

109 年度初級巨量資料分析師能力鑑定試題

科目 2：資料處理與分析概論

考試日期：109 年 11 月 28 日

第 1 頁，共 12 頁

單選題 50 題 (佔 100%)

B	<p>1. 在分析資料前，通常需要先清理資料。當數字與文字混合在一起時，但我們僅需要提取出數字時，若以逐筆資料提取十分曠日廢時，在 Python 語法中的套件 re 可以處理大部分的此類問題，例如語法： re.findall(pattern, string)，當 pattern = '\d'，可以提取出 string 中所有單一數字；pattern = '\d\d'，可以提取出 string 中所有 2 個相連數字；pattern = '\d+'，可以提取出 string 中所有任意相連個數的數字。請問當 string = '王大明手機號碼:0912334567,地址...'時，下列何者語法無法提取出 0912334567？</p> <p>(A) re.findall('\d+', string) (B) re.findall('\d\d\d\d\d\d\d\d', string) (C) re.findall('\d\d+', string) (D) re.findall('\d\d\d\d\d\d\d+', string)</p>
D	<p>2. 下列何者「不」是資料前處理該進行的程序？</p> <p>(A) 資料清理 (data cleaning) (B) 資料轉換 (data transform) (C) 屬性挑選 (feature selection) (D) 資料建模 (data modeling)</p>
D	<p>3. 關於遺缺值 (missing value) 的處理方式，下列何者較「不」恰當？</p> <p>(A) 將有遺缺值的那筆樣本刪去 (B) 使用平均值來填補遺缺值 (C) 使用 k 近鄰演算法來填補遺缺值 (D) 將有遺缺值的那個欄位刪去</p>
C	<p>4. 關於資料具有離群值 (outlier)，進行資料標準化時，下列敘述何者較為適合？</p> <p>(A) 可採用 Z-分數法 (Z-score) (B) 可採用最小最大正規化法 (min-max normalization) (C) 可採用穩健縮放法 (robust scaler) (D) 可採用最大絕對值縮放法 (maximum absolute scaler)</p>
C	<p>5. 參考附圖，Python 語言中，關於使用 numpy 套件處理遺缺值 (missing value)，下列敘述何者「不」正確？</p> <pre>import numpy as np</pre> <p>(A) np.nan 執行結果為 nan (B) np.isnan(np.nan)執行結果為 True (C) np.NaN 執行結果為 NaN (D) np.isnan(np.NaN)執行結果為 True</p>

109 年度初級巨量資料分析師能力鑑定試題

科目 2：資料處理與分析概論

考試日期：109 年 11 月 28 日

第 2 頁，共 12 頁

C	<p>6. 關於資料彙總 (data aggregation)，下列敘述何者最為正確？</p> <p>(A) 可降低資料尺度、資料偏斜性對於模型的不良影響</p> <p>(B) 是運用推論統計學，比較兩筆樣本的差異</p> <p>(C) 是以摘要的形式收集或呈現資訊的任何過程</p> <p>(D) 是統整不同連續屬性間的數值分佈</p>
A	<p>7. 參考附圖，R 語言中，關於 aggregate 資料群組計算，下列敘述何者「不」正確？</p> <pre> > head(warpbreaks) breaks wool tension 1 26 A L 2 30 A L 3 54 A L 4 25 A L 5 70 A L 6 52 A L > str(warpbreaks) 'data.frame': 54 obs. of 3 variables: \$ breaks : num 26 30 54 25 70 52 51 26 67 18 ... \$ wool : Factor w/ 2 levels "A","B": 1 1 1 1 1 1 1 1 1 1 ... \$ tension : Factor w/ 3 levels "L","M","H": 1 1 1 1 1 1 1 1 1 2 ... > > aggregate(breaks ~ ., data=warpbreaks, mean) </pre> <p>(A) aggregate 執行結果有 8 筆資料</p> <p>(B) aggregate 函數中的 breaks 表示對 breaks 欄位進行計算</p> <p>(C) 本題 aggregate 函數功能是計算各群組的平均值</p> <p>(D) aggregate 函數的「.»表示除了 breaks 以外的所有欄位為群組欄位</p>
A	<p>8. 下列何種圖形常用於視覺化年齡與收入的關係？</p> <p>(A) 散佈圖 (scatter plot)</p> <p>(B) 甘特圖 (Gantt chart)</p> <p>(C) 流程圖 (flow chart)</p> <p>(D) 樹狀圖 (tree diagram)</p>
D	<p>9. R 語言中，下列何者為專門在處理群組與摘要的函數？</p> <p>(A) spread()</p> <p>(B) sort()</p> <p>(C) gather()</p> <p>(D) aggregate()</p>
D	<p>10. R 語言中，關於資料排序，下列敘述何者「不」正確？</p> <p>(A) order()回傳排序後的觀測值編號</p> <p>(B) sort()回傳排序後的字串值或數值</p> <p>(C) rank()回傳每個元素的排名值</p> <p>(D) permute()回傳排序後的字串值或數值</p>
B	<p>11. 下列何者「不」是屬性挑選 (feature selection) 的特性？</p> <p>(A) 降低計算時間</p>

