

經濟部所屬事業機構 106 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、k-均值聚類法(k-Means Clustering)是資料探勘常使用之分群(Clustering)方法，請回答下列問題。(10 分，每小題 5 分)

(一)此方法除了須事先設定 k 值外，通常還有哪些缺點？

(二)在無法事先決定 k 值下，對數值資料進行分群，當反覆嘗試各種 k 值時，如何判斷該 k 值已經是最適當之數值？

二、某公司之關聯式人事資料庫如【附圖】所示，請回答下列問題：(15 分)

員工

員工編號	姓名	年資	年齡
22	林峯	10	45
29	王五	1	33
31	章三	8	55
32	李四	2	25
58	張飛	4	35
64	陳璇	7	37
71	周潔	5	36
74	楊忠	3	28

任職紀錄

記錄編號	員工編號	到職年月	部門編號
001	22	10/2007	101
002	22	8/2012	104
003	29	7/2016	102
004	31	10/2009	102
005	31	11/2015	103
006	32	11/2015	103
007	58	2/2013	104
008	64	9/2010	101
009	71	9/2012	102
010	74	9/2014	103

部門

部門編號	部門名稱	主管編號
101	研發處	64
102	行銷處	71
103	人事室	31
104	產品處	22

【附圖】

(一)是否可從【附圖】部門資料表中，直接刪除「人事室」這筆資料，並請說明原因。(5 分)

(二)請寫出將所有員工的年資及年齡都增加 1 的 SQL 語法。(3 分)

(三)請寫出可查詢曾任職兩個(含)以上不同部門的員工，其員工編號及姓名的 SQL 語法。(7 分)

三、【表 1】是某一簡化後的信用卡交易紀錄表，請試用 k-近鄰算法(k-Nearest Neighbors Algriothm, kNN 或稱案例推論法)回答下列問題。(25 分)

【表 1】

交易編號	退貨與否	婚姻狀況	信用等級	月薪(萬)	盜刷與否
001	否	已婚	高	10	否
002	否	單身	中	7	否
003	是	單身	高	12	否
004	否	離婚	高	9	是
005	是	單身	低	3	否
006	否	單身	中	8	是
007	是	離婚	高	23	否
008	否	已婚	中	8	否

(一)請說明何謂 kNN 方法。(5 分)

(二)請寫出【表 1】中各欄位(交易編號除外)分別屬於符號型(Nominal)、次序型(Ordinal)、亦或數值型(Numerical)欄位，以及其適合之相似度(Similarity)計算方式。(10 分)

(三)假設 k=3 且各欄位的重要性都一樣，請使用上題回答之計算方式，判斷下列新產生之刷卡紀錄為正常或被盜刷。(未列出算式不計分)(10 分)

交易編號=011、退貨與否=否、婚姻狀況=已婚、信用等級=高、月薪=10 萬

四、【表 2】為某工程的工作計畫，請以頂點工作網路(Activity On Vertex Network)的應用，回答下列問題。(未列出算式不計分)(10 分，每小題 5 分)

【表 2】

工作項目編號	工作項目名稱	工作項目所需天數	須完成之前項工作項目
A	準備計畫	10	無
B	申請執照	5	A
C	選擇建商	2	A
D	基礎工程	8	B、C
E	外部油漆工程	7	D
F	內部油漆工程	4	D
G	內部裝潢施作	3	F
H	外部廣場佈置	7	E
I	工程驗收	2	G、H

(一)請畫出本工程 AOE(Activity On Edge)網路圖。

(二)請計算本工程最快幾天後可完成工程驗收。

五、試以設計線上英文字典查詢的搜尋法為例，假設其資料分布均勻，請回答下列問題：(15 分)

(一)常見之搜尋法中，哪一種最適合應用此搜尋，且搜尋時間最有效率(3 分)?並請說明此搜尋法的時間複雜度(3 分)。

(二)有一數列：15、1、3、100、50、44、58、46、27、38，如以 50 作為欲搜尋之鍵值，請以上述回答的搜尋法，依序列出於搜尋成功前，各次與鍵值比較的值為何。(未列出算式不計分) (9分)

六、【表 3】是某籃球隊 10 位隊員參加 2 場比賽的得分統計表，在考慮平均執行時間的前提下，請以內部排序法中最好的方法，回答下列問題。(25分)

【表 3】

場次 \ 隊員	隊員 A	隊員 B	隊員 C	隊員 D	隊員 E	隊員 F	隊員 G	隊員 H	隊員 I	隊員 J
第 1 場 得分	16	2	17	0	30	5	37	10	22	5
第 2 場 得分	10	3	20	1	31	6	22	5	26	14

(一)欲將 2 場比賽各隊員的總得分由小至大排列完成，請寫出此排序法為何 (3 分)？並說明此排序的作法 (7 分)。

(二)承上題，請列出各次排序(以總得分排序)的結果為何。(15分)