107.05.09

電子商務技術 期中考

*** 請按照題號順序作答，不會的題目也請寫上題號
所有的計算過程都應寫出，否則扣分

1. 右下圖是 weather 測試集使用 ID3 後產生的結果，試據此回答以下問題：

Relation: weather.symbolic

No.	outlook Nominal	temperature Nominal	humidity Nominal	windy Nominal	play Nominal
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	cool	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
14	rainy	mild	high	TRUE	no

```
outlook = sunny
|  humidity = high: no
|  humidity = normal: yes
outlook = overcast: yes
outlook = rainy
|  windy = TRUE: no
|  windy = FALSE: yes
```

(1.1) 為何根節點（root）為 outlook？(15%)
(1.2) 如果使用 J48 是否根節點仍為 outlook？(10%)

2. 某訓練出來的預測模型於 4000 筆資料的測試後得到 75%的正確率，試問它在 90%信心度條件下的真實正確率為何？(10%)

Table 5.1 Confidence Limits for the Normal Distribution

Pr[X ≥ z]	z
0.1%	3.09
0.5%	2.58
1%	2.33
5%	1.65
10%	1.28
20%	0.84
40%	0.25

3. 假設 support=30%, confidence=90%, 使用 Apriori Algorithm 回答以下問題：

Relation: weather.symbolic

No.	outlook Nominal	temperature Nominal	humidity Nominal	windy Nominal	play Nominal
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	cool	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
14	rainy	mild	high	TRUE	no

(3.1) 所有 Frequent Itemsets。(10%)
(3.2) 3 條關聯規則。(10%)

4. 使用 single-linkage agglomerative clustering 將以下 9 點分為三群: (15%)
$p_1(1,0)$, $p_2(1,1)$, $p_3(2,2)$, $p_4(4,0)$, $p_5(5,2)$, $p_6(6,3)$, $p_7(4,5)$, $p_8(4,4)$, $p_9(3,4)$

5. 下圖為資料集與利用 NaiveBayes 訓練 watch-promo 後的預測結果，請依序回答以下問題：

No.	Income-Range Nominal	Watch-Promo Nominal	Credit-Card-Ins Nominal	Sex Nominal	Age Numeric
1	40-50000	No	No	Male	45.0
2	30-40000	Yes	No	Female	40.0
3	40-50000	No	No	Male	42.0
4	30-40000	Yes	Yes	Male	43.0
5	50-60000	No	No	Female	38.0
6	20-30000	No	No	Female	55.0
7	30-40000	No	Yes	Male	35.0
8	20-30000	Yes	No	Male	27.0
9	30-40000	No	No	Male	43.0
10	30-40000	Yes	No	Female	41.0
11	40-50000	Yes	No	Female	43.0
12	20-30000	Yes	No	Male	29.0
13	50-60000	Yes	No	Female	39.0
14	40-50000	Yes	No	Male	55.0
15	20-30000	No	Yes	Female	19.0

```
=== Predictions on training set ===

inst#,     actual, predicted, error, probability distribution
     1        2:No      1:Yes      +  *0.546  0.454
     2       1:Yes      1:Yes         *0.692  0.308
     3        2:No      1:Yes      +  *0.565  0.435
     4       1:Yes       2:No      +   0.464 *0.536
     5        2:No      1:Yes      +  *0.627  0.373
     6        2:No       2:No          0.471 *0.529
     7        2:No       2:No          0.47  *0.53
     8       1:Yes       2:No      +   0.492 *0.508
     9        2:No      1:Yes      +  *0.634  0.366
    10       1:Yes      1:Yes         *0.684  0.316
    11       1:Yes      1:Yes         *0.619  0.381
    12       1:Yes      1:Yes         *0.53   0.47
    13       1:Yes      1:Yes         *0.627  0.373
    14       1:Yes       2:No      +   0.416 *0.584
    15        2:No       2:No          0.28  *0.72
```

(5.1) Confusion matrix? (5%)

(5.2) TP rate of class No? (5%)

(5.3) F-measure of class Yes? (5%)

(5.4) Kappa statistic? (5%)

(5.5) 試繪出 Lift chart of class No。(10%)