

經濟部所屬事業機構 104 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、請列出在資料探勘中，以決策樹為基礎的分類方法有哪些？（10 分）

二、假設現在有一些文件需要被分類成+和-兩類，為簡化起見，我們僅考慮 3 個詞，分別是"資料庫"、"資料探勘"和"電子商務"，我們共有 2100 份文件的訓練資料集，它們對於這 3 個詞和類別的資訊如【表 1】所示，其中 T 表示有出現，F 表示無出現：

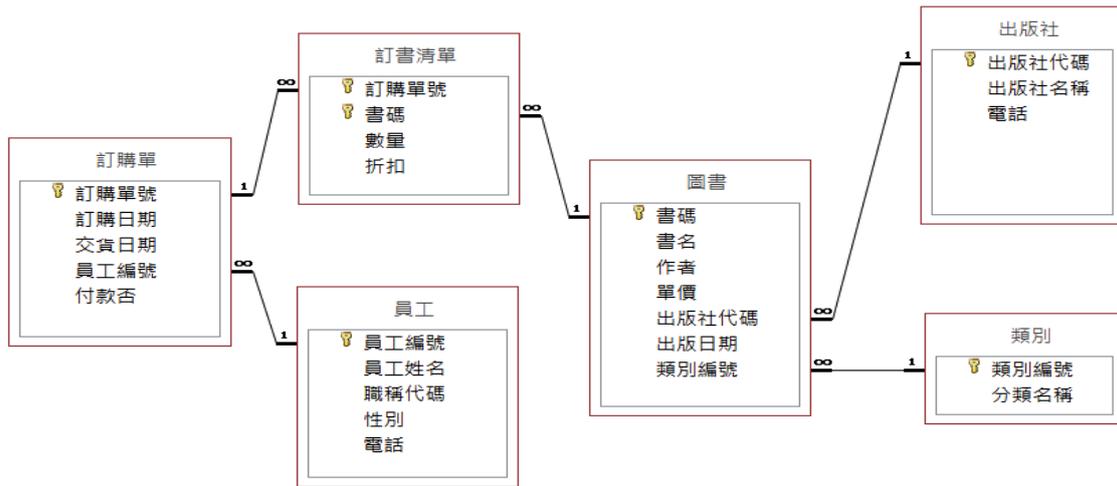
資料庫	資料探勘	電子商務	文件數量	
			+	-
T	T	T	100	100
F	T	T	10	10
T	F	T	30	0
F	F	T	490	100
T	T	F	100	200
F	T	F	500	200
T	F	F	70	50
F	F	F	100	40

【表 1】

比如，第 1 列的資料顯示在有出現"資料庫"、"資料探勘"和"電子商務"3 個詞的文件裡，有 100 份的類別標籤是+，而另 100 份的類別標籤是-。

現在有 1 份新文件，它沒有出現"資料庫"，但有出現"資料探勘"和"電子商務"，請利用簡易貝式分類器(Naïve Bayesian Classifier)來預測新文件的類別，須列出詳細計算式（15 分），以及簡易貝式分類器有何假設（5 分）？

三、請依據【圖 1】資料庫之實體關聯圖，回答下列查詢(請使用 SQL 語法)(各小題 5 分，共 20 分)



【圖 1】

- (一)請將類別為散文的書籍，單價調漲兩成。
- (二)請列出從未接過訂單的員工編號、姓名、電話。
- (三)請列出每筆訂單的訂購單號、交貨日期、小計金額。
- (四)請列出銷售量大於 1000 本書的出版社名稱與銷售量，並依據銷售量遞減排序。

四、在物件導向程式設計中：

- (一)請說明何謂抽象類別及其特性。(4 分)
- (二)請宣告一內含抽象方法之抽象類別，同時宣告一子類別繼承此抽象類別，並在子類別中實作在抽象類別內宣告之抽象方法(請註明使用何種程式語言或虛擬碼)。(6 分)

五、如欲以結構化程式設計方式利用堆疊設計一簡易撲克牌遊戲發牌系統，包含洗牌、切牌及發牌，以數字 1~52 代表撲克牌 52 張牌，堆疊陣列 R[0..51]儲存代表 52 張撲克牌之數字，R[51]為底部第 52 張牌，R[0]為頂部第 1 張牌：

- (一)洗牌部分，先用亂數方式隨機取出上述 52 個數字其中 1 個數字(代表隨機取一張牌)以堆疊方式填入陣列 R 之底部，再繼續取下 1 個數字(下一張牌，須判斷亂數取出之數是否已被取出過，若已取過須重取)以堆疊方式依序填入陣列 R，直至 52 個數字(52 張牌)皆被取出填入陣列 R 為止，試畫出其流程圖。(5 分)
- (二)已用隨機方式選擇從陣列 R 第 20 張牌(R[19])切牌，並開始取出發牌，請問取出發牌的順序為何？請以 R 陣列內含引數 R[i]方式表示。(5 分)
- (三)開始發牌後，如何判斷已發牌至堆疊陣列 R 之底部，以及如何繼續取出下一張牌至整副牌發完？試以流程圖表示。(10 分)
- (四)洗牌及切牌均須隨機從 1~52 取出一數字，今有一無參數之亂數函數 rand()可隨機傳回值 K，其中 $0 \leq K < 1$ ，請問如何利用 rand()隨機取出 1~52，試寫出其程式表示式。(5 分)

六、有一數列：30、5、37、2、61、8、55、19、46、23，欲以快速排序法進行排序，以 30 作為第一次比較之鍵值：

- (一)請將數列由小至大排列完成為止，寫出排序的步驟。(10 分)
- (二)請將每次做比較之鍵值列出。(5 分)
 - 第一次比較之鍵值：30
 - 第二次比較之鍵值：
 - 第三次比較之鍵值：
 - 依此類推。