

經濟部所屬事業機構 110 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。
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、BIRCH 是一個 hierarchical clustering 方法，可以處理大量資料，以及避免雜訊(noisy)資料的問題，請簡答以下題目：(4 題，每題 5 分，共 20 分)

(一)此方法適用何種資料型態？

(二)此方法利用一種 tree 資料結構，請說明 internal 節點(除 pointer 之外)儲存何種資料？

(三)此資料結構透過何種機制，達成可處理大量資料，不受記憶體大小限制？

(四)BIRCH 建構完成 tree 後，可再採用其他分群組方法，假設採用 k-means clustering，以 elbow curve method 決定群組個數 k，請說明 elbow curve method 之做法。

二、資料探勘常用歐幾里得距離來定義相似度(或不相似度)，資料前處理時，常需對每一欄位分別做正規化(或稱標準化)。(共 2 題，共 10 分)

(一)請說明正規化之目的。(1 分)

(二)請說明常用之 3 種正規化方法。(每個答案 3 分，共 9 分)

三、請針對下表的資料進行關聯規則分析，資料為購買商品的交易紀錄，設定最小支持度 $\text{min_sup} = 30\%$ 與最小信心度 $\text{min_conf} = 80\%$ 為強關聯規則的最低門檻。請寫出以下商品的所有強關聯規則(僅列出「若 item1 與 item2，則 item3」之強關聯規則即可，如：「 $AB \rightarrow C$ 」)。(每個答案 5 分，共 20 分)

TID	ITEMS_BOUGHT
T1	A, B, E
T2	B, C, D, E
T3	B, F
T4	A, D, E
T5	A, B, D, E
T6	C, D, G

四、下列問題，請用遞迴(Recursive)的方式來撰寫：（共 2 題，共 15 分）

(一)實驗室培養某細胞，此細胞區分為成年與幼年細胞。每個成年細胞於 1 個月後會生出 2 個幼年細胞，幼年細胞於 1 個月後會變為成年細胞。請用遞迴方式撰寫一函式 CellNumber，輸入為 N 個月後，其傳回為 N 個月後細胞的總數。（8 分）

(二)請用遞迴方式撰寫一函式 GCD，輸入為 2 個正整數，其傳回為此 2 個正整數之最大公因數。（7 分）

五、某銀行有多台取號機供民眾取號，每台取號機提供多項服務的取號，同一服務會依據各取號機之取號的先後時間，依序給予連續的號碼，且最小的號碼為 1。請使用下列所有的函式來撰寫一取號函式 GetNumber，輸入為服務代碼與鎖定號碼變數失敗的上限次數，其傳回為此服務的下一個號碼，但當鎖定號碼變數失敗超過其上限次數時，則傳回 0。（15 分）

int lock(int serviceid) /* 鎖定指定服務代碼的號碼變數。輸入為服務代碼，其傳回為是否鎖定成功，當鎖定成功時，則傳回 1，當鎖定失敗時，則傳回 0。 */

unlock(int serviceid) /* 解鎖指定服務代碼的號碼變數。輸入為服務代碼。 */

int get(int serviceid) /* 取得指定服務代碼的號碼變數之值。輸入為服務代碼。 */

set(int serviceid, int value) /* 設定指定服務代碼的號碼變數之值。輸入為服務代碼與此服務代碼的號碼變數欲設定之值。 */

sleep() /* 執行暫停 0.1 秒 */

六、某電梯管理系統控制多台電梯，當某樓層之等待區按下按鈕時，會將其樓層號碼傳送到系統之共用 Queue 中，系統會定時自共用 Queue 中讀取一筆目標樓層號碼，並尋找離此目標樓層最近的電梯來提供服務。請使用下列所有的函式，來撰寫一選取電梯函式 SelElev，輸入為電梯的總數，其傳回為離目標樓層最近的電梯號碼，但當未選擇到電梯時，則傳回 0。另外，當電梯總數為 N 時，電梯號碼為 1 到 N；樓層號碼為 1 代表 1 樓、2 代表 2 樓、0 代表地下 1 樓、-1 代表地下 2 樓(依此類推)。（20 分）

int getqueue() /* 自共用 Queue 中取得下一筆目標樓層號碼。其傳回為下一筆目標樓層號碼，當無資料時，則傳回 999。 */

int getcurrfloor(int elevid) /* 取得指定電梯目前所在的樓層。輸入為電梯號碼，其傳回為此電梯目前所在的樓層，但當此電梯不提供服務時，則傳回 999。 */