COMP9318 (18S1) Assignment 1

Bihe Zhu

z5149413

Q1.

(1)

Location	Time	Item	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)

```
SELECT * FROM Sales
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	Quantity
Sydney	2006	ALL	2000
Sydney	ALL	PS2	2900
Sydney	ALL	ALL	3400
ALL	2006	ALL	2000
ALL	2005	ALL	3100
ALL	ALL	PS2	2900
ALL	ALL	ALL	5100

(4)

Step 1: mappings

1 if x = `Sydney'; 2 if x = `Melbourne'; 0 if x = ALL:

1 if x = 2005; 2 if x = 2006; 0 if x = ALL:

1 if x = `PS2'; 2 if x = `XBox 360'; 3 if x = `Wii'; 0 if x = ALL:

Location	Time	Item	Quantity
1	1	1	1400
1	1	0	1400
1	2	1	1500
1	2	3	500
1	2	0	2000
1	0	1	2900
1	0	3	500
1	0	0	3400
2	1	2	1700
2	1	0	1700
2	0	2	1700
2	0	0	1700
0	1	1	1400
0	1	2	1700
0	1	0	3100
0	2	1	1500
0	2	3	500
0	2	0	2000
0	0	1	2900
0	0	3	500
0	0	2	1700
0	0	0	5100

Step 2: An injective map from cell to offset

f(location, time, item) = 16*location + 4*time + item

Location	Time	Item	Quantity	Offset
1	1	1	1400	21
1	1	0	1400	20
1	2	1	1500	25
1	2	3	500	27
1	2	0	2000	24
1	0	1	2900	17
1	0	3	500	19
1	0	0	3400	16
2	1	2	1700	38
2	1	0	1700	36
2	0	2	1700	34
2	0	0	1700	32
0	1	1	1400	5
0	1	2	1700	6
0	1	0	3100	4
0	2	1	1500	9
0	2	3	500	11
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

Step 3:

ArrayIndex	Value
0	5100
1	2900
2	1700
3	500
4	3100
5	1400
6	1700
8	2000
9	1500
11	500
16	3400
17	2900
19	500
20	1400
21	1400
24	2000
25	1500
27	500
32	1700
34	1700
36	1700
38	1700

Q2.

(1)

Suppose if feature vector = $\vec{x}$, then y=0. So, $P(y=0|\vec{x}) >= P(y=1|\vec{x})$

According to Bayesian Theorem, P(h|x)P(x)=P(x|h)P(h).

Then we get, $\frac{P(\vec{x}|y=0)P(y=0)}{P(\vec{x})} >= \frac{P(\vec{x}|y=1)P(y=1)}{P(\vec{x})}$, which can be simplified to

$P(\vec{x}|y=0)P(y=0) >= P(\vec{x}|y=1)P(y=1)$.

According to Naive Bayes Classifier, $P(\vec{x}|C_i) = \prod_{k=1}^{n} P(x_k|C_i)$.

We get

$$\frac{P(y=0)}{P(y=1)} * \prod_{k=1}^{d} \frac{P(x_k|y=0)}{P(x_k|y=1)} >= 1.$$

Because the x is a binary vector, so x is either 0 or 1.

Assume that

$$p(x_i=0|y=1)=p1, p(x_i=1|y=1)=1-p1$$
$$p(x_i=0|y=0)=p2, p(x_i=1|y=0)=1-p2'$$

So,

$$p(x_k|y=1)=p1^{(1-x_k)}(1-p1)^{x_k},$$
$$p(x_k|y=0)=p2^{(1-x_k)}(1-p2)^{x_k}.$$

So,

$$\frac{P(y=0)}{P(y=1)} * \prod_{k=1}^{d} \frac{P(x_k|y=0)}{P(x_k|y=1)} >= 1$$
$$\Rightarrow \frac{P(y=0)}{P(y=1)} * \prod_{k=1}^{d} \frac{p2^{(1-x_k)}(1-p2)^{x_k}}{p1^{(1-x_k)}(1-p1)^{x_k}} >= 1,$$

Using log to simplify it,

$$\log\left(\frac{P(y=0)}{P(y=1)} * \prod_{k=1}^{d} \frac{p2^{(1-x_k)}(1-p2)^{x_k}}{p1^{(1-x_k)}(1-p1)^{x_k}}\right) >= \log(1)$$

$$\Rightarrow \log\left(\frac{P(y=0)}{P(y=1)}\right) + \log\left(\prod_{k=1}^{d} \frac{p2^{(1-x_k)}(1-p2)^{x_k}}{p1^{(1-x_k)}(1-p1)^{x_k}}\right) >= 0$$

