

經濟部所屬事業機構 107 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

1. 本試題共 2 頁(A4 紙 1 張)。
2. 可使用本甄試簡章規定之電子計算器。
3. 本試題分 6 大題，每題配分於題目後標明，共 100 分。須用藍、黑色鋼筆或原子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。
4. 本試題採雙面印刷，請注意正、背面試題。
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。
6. 考試時間：120 分鐘。

一、針對各類資料探勘工具的使用與計算，請回答下列問題：（每小題 5 分，共 15 分）

(一) 假設使用 feed-forward back-propagation neural network 建構了一個分類器，後來發現此分類器有 overfitting 的現象發生。如果想透過修改此類神經網路的拓譜(topology)來避免分類器的 overfitting 現象，請問可以如何做？

(二) 在相關規則的計算中，假設 A、B、C、D 與 E 代表各 item，如果在所有 2-itemset 中，只有 {A,B}、{B,C}、{A,D}、{A,C}、{C,D} 是頻繁項目集(frequent itemset)，請問 3-itemset 有哪些可能是 frequent itemset？

(三) 針對一交易資料集，假設項目 A、B、C 的 support 值是 A=50%、B=80%、C=60%，2-itemset 的 support 值是 {A,B}=35%、{B,C}=32%、{A,C}=10%，如果設定 minimum support=25%、minimum confidence=50%，請問此資料集有哪些相關規則？

二、附表為某公司之關聯式人事資料庫，其中部門表的「經理」欄位是外來鍵(foreign key)，參考員工表的「編號」欄位，員工表的「部門」欄位是外來鍵，參考部門表的「編號」，請回答下列問題：（每小題 5 分，共 15 分）

部門表

| 編號 | 名稱 | 經理 | 預算 | 總薪資 |
|-----|----|-----|------|------|
| D01 | 會計 | 002 | 20K | 200K |
| D02 | 資訊 | 018 | 70K | 900K |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| D05 | 客服 | 200 | 300K | 600K |

員工表

| 編號 | 姓名 | 部門 | 薪資 |
|-----|----|-----|-----|
| 001 | 張山 | D02 | 30K |
| 002 | 李嗣 | D02 | 80K |
| ⋮ | ⋮ | ⋮ | ⋮ |
| 200 | 王武 | D03 | 50K |

(一) 員工李嗣離職後，會刪除員工李嗣的資料，除了「限制刪除」，還有哪些作法可維護參考完整性(Referential Integrity)？

(二) 假設員工表資料量大，針對查詢 `select 姓名,薪資 from 員工表 where 部門=' D02'`，請以 SQL 語法建立應有的索引(Index)名為 Idx1，以加速此查詢。

(三) 請以 SQL 語法列出各部門員工人數大於 200 位的部門編號與員工人數。

三、以下是某公司需要記錄並儲存的資料，請根據 A~G 的描述設計關聯式資料庫結構 (relational database schema)，並標出主鍵(primary key)和外來鍵(foreign key)，以及每一外來鍵所參考的主鍵。(20分)

- A. 員工資料需記錄員工編號、姓名、性別和地址，其中員工編號是主鍵。
- B. 部門資料需記錄部門編號和部門名稱，其中部門編號是主鍵。
- C. 客戶資料需記錄客戶編號、姓名、地址和電話，其中客戶編號是主鍵。
- D. 商品資料需記錄商品編號和商品名稱，其中商品編號是主鍵。
- E. 需記錄員工在哪一個部門工作以及工作的月薪，一個員工只能在一個部門工作，一個部門會有多個員工。
- F. 需記錄哪一個部門賣哪些商品，一個部門可賣多個商品，不同部門不可賣相同商品。
- G. 需記錄哪一個客戶購買了哪些商品以及購買的日期，一個客戶可購買多個商品，不同客戶也可購買相同的商品。

四、以 PHP、JSP、ASP.NET 等程式語言開發網站時，常利用 Session、Cookie、POST Method、GET Method 儲存網頁資訊或傳遞參數，請說明其運作方式與特性(安全性、速度...等)。(16分)

五、請以 Perl、Javascript 等程式語言或虛擬碼(Pseudo code)檢查使用者輸入之字串是否符合身分證字號格式，並分別以下列 2 種處理方式撰寫。(每小題 5 分，共 10 分)

(一)多條件式判斷(IF conditions)

(二)正規表示法(Regular Expression)

六、何謂環狀佇列(Circular Queue) (4分)？請使用 C、Java 等程式語言或虛擬碼(Pseudo Code)以陣列(Array)方式實作環狀佇列，其中須包含新增與刪除佇列元素之方法 (20分)。