

99年公務人員特種考試警察人員考試及
99年特種考試交通事業鐵路人員考試試題

代號：20220 全一張
(正面)

等 別：二等考試
類 科：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、內政部警政署委託某一大學舉辦高階資料探勘研習班，研習過程中學員必須完成指定作業並參加模擬測驗。研習通過與否以期末考成績為唯一標準。該班學員資料如下表所列：

學號 (ID)	電話 (Phone)	語言 (Language)	完成所有作業 (Assignments)	模擬考成績 (Test)	通過期末考 (Exam)
0001	3231-5483	Java	No	31	Yes
0002	3231-5482	Java	No	20	No
0003	3231-4283	C++	Yes	35	Yes
0004	3231-4729	Python	Yes	25	Yes
0005	3231-4214	Java	Yes	39	No
0006	3231-4782	C++	No	29	No
0007	3231-4824	Java	No	19	No
0008	3231-3845	Python	Yes	32	Yes

(一)請利用上表資料，設計並描述一個合理的 K-NN classifier (K-Nearest Neighbor 分類器) 來預測學員是否會通過最終的期末考。(20分)

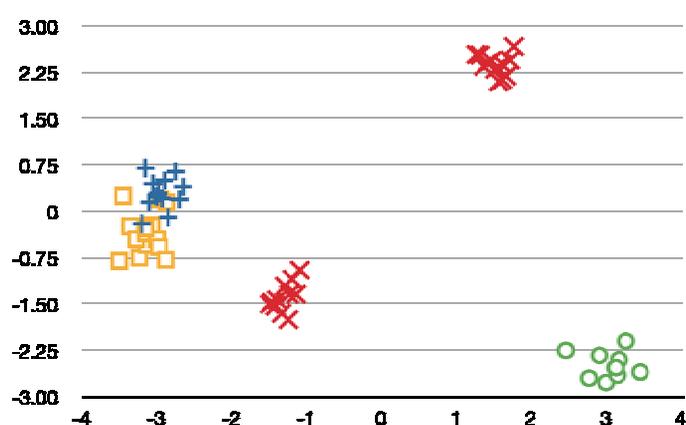
(二)若學員 0009 的研習記錄如下所示，請用你所設計的 K-NN classifier 預測該學員是否有可能通過最後的期末考。請寫出相關計算與分析。(10分)

學號 (ID)	電話 (Phone)	語言 (Language)	完成所有作業 (Assignments)	模擬考成績 (Test)	通過期末考 (Exam)
0009	3231-5423	C++	Yes	30	?

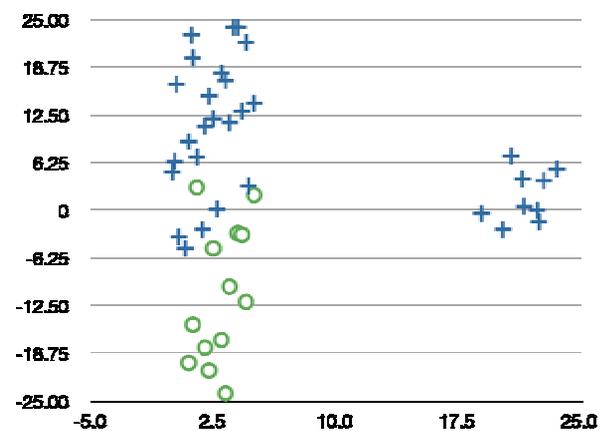
二、下圖(一)與圖(二)分別是用 K-means 分群 (clustering) 的結果。請分別分析為什麼分群結果會不甚理想，並提出改進之道 (但必須保有 K-means 之精神)。

(一) (10分)

(二) (10分)



圖(一)



圖(二)

(請接背面)

99年公務人員特種考試警察人員考試及
99年特種考試交通事業鐵路人員考試試題

代號：20220

全一張
(背面)

等 別：二等考試

類 科：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】）

三、請比較 Naïve Bayesian, Support Vector machine, K-nearest neighbor 這三種常見的資料分群方法（classification methods）。

(一)那一種方法最容易修改成可以處理動態資料流（Dynamically changing data streams）？請敘明理由以及改變方法。（15分）

(二)那一種方法最不容易修改成可以處理動態資料流？請敘明理由。（10分）

四、請說明一個兩群組的二維資料集（2-dimensional dataset）如何可以用一個線性 SVM 完全正確的分群，但是卻無法用 3-NN classifier 完全正確的分群。請舉例並以圖示輔助說明。（25分）