$$\Rightarrow \log\left(\frac{P(y=0)}{P(y=1)}\right) + \sum_{k=1}^{d}((1-x_k)\log p2 + x_k \log(1-p2) - (1-x_k)\log p1 - x_k \log(1-p1)) >= 0$$

$$\Rightarrow \log\left(\frac{P(y=0)}{P(y=1)}\right) + \sum_{k=1}^{d}(x_k(-\log p2 + \log(1-p2) + \log p1 - \log(1-p1) + \log p2 - \log p1) >= 0$$

$$\Rightarrow \log\left(\frac{P(y=0)}{P(y=1)}\right) + \sum_{k=1}^{d}(x_k \log\frac{p1(1-p2)}{p2(1-p1)} + \log\frac{p2}{p1}) >= 0$$

$$\Rightarrow \log\left(\frac{P(y=0)}{P(y=1)}\right) + \sum_{k=1}^{d}\log\frac{p2}{p1} + \sum_{k=1}^{d}(x_k \log\frac{p1(1-p2)}{p2(1-p1)}) >= 0$$

$$\Rightarrow \log\left(\frac{P(y=0)}{P(y=1)} * \prod_{k=1}^{d}\frac{p2}{p1}\right) + \sum_{k=1}^{d}(x_k \log\frac{p1(1-p2)}{p2(1-p1)}) >= 0$$

so, we could suppose that $\log\left(\frac{P(y=0)}{P(y=1)} * \prod_{k=1}^{d}\frac{p2}{p1}\right)$ is b, $\log\frac{p1(1-p2)}{p2(1-p1)}$ is w_k, then

we get $b + x_k w_k >= 0 \Rightarrow b + \vec{x}\vec{w} >= 0$,so Naive Bayes Classifier is a linear classifier.

$\vec{w}$ is $(w_1, w_2, ..., w_d)$, where w_i is $\log\frac{p(x_i=0 \mid y=1) * p(x_i=1 \mid y=0)}{p(x_i=0 \mid y=0) * p(x_i=1 \mid y=1)}$.

(2)

For logistic Regression, it only allocates the probabilities to the only cases it observed.

Naive Bayes Classifier is a generative classifiers which learn a model of joint probabilities p(x, y) and use Bayes rule to calculate p(x, y) to make a prediction.

Logistic Regression is a discriminative models learn the posterior probability p(x, y) "directly".

In discriminative models, you have "less assumptions", when you have very little data that a generative model can beat a discriminative model.

Q3.

(1)

We know that,

$$p = \frac{e^{w^T x}}{1+e^{w^T x}} = \sigma(w^T x),$$

$$\log-likelihood : L(w) = \sum_{i=1}^{n} y^{(i)} \log p(x^{(i)}) + (1-y^{(i)}) \log(1-p(x^{(i)}))$$

So, we get,

$$L(w) = \sum_{i=1}^{n} y^{(i)} \log \frac{e^{w^T x_i}}{1+e^{w^T x_i}} + (1-y^{(i)}) \log(1-\frac{e^{w^T x_i}}{1+e^{w^T x_i}})$$

$$= \sum_{i=1}^{n} y^{(i)} \log \frac{e^{w^T x_i}}{1+e^{w^T x_i}} + (1-y^{(i)}) \log \frac{1}{1+e^{w^T x_i}}$$

$$= \sum_{i=1}^{n} y^{(i)} (\log e^{w^T x_i} - \log(1+e^{w^T x_i})) + (1-y^{(i)})(-\log(1+e^{w^T x_i}))$$

$$= \sum_{i=1}^{n} y^{(i)} \log e^{w^T x_i} - y^{(i)} \log(1+e^{w^T x_i}) - (1-y^{(i)}) \log(1+e^{w^T x_i})$$

$$= \sum_{i=1}^{n} y^{(i)} \log e^{w^T x_i} - \log(1+e^{w^T x_i})$$

$$= \sum_{i=1}^{n} y^{(i)} w^T x_i - \log(1+e^{w^T x_i})$$

Since it’ s the log-likelihood which need to be maximized, so with the negative sign, it will need to be minimized which is actually a loss function.

So we get,

$$l(w) = -(\sum_{i=1}^{n} y^{(i)} w^T x_i - \log(1+e^{w^T x_i}))$$

$$= \sum_{i=1}^{n} (-y^{(i)} w^T x_i + \ln(1+\exp(w^T x_i)))$$

(2)

$$l(w) = -\sum_{i=1}^{n} (y^{(i)} \log f(w^T x_i) + (1-y^{(i)}) \log(1-f(w^T x_i)))$$