

臺灣銀行 109 年新進人員甄試試題

進用職等／甄試類別【代碼】：7 職等／電子金融-數據分析人員【Q6118】

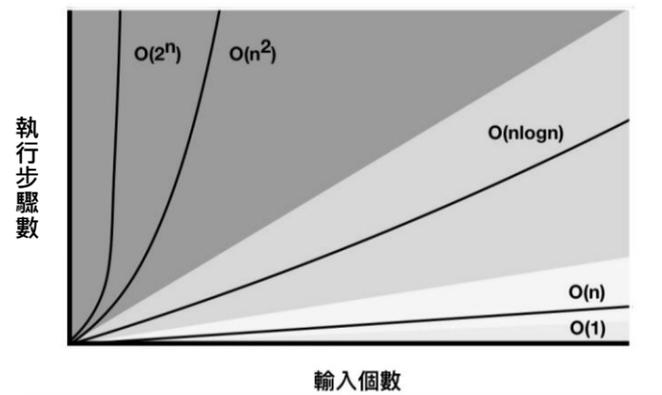
科目二：數據分析

\*入場通知書編號：\_\_\_\_\_

注意：①作答前先檢查答案卷，測驗入場通知書編號、座位標籤、應試科目是否相符，如有不同應立即請監試人員處理。使用非本人答案卷作答者，該節不予計分。  
②本試卷為一張雙面，非選擇題共 4 大題，每題 25 分，共 100 分。  
③非選擇題限以藍、黑色鋼筆或原子筆於答案卷上採橫式作答，並請依標題指示之題號於各題指定作答區內作答。  
④請勿於答案卷上書寫應考人姓名、入場通知書編號或與答案無關之任何文字或符號。  
⑤本項測驗僅得使用簡易型電子計算器（不具任何財務函數、工程函數、儲存程式、文數字編輯、內建程式、外接插卡、攝（錄）影音、資料傳輸、通訊或類似功能），且不得發出聲響。應考人如有下列情事扣該節成績 10 分，如再犯者該節不予計分。1.電子計算器發出聲響，經制止仍執意續犯者。2.將不符規定之電子計算器置於桌面或使用，經制止仍執意續犯者。  
⑥答案卷務必繳回，未繳回者該節以零分計算。

第一題：

【圖（一）】為演算法設計時所需考量的時間複雜度比較圖，請回答下列問題：



【圖（一）】

（一）請簡述何謂時間複雜度？時間複雜度對於大數據處理有何關聯性？【5 分】

（二）請計算下列 Python 程式碼的「執行步驟數」為多少次？【10 分】

```
y = [10, 9, 8]
for each_value in y:
    print(each_value)
```

（三）請詳述【圖（一）】之 Y 軸標題為何不是程式「執行時間」反而是指程式「執行步驟」？【10 分】

第二題：

在建立資料預測模型時，通常會將手頭上所持有的資料區分為兩大類，包含訓練資料集 (training set) 與測試資料集 (test set)，請回答下列問題：

（一）請簡述「訓練資料集」與「測試資料集」的各自用途為何？【5 分】

（二）Precision 與 Recall 是在機器學習訓練中用來判斷模型好壞的指標，請舉例並解釋此兩項指標的意涵。【10 分】

（三）何謂 F1-score？它與小題（二）中的 Precision、Recall 兩者關係為何？【10 分】

【請接續背面】

**第三題：**

在預測（分類）某人是否會辦理信用卡（視為變數 C）的問題中，會辦理信用卡者視為類別 1，不會辦理信用卡者視為類別 0。評估預測方法好壞會利用到混淆矩陣(confusion matrix)。其中有四種情況，如下：

	C 預測為類別 1	C 預測為類別 0
C 實際為類別 1	True Positive(TP)	False Negative(FN)
C 實際為類別 0	False Positive(FP)	True Negative(TN)

請分別回答下列問題：

- (一) 請依照問題情境，以是否辦理信用卡的方式說明 ROC(Receiver Operating Characteristic)曲線圖中，X 軸、Y 軸及圖中四個端點的意義。【10 分】
- (二) 若分類時，我們考慮兩個變數 A（有兩個類別 1 與 0）、B（有三個類別 1、2 與 3），其條件機率如下：  
 $P(A=1|C=0)=1/4$ ,  $P(A=0|C=0)=3/4$ ,  $P(A=1|C=1)=1/7$ ,  $P(A=0|C=1)=6/7$   
 $P(B=1|C=0)=2/7$ ,  $P(B=2|C=0)=1/7$ ,  $P(B=3|C=0)=4/7$   
 $P(B=1|C=1)=2/3$ ,  $P(B=2|C=1)=1/3$ ,  $P(B=3|C=1)=0$   
 $P(C=0)=0.3$ ,  $P(C=1)=0.7$   
 若某人其變數 A 類別 0、變數 B 為類別 2，請利用單純貝氏法(naïve Bases)，計算出相關機率，並分類其是否會辦理信用卡？【10 分】
- (三) 使用單純貝氏法需要哪些前提假設？若這些假設不成立，一般會用哪些解決方式？【5 分】

**第四題：**

某資料中有四個類別變數(Class, Sex, Age, Survived)，各類別及其出現的個數如下：

變數名稱							
Class		Sex		Age		Survived	
類別	個數	類別	個數	類別	個數	類別	個數
1st	325	Female	470	Adult	2092	No	1490
2nd	285	Male	1731	Child	109	Yes	711
3rd	706						
Crew	885						

今欲以關聯性分析方法分析其關聯，請回答下列各題：

- (一) 某關聯 A 為 {Sex=Female, Age=Adult} => {Survived=Yes}。請以機率的角說明明何謂關聯 A 的支持度(Support)、信賴度(Confidence)及提升度(Lift)。【15 分】
- (二) 若此四個變數的四維列聯表如下，請算出關聯 A 的支持度、信賴度及提升度。【10 分】

, , Age = Adult, Survived = No		
	Sex	
Class	Female	Male
1st	4	118
2nd	13	154
3rd	89	387
Crew	3	670

, , Age = Adult, Survived = Yes		
	Sex	
Class	Female	Male
1st	140	57
2nd	80	14
3rd	76	75
Crew	20	192

, , Age = Child, Survived = No		
	Sex	
Class	Female	Male
1st	0	0
2nd	0	0
3rd	17	35
Crew	0	0

, , Age = Child, Survived = Yes		
	Sex	
Class	Female	Male
1st	1	5
2nd	13	11
3rd	14	13
Crew	0	0