109 年度初級巨量資料分析師能力鑑定試題

科目 2：資料處理與分析概論

考試日期：109 年 11 月 28 日

第 3 頁，共 12 頁

	<p>(B) 提高資料維度</p> <p>(C) 降低模型複雜度</p> <p>(D) 增加模型穩定性</p>
D	<p>12. 關於主成分分析 (Principal Component Analysis, PCA) 與奇異值分解 (Singular Value Decomposition, SVD)，下列敘述何者正確？</p> <p>(A) PCA 較 SVD 更一般化</p> <p>(B) SVD 將資料矩陣分解出只有橫列的基底向量</p> <p>(C) PCA 分解出資料矩陣之縱行與橫列的基底向量</p> <p>(D) PCA 可以基於資料矩陣之相關係數方陣或共變異數方陣進行計算</p>
B	<p>13. 屬性轉換 (feature transform) 為資料前處理的重要步驟，下列敘述何者「不」正確？</p> <p>(A) 可調整屬性的尺度</p> <p>(B) 會破壞模型的預測能力</p> <p>(C) 可改變屬性的偏斜性</p> <p>(D) 可降低離群資料對於模型的不良影響</p>
C	<p>14. 下列何者為以均值正規化 (mean normalization) 的方式，計算數列[1, 2, 2, 4, 5]之結果？</p> <p>(A) [0.00, 0.25, 0.25, 0.75, 1.00]</p> <p>(B) [0.2, 0.4, 0.4, 0.8, 1.0]</p> <p>(C) [-0.45, -0.20, -0.20, 0.30, 0.55]</p> <p>(D) [-0.5, 0.0, 0.0, 1.0, 1.5]</p>
B	<p>15. 下列何者「並非」屬性萃取 (feature extraction) 的方法？</p> <p>(A) 主成分分析 (Principal Component Analysis, PCA)</p> <p>(B) 正規化 (normalization)</p> <p>(C) 自編碼器 (autoencoder)</p> <p>(D) 偏最小平方迴歸 (Partial Least Squares Regression, PLSR)</p>
C	<p>16. 關於巨量資料運算平台，下列敘述何者「不」正確？</p> <p>(A) 採用水平式擴充 (scale-out)，並大量分散式處理</p> <p>(B) 具有高容錯性，採平行化運算</p> <p>(C) 採用分片機制 (sharding)，將資料分片儲存於各節點，以利多次讀取與多次寫入</p> <p>(D) 資料本身可分散到各資料節點，以增加運算效能，也可透過網路傳輸將執行程式送到資料端進行運算</p>
D	<p>17. 為了有效利用多台電腦對大量的資料進行計算，並避免某一電腦壞掉而對計算結果造成影響，請問該採用下列何種方式進行？</p> <p>(A) 集中計算 (centralized computing)</p>

109 年度初級巨量資料分析師能力鑑定試題

科目 2：資料處理與分析概論

考試日期：109 年 11 月 28 日

第 4 頁，共 12 頁

	<p>(B) 多執行緒 (multithreading)</p> <p>(C) 平行運算 (parallel computing)</p> <p>(D) 分散式運算 (distributed computing)</p>
C	<p>18. 為了從網路上蒐集外部資訊，往往需要撰寫爬蟲 (web scraping) 程式取得資料。下列何種情況與分散式的爬蟲架構較無關係？</p> <p>(A) 獲取數十個新聞網站的新聞作者、內容與圖片</p> <p>(B) 取得某一電商平台的所有商品名稱、價格與賣家訊息等</p> <p>(C) 從某一交易所的當日盤後資訊頁面取得單一統計表</p> <p>(D) 從多個技術相關文章的網站中獲取文章內容</p>
A	<p>19. 關於 MapReduce 架構，下列敘述何者「不」正確？</p> <p>(A) 當節點 (nodes) 失效時，其他節點無法接管失效的任務</p> <p>(B) 可運行在不可靠的低階電腦叢集 (clusters) 上</p> <p>(C) 映射 (map) 是將任務分配到不同節點 (nodes) 進行計算</p> <p>(D) 化簡 (reduce) 是將處理完的結果重新組合</p>
B	<p>20. 附圖是藉由 MapReduce 進行詞頻統計 (word count) 工作的流程示意圖。請問圖中對應的動作組合應為下列何者？</p> <div data-bbox="379 1037 1345 1435" data-label="Diagram"> </div> <p>(A) Mapping -> Splitting -> Shuffling -> Reducing</p> <p>(B) Splitting -> Mapping -> Shuffling -> Reducing</p> <p>(C) Reducing -> Splitting -> Mapping -> Shuffling</p> <p>(D) Splitting -> Shuffling -> Mapping -> Reducing</p>
D	<p>21. 關於相關性 (correlation)，下列敘述何者為正確？</p> <p>(A) 如果一個自變數(X)與反應變數(Y)高度相關，則此自變數應該是不重要，不用加入模型之中</p> <p>(B) 皮爾遜相關係數 (Pearson correlation coefficient) 可以用於度量兩個變數 X 和 Y 之間的相關程度，其值介於 0 與 1 之間</p> <p>(C) 皮爾遜相關係數可以用來評估非線性關係</p> <p>(D) 如果二個變數之相關係數為零，此二個變數可能具有非線性關係</p>
C	<p>22. 某研究人員想檢定國中生的性別與戴眼睛是否有顯著相關性，隨機抽</p>

