

104年公務人員特種考試警察人員、一般警察人員考試及104年
特種考試交通事業鐵路人員、退除役軍人轉任公務人員考試試題 代號：20350 全一頁

等 別：二等一般警察人員考試

類 科 別：刑事警察人員犯罪分析組

科 目：資料探勘技術（包括資料庫管理與運用、線上交易處理【OLTP】、
資料倉儲【Data Warehouse】、資料探勘【Data Mining】）

考試時間：2 小時

座號：_____

※注意：(一)禁止使用電子計算器。

(二)不必抄題，作答時請將試題題號及答案依照順序寫在試卷上，於本試題上作答者，不予計分。

一、下列關聯表 SJT 限制如下(1)每一門科目的每位學生只能被一位老師教，(2)每位老師只能教一門科目（但一門科目可由多位老師教）。請問關聯表 SJT 是否是廣義第三正規化型式（BCNF）？並說明理由。假如不是，該如何切割（decompose）關聯表 SJT？切割後有何功能相依性（functional dependency）會遺失？（20 分）

學生 (S)	科目 (J)	老師 (T)
張三	數學	陳老師
張三	物理	黃老師
李四	數學	陳老師
李四	物理	林老師

二、針對資料探勘（data mining），請描述其資料前置作業（data preprocessing）的主要步驟。（20 分）

三、下列資料庫有四筆交易，如果最小支持度（minimum support）為 60%、最小信心水準（minimum confidence）為 80%。請找出所有的頻繁項目集（frequent itemsets）和所有符合下列型態的關聯法則（association rule），同時也要列出其支持度（support）和信心水準（confidence）。（20 分）

{項目 1, 項目 2} $\Rightarrow$ 項目 3 [支持度（support），信心水準（confidence）]

TID	項目
T100	{K,A,D,B}
T200	{D,A,C,E,B}
T300	{C,A,B,E}
T400	{B,A,D}

四、用下列分類資料呈現混淆矩陣表（confusion matrix），並算出精確率（precision）、回想率（recall）及正確率（accuracy）。（20 分）

真實結果：

癌症樣本數目 200
非癌症樣本數目 800

分類結果：

預測真實癌症數目 150
預測真實非癌症數目 700

五、線上分析處理（OLAP）在高維度下會發生什麼問題？該如何解決？（20 分）