

經濟部所屬事業機構 113 年新進職員甄試試題

類別：統計資訊

節次：第三節

科目：1. 資料庫及資料探勘 2. 程式設計

注意事項	<p>1. 本試題共 2 頁(A4 紙 1 張)。</p> <p>2. 可使用本甄試簡章規定之電子計算器。</p> <p>3. 本試題分 6 大題，每題配分於題目後標明，共 100 分。須用黑色或藍色原子筆或鋼筆在答案卷指定範圍內作答，不提供額外之答案卷，作答時須詳列解答過程，於本試題或其他紙張作答者不予計分。</p> <p>4. 本試題採雙面印刷，請注意正、背面試題。</p> <p>5. 考試結束前離場者，試題須隨答案卷繳回，俟本節考試結束後，始得至原試場或適當處所索取。</p> <p>6. 考試時間：120 分鐘。</p>
------	--

一、某一醫院資料庫系統由下列 4 個資料表組成，請以 SQL 語法回答下列問題：
(4 題，每題 4 分，共 16 分)

Doctor (醫師資料表)

D#	D_NAME	D_ADDR
1	Smith	Gosforth
2	Jones	Jesmond
3	Roberts	Whitley Bay

Patient (病人資料表)

P#	P_NAME	P_ADDR
1	Giles	Gosforth
2	Peters	Benton
3	Jackson	Cowgate

Visits (看診資料表)

P#	D#	VDATE	VTIME	DIAGNOSIS
1	1	1-5-88	09:30	flu
1	2	2-5-88	17:30	bronchitis
1	3	3-5-88	18:00	pneumonia
2	3	1-5-88	09:30	flu
3	1	3-5-88	18:00	broken leg

Prescriptions (處方箋資料表)

P#	VDATE	VTIME	PRES#	MEDICINE
1	1-5-88	09:30	1	aspirin
1	1-5-88	09:30	2	linctus
1	2-5-88	17:30	1	penicillin
3	3-5-88	18:00	1	aspirin

- (一) 查詢在 1988 年 5 月 3 日(3-5-88)拜訪編號 1 號醫師(D#)的病人編號(P#)。
- (二) 查詢在 1988 年 5 月 3 日(3-5-88)拜訪 Smith 醫師的病人編號(P#)。
- (三) 查詢在 1988 年 5 月 3 日(3-5-88)拜訪 Smith 醫師的病人姓名(P_NAME)與住址(P_ADDR)。
- (四) 查詢病人 Giles 的處方箋(MEDICINE)。

二、使用關聯分析來搜尋關聯規則時，須設定最小支持度(support)和最低信賴度(confidence)，請分別說明設定以該 2 項為搜尋門檻值(threshold)用意為何？(8 分)

三、當一個資料集中有兩個類別的資料，且其中一個類別的資料筆數遠大於另一個類別的資料筆數時，則稱其為不平衡(imbalanced)資料集，在這類資料集中，一般較受到關注的是少數類別的資料。請回答下列問題：(2 題，每題 13 分，共 26 分)

(一)請論述分類方法決策樹(Decision Tree)、支持向量機(Support Vector Machine)和基於規則演算法(Rule-based Algorithm)，何者較適合用來做不平衡資料的分類？

(二)請論述正確率(accuracy)、F 測度(F measure)和召回率(recall)，何者較適合用來評估分類方法在不平衡資料上的效能？

四、請依下列條件畫出樹狀圖：(2 題，每題 6 分，共 12 分)

(一)依據前序(prefix)表示法 JBHCDIGAEF 及中序(infix)表示法 CHBIDJEAGF，畫出唯一的二元樹。

(二)在 1 個空的 AVL 樹，依序插入 53、68、72、5、47、14、36、21，畫出完成後的 AVL 樹。

五、請實作下列函式以完成設計 1 個插入排序法(Insertion Sort)，據以依參數值決定排序方式採遞增或遞減。(18 分)

```
bool isInverse(int x, int y, bool isAsc); //判斷傳入的 x、y 是否反序
```

```
void InsertionSort(int *arr, int len, bool isAsc); //插入排序
```

(註：參數 arr 為傳入的整數陣列；參數 len 為整數陣列的長度；參數 isAsc 為是否遞增，函式 InsertionSort 應呼叫函式 isInverse。)

六、請使用程式語言設計 1 個低時間複雜度的函式，傳入 a、b 皆為不大於 2^{30} 的正整數後，可判斷出 $|a - b|$ 的值是否只有 3 個因數，並註明所使用的程式語言。(20 分)

(註：該函式只允許呼叫平方根函式。)