

經濟部所屬事業機構 109 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	<ol style="list-style-type: none">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、某公司欲建立【考績】的預測模型，假設【表 1】是 10 名員工的資料，請以此為例，依序回答下列問題。（共 2 題，共 25 分）

【表 1】

工號	年齡	性別	部門	考績
021	22	女	行政	乙
022	24	女	行政	乙
023	30	男	銷售	乙
024	33	男	行政	甲
025	35	女	行政	乙
026	36	女	銷售	乙
027	38	男	研發	甲
028	48	女	研發	乙
029	50	男	研發	乙
030	52	男	研發	甲

(一)採用 k-均值分群法(k-means clustering)將【年齡】分成青年與壯年二族群，以標籤取代數值，離散化(discretization)或稱二分法(dichotomization)的結果為何？請簡述計算過程，而且要明確指出哪些員工被歸類為青年，哪些屬於壯年。（10 分）

(二)延續上題，進行【年齡】離散化後，若用單純貝氏分類器(Naïve Bayes)來建立【考績】的預測模型，新進員工【工號 = 031, 年齡 = 50, 性別 = 男, 部門 = 研發】的預測考績為何？請簡述計算過程，並明確指出預測結果的考績為何。（15 分）

二、某公司的關聯式資料庫中包含下列 3 個表格，請用 SQL 語言列出下列查詢。(3 題，每題 5 分，共 15 分)

供應商資料表 Suppliers (SupplierID, SupplierName, ContactName, Address, City, PostalCode, Country, Phone)，主鍵 SupplierID 為供應商編號, SupplierName 為供應商名稱, ContactName 為聯絡人, Address 為地址, City 為城市, PostalCode 為郵遞區號, Country 為國家, Phone 為電話

產品資料表 Products (ProductID, ProductName, SupplierID, CategoryID, QuantityPerUnit, UnitPrice, UnitsInStock, UnitsOnOrder, ReorderLevel, Discontinued)，主鍵 ProductID 為產品編號, ProductName 為產品名稱, SupplierID 為供應商編號, CategoryID 為分類編號, QuantityPerUnit 為每單位數量, UnitPrice 為單價, UnitsInStock 為在庫量, UnitsOnOrder 為訂購量, ReorderLevel 為再訂購水準, Discontinued 為是否停產

分類資料表 Categories (CategoryID, CategoryName, Description)，主鍵 CategoryID 為分類編號, CategoryName 為分類名稱, Description 為分類的描述

(一)找出位在相同城市的供應商編號，供應商名稱，以及城市。

(二)找出最大在庫量的產品編號，產品名稱，分類名稱，供應商編號，供應商名稱，在庫量，聯絡人。

(三)找出已經停產(Discontinued的值為1)的產品，但是其訂購總量不等於0的供應商編號，供應商名稱，以及其所停產的產品的品項數。

三、若資料庫中有一客戶資料表(Table) CUSTOMER 如下。今利用此資料表訓練一倒傳遞類神經網路(Back Propagation Neural Network)，使其可依客戶購買行為判別客戶性別，則應如何設計此類神經網路？請回答下列問題：(共 3 題，共 10 分)

(一)應於輸入層設計幾個神經元？為什麼？(3 分)

(二)應於輸出層設計幾個神經元？為什麼？(3 分)

(三)如果資料庫中有 10 萬筆此種客戶資料，應如何運用這些資料做訓練及測試？(4 分)

CUSTOMER

ID	Freq	UnitPrice	Amount	Discount	Payment	Sex
----	------	-----------	--------	----------	---------	-----

ID：客戶編號

Freq：購物頻率(平均每月購買次數)

UnitPrice：單價(購買商品平均單價)

Amount：每筆購物總價(平均單次購物總價)

Discount：購物折扣率(所購商品平均折扣率)

Payment：付款方式(主要付款方式：現金、信用卡、行動支付...)

Sex：性別

四、給定物件導向程式碼如下：（共 2 題，共 10 分）

<pre>class A { public int a0; private int a1; public void m1() {a1 = 1;} public void m1(int x) {a1 = x;} public void m2() {a1 = 3;} }</pre>	<pre>class B extends A { private int b1; private int b2; public void m2() {b1 = 5; b2 = 2;} public void m3() {a0=m1(b1);} }</pre>
---	---

(一)請以 class A 說明以下概念：（6 分）

(1)封裝(encapsulation) (3 分)

(2)繼承(inheritance) (3 分)

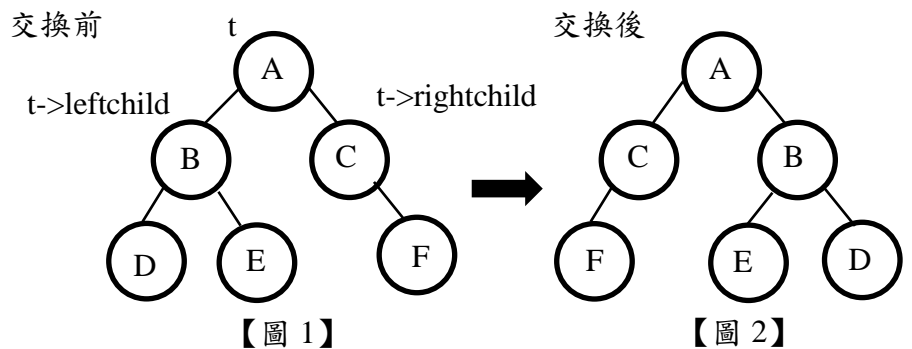
(二)請以 class A、class B 說明以下概念：（4 分）

(1)覆寫(overriding) (2 分)

(2)超載(overloading) (2 分)

五、請於下方標示之【待填入程式區塊】中，以 6 行內為限填入虛擬程式碼(pseudo-code)，實現將二元樹(以鏈結串列方式儲存)中每一節點的左子樹、右子樹皆調換之功能(示意如下，交換前二元樹如【圖 1】，交換後二元樹如【圖 2】，以 t 表示父節點之指標，t->leftchild 表示 t 的左子節點指標，t->rightchild 表示 t 的右子節點指標)。（15 分）

```
Swaptree ( t pointer)
{
if ( t <> NULL) then
{
【待填入程式區塊】
}
} /* Swaptree 函式結尾
```



六、給定一陣列名稱為 NUM，包含 n 個不重複整數(n > 2)，請撰寫虛擬程式碼找出該陣列中元素兩兩乘積最大者(即 Maximum pairwise product，變數名稱為 maxprod，maxprod = maximum (NUM[i] * NUM[j], i <> j))，完成下列 2 項子題。（共 2 題，共 25 分）

(一)請說明欲撰寫之虛擬程式碼的主要程式邏輯。（10 分）

(二)請在演算法時間複雜度須為 O(n)的限制下，撰寫虛擬程式碼。(請注意，如作答內容之演算法時間複雜度經分析為 O(n²)，本子題僅給 5 分) (15 分)