

巨量資料分析師-中級考試說明

時間	科目	題型	鑑定方式
90 分鐘	(學科)進階機器學習	單選題 50 題	電腦測驗
180 分鐘	(術科)巨量資料分析實作	共 4 大題組，每題 3~5 小題	術科實作



學科考試重點：

考科 1 (學科)

評鑑主題	評鑑內容	考試重點內容
1.模型參數最佳化 (30%)	1-1.參數估計最佳化與 求解方法 (15%)	● 各機器學習數學模型公式 ✓ 邏輯斯迴歸(Logistic regression)、線性迴歸(Linear regression) ✓ 決策樹：最大熵(Entropy)、資訊增益(Information Gain)、獲利率(Gain Ratio) ✓ 類神經參數估計 ✓ 樹狀模型
	1-2.模型參數調校與選擇 (15%)	● 模型績效評量 ✓ 迴歸：MSE (Mean Squared Error) ✓ 二元分類：ROC 曲線、混淆矩陣 (Confusion Matrix) ● 訓練與測試機制、重抽樣方法

		✓ 保留法、交叉驗證、拔靴法 ● FNMP
2.進階非監督式與監督式學習(40%)	2-1.進階集群與關聯型態分析 (15%)	● 報表結果解讀 ✓ K-means 及其延伸：誤差平方和 (SSE)、側影係數 (Silhouette Coefficient) ✓ 階層式集群 ✓ 關聯式分析 ● 集群結果的評估
	2-2.異常值偵測 (10%)	● 異常值偵測方法種類： ✓ 基於迴歸的方法、基於時間序列的方法、control chart、T-test、盒鬚圖
	2-3.非線性分類與迴歸概念 (15%)	● KNN、SVM、多項式迴歸、樹狀模型、神經網路
3.薈萃式學習 (30%)	3-1.拔靴集成法 (10%)	● 薈萃式學習沿革 ● 各種薈萃式學習方法細節及其異同、搭配的重抽樣方法
	3-2.效能提升法 (10%)	
	3-3.隨機森林 (10%)	

註：評鑑內容百分比依實際出題些微調整

術科考試重點及方式

1. 考試僅連內網。
2. 題目可使用 R 或 Python 進行作答。
3. 考生需熟悉常見機器學習演算法及套件操作，含：演算法：k-means, 簡單線性迴歸, 決策樹, 隨機森林分類樹與迴歸樹。
4. 需具備文字資料探勘基礎能力及常用中文斷詞套件「結巴」。
5. 考場將預載 R, Python 之基本套件。
6. 交卷方式：於考場建置交卷系統，於考場伺服器建置 FTP(Filezilla)交卷系統，考生作答後進行檔案上傳。

術科參考樣題

範例題型(一)

- 載入 data 資料夾所附的 iris.csv，產生所有量化欄位之相關係數矩陣，說明具高度正相關的兩個欄位 Petal.Length 與 Petal.Width 其相關係數值為何？請建立 Petal.Length 為反應變數(response variable)，Petal.Width 為預測變數(predictor)的簡單線性迴歸模型。當 Petal.Width 為 0.6 時，Petal.Length 的預測值為何？其殘差值又為何？

評分標準：

項次	評分標準說明	佔比	分數
1	相關係數矩陣計算	20%	
2	簡單線性迴歸模型建立	20%	
3	Petal.Width 為 0.6 時，Petal.Length 的實際值	20%	
4	Petal.Width 為 0.6 時，Petal.Length 的預測值	20%	
5	殘差值	20%	
合計		100%	

範例題型(二)

- 載入 data 資料夾所附的 Smarket.csv，它是從 2001 到 2005 年的 S&P 500 指數資料集，每筆資料為日觀測值包含當天百分比報酬(Today)、前五天百分比報酬(Lag1-Lag5)、成交量(Volumn)，以及當天報酬的正負方向(Direction)。建模的目標是以過去五天的百分比報酬，預測當天 S\&P 500 漲跌方向(Direction 是 Up 或 Down)的分類模型。預測建模的必備流程如下：請以 75%和 25%切分訓練集與測試集，運用 75%訓練樣本進行重複三次的十摺交叉驗證(10-fold cross validation repeated three times)模型參數調校訓練，找出分類正確率平均值最高(佳)的模型，最後以妳/你的最佳模型預測 25%的測試樣本，並敘述最佳模型的參數，以及測試集的正確率(小數點下四位)。

評分標準：

項次	評分標準說明	佔比	分數
1	資料切分是否 75% vs 25%	10%	
2	預測變數標準化	10%	
3	重複三次 10 摺交叉驗證找最佳模型	10%	
4	25%測試集預測結果與小數點下四位數正確率	10%	
5	就所有考生之模型績效指標分佈範圍給分(20%, 40%, 60%, 80%)各區間得點 12%	60%	
合計		100%	