

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

1. 下圖為資料集 contact-lenses，請回答以下兩個問題：

Relation: contact-lenses					
No.	age Nominal	spectacle-prescrip Nominal	astigmatism Nominal	tear-prod-rate Nominal	contact-lenses Nominal
1	young	myope	no	reduced	none
2	young	myope	no	normal	soft
3	young	myope	yes	reduced	none
4	young	myope	yes	normal	hard
5	young	hypermetrope	no	reduced	none
6	young	hypermetrope	no	normal	soft
7	young	hypermetrope	yes	reduced	none
8	young	hypermetrope	yes	normal	hard
9	pre-presbyopic	myope	no	reduced	none
10	pre-presbyopic	myope	no	normal	soft
11	pre-presbyopic	myope	yes	reduced	none
12	pre-presbyopic	myope	yes	normal	hard
13	pre-presbyopic	hypermetrope	no	reduced	none
14	pre-presbyopic	hypermetrope	no	normal	soft
15	pre-presbyopic	hypermetrope	yes	reduced	none
16	pre-presbyopic	hypermetrope	yes	normal	none
17	presbyopic	myope	no	reduced	none
18	presbyopic	myope	no	normal	none
19	presbyopic	myope	yes	reduced	none
20	presbyopic	myope	yes	normal	hard
21	presbyopic	hypermetrope	no	reduced	none
22	presbyopic	hypermetrope	no	normal	soft
23	presbyopic	hypermetrope	yes	reduced	none
24	presbyopic	hypermetrope	yes	normal	none

(1.1) 試找出使用 ID3 建立的決策樹的根節點，並詳列計算過程。(15%)

(1.2) 如果改用 J48，根節點是否維持不變，為什麼？(5%)

2. 使用 Apriori algorithm，假設 support=35% (5 instances)，confidence=70%，請回答以下問題：

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

(2.1) 試找出所有 Frequent itemsets。(10%)

(2.2) 列出所有可能的關聯規則 (10%)

3. 使用 single-linkage agglomerative clustering 將以下 9 點分為三群: (15%)

$p_1(1,0), p_2(1,1), p_3(2,2), p_4(4,1), p_5(5,2), p_6(6,4), p_7(4,5), p_8(4,4), p_9(3,4)$

4. 試依據底下的 Confusion matrix 回答問題：

		Predicted Class		
		a	b	c
Actual Class	a	88	10	2
	b	14	40	6
	c	18	10	12

(4.1) TP rate of class a? (5%)

(4.2) F-measure of class b? (5%)

(4.3) Kappa statistic? (5%)

5. 試說明 Model tree 與 Regression tree 之間的差異。(5%)

6. 何謂 SVM 分類？它的優勢在哪裡？(10%)

7. 某資料集共 150 instances，標示為兩類：yes 50 instances, no 100 instances。下表為經過 DM 處理後，依照 yes 類預測後的機率排序，試回答以下問題：

Rank	Predicted	Actual Class
1	0.95	yes
2	0.93	yes
3	0.93	no
4	0.88	yes
5	0.86	yes
6	0.85	yes
7	0.82	yes
8	0.80	yes
9	0.80	no
10	0.79	yes
11	0.77	no
12	0.76	yes
13	0.73	yes
14	0.65	no
15	0.63	yes
16	0.58	no
17	0.56	yes
18	0.49	no
19	0.48	yes
...

(7.1) 取樣 10%時的 Lift 值為何? (5%)

(7.2) 寫出取樣 10%的 Confusion matrix (5%)

(7.3) 寫出 ROC curve 上，左起前三個點的座標 (5%)