

經濟部所屬事業機構 112 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 3 頁(A3 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、卷積神經網路(convolution neural network)可由二維矩陣中取出一維特徵向量，以利後續建立單層或深度神經網路。有一維度為 5×5 之二維矩陣如【圖 1】所示，有一維度為 3×3 之卷積核如【圖 2】所示，請以【圖 2】之卷積核對【圖 1】之二維矩陣進行步長為 1 的卷積操作(convolution operation)，所得之二維矩陣再以維度為 2×2 之最大池化操作(max-pooling operation)同樣以步長為 1 進行處理。請分別列出經卷積操作及最大池化操作之二維矩陣。(20 分)

35	21	71	12	34
22	14	81	6	15
41	24	11	61	17
65	64	23	45	48
73	15	88	51	27

【圖 1】

0	1	0
-1	0	-1
0	1	0

【圖 2】

- 二、某資料庫之交易資料如【表 1】所示，其中每 1 列為 1 筆交易，TID 欄位為交易序號，items 欄位為 1 筆交易所包括之項目，請以 Apriori 演算法(設定最小支持度為 4)求出所有頻繁項目集(frequent itemsets)並詳列其演算過程。(15 分)

【表 1】

TID	items
T100	I2, I3, I4, I5, I6
T200	I1, I2, I3, I4
T300	I1, I2, I5
T400	I1, I2, I4, I6
T500	I1, I2, I3, I4
T600	I1, I4, I6
T700	I1, I4

三、某一維資料集有 6 筆資料：(10,25,45,55,80,90)，若以 k-means 演算法，分為 A、B、C 群集，群集中心起始值分別為(0,40,100)，如【表 2】所示。請列出前 3 次迭代(iteration)結果(計算至整數位，以下四捨五入)。(15 分)

【表 2】

Iter #	群集中心位置			各筆資料之群集所屬					
	A	B	C	10	25	45	55	80	90
0	0	40	100	A	B	B	B	C	C
1									
2									
3									

四、下列程式經審閱後有嚴重的安全性威脅，可讓使用者執行不被允許的指令，請回答下列問題：(3 題，共 15 分)

```
#include <stdio.h>
#include <string.h>

int main() {
    char username[16];
    char password[16];

    printf("Enter your username: ");
    gets(username);

    printf("Enter your password: ");
    gets(password);

    if (strcmp(username, "testuser") == 0 && strcmp(password, "testpwd ") == 0) {
        printf("Access granted!\n");
    } else {
        printf("Access denied!\n");
    }

    return 0;
}
```

(一)請說明該程式具有何種安全性威脅？(3 分)

(二)請說明此安全性威脅如何造成危害及其運作原理？(6 分)

(三)為改善上述程式，請提出修訂作法。(6 分)

五、在計算機科學中，「stack」是1種常見的資料結構，請回答下列問題：（3題，共15分）

(一)請說明何謂「stack」？（3分）

(二)請以2個實際應用案例具體說明如何使用「stack」？（6分）

(三)請基於「stack」概念就以下問題設計1個程式（若程式非基於「stack」概念進行實作，將不予計分）：

將4個浮點數(20.22、12.21、70.25和58.29)插入1個stack中，然後逐一從該stack中移除它們，且必須能檢測該stack是否已為空。（6分）

六、隨著資訊技術的進步，資料的重要性與日俱增，如何善用資料強化競爭力，已是企業刻不容緩的經營議題。現在請您扮演資料科學家的角色，就以下需求設計1個程式，將檔案data.txt讀取數據後，執行資料清理，最後將整理好的資料寫入資料庫中：（20分）

A. 檔案data.txt中包含未清理的數據資料，每1行代表1條記錄。每條記錄包含多個字段，字段之間使用逗號進行分隔。每條記錄的字段數可能不同，但都包含以下字段：

- 使用者號碼 (User ID)
- 姓名 (Name)
- 年齡 (Age)
- 地址 (Address)
- 電子郵件地址 (Email)
- 手機號碼 (Phone)

B. 資料清理包括以下項目：

- 刪除任何缺失字段的記錄
- 清除姓名和地址字段中任何不必要的空格
- 將所有電子郵件地址轉換為小寫
- 驗證使用者號碼，確保它是1個8位數字
- 驗證年齡字段，確保它是1個合法的正整數
- 驗證手機號碼，確保它是1個10位數且為09開頭

C. 最後，使用任何您常用的資料庫系統（例如 SQLite、MySQL、PostgreSQL 等），將整理好的資料寫入資料庫，並於首次寫入時先建立1個名為user_data的資料表。