * \prod_{j=0}^{d} \frac{1-M_j}{1-N_j})$ has no relation with x_j

$\therefore$ it's a constant C

Let $W_j = log(\frac{M_j}{N_j} * \frac{1-N_j}{1-M_j}) \therefore \sum_{j=0}^{d} (x_j W_j) plus; C \geq 0$

$\therefore$ The Naïve Bayes classifier is a linear classifier in this situation

$W_j = log(\frac{P(x_j=1|y=1)}{P(x_j|y=0)} * \frac{1-P(x_j=1|y=0)}{1-P(x_j=1|y=1)})$

2.(2)

The learning W_{nb} is much easier than learning W_{lr} because W_{nb}s are learned independently of each other which means each of them only be calculated once, however, the W_{lr}s learned jointly, gradient ascent will be taken a lot of times.

3.(1) According from the question, we can get:

$$P_{i,j} = P(O_j|S_i); P(S_i) = q_i; P(O_j) = u_j$$

$$L(w) = MAX \prod_{j=1}^{m} P_{i,j} = \prod_{j=1}^{m} q_i^{u_j} * (1 - q_i^{1-u_j})$$

The likelihood function is:

$$\therefore l(w) = \sum_{j=1}^{m} ((u_j * log(q_i) + (1 - u_j) * log(1 - q_i)))$$

$$= \sum_{j=1}^{m} (u_j * log(\frac{q_i}{1 - q_i}) + log(1 - q_i))$$

3.(2)

$$l(w)' : u_j * \frac{1}{q_i} - \frac{1}{1 - q_i} + u_j * \frac{1}{1 - q_i} = 0$$

The MLE of q1 and q2 is when the derivative of l(w) equals 0. Therefore, MLE of q1 = 0.5, q2 = 0.5.