109 年度初級巨量資料分析師能力鑑定試題

科目 2：資料處理與分析概論

考試日期：109 年 11 月 28 日

第 5 頁，共 12 頁

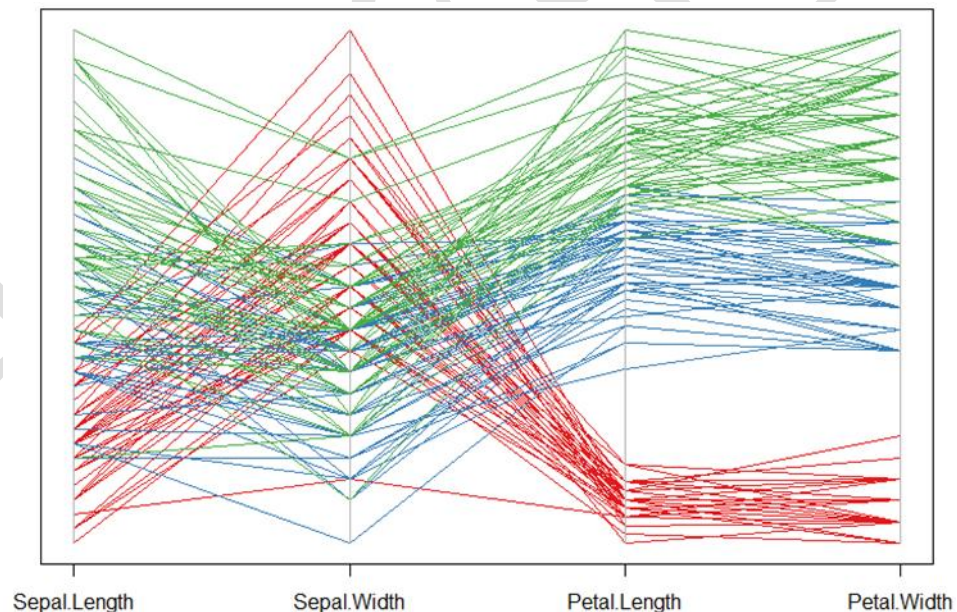
樣取得附圖資料，參考附圖 R 語言執行，下列敘述何者「不」正確？

	男	女	小計
有戴眼鏡	92	80	172
無戴鏡	10	18	28
小計	102	98	200

```
> mydf
      male female
with_glasses    92    80
without_glasses  10    28
> mydf.indep <- chisq.test(mydf)
> mydf.indep$p.value
[1] 0.004320456
```

- (A) 虛無假設為 H_0 ：戴眼鏡與性別為獨立
- (B) 對立假設為 H_1 ：戴眼鏡與性別為相關
- (C) 考慮 $\alpha=0.05$ ，則獨立性檢定之卡方分配自由度為 2
- (D) R 語言執行結果為拒絕接受 H_0 ，即戴眼鏡與性別為相關

D 23. 參考附圖 iris 資料集的視覺化結果，下列敘述何者正確？



- (A) 此視覺化結果稱為散佈圖（scatter plot）矩陣
- (B) Sepal.Length 變數最能夠將資料明顯區分成 3 個類別
- (C) 該圖使用互相垂直的軸來表示不同的維度
- (D) Petal.Length 變數較 Sepal.Width 變數更能夠將資料區分成 3 個類別

D 24. R 語言中，下列何者為可以回傳百分位數資訊的函數？

- (A) var()
- (B) sd()

109 年度初級巨量資料分析師能力鑑定試題

科目 2：資料處理與分析概論

考試日期：109 年 11 月 28 日

第 6 頁，共 12 頁

	(C) mean() (D) quantile()
B	25. 下列何者「不」代表獨立性？ (A) $P(A B) = P(A)$ (B) $P(A \cap B) = 0$ (C) $P(A \cap B) = P(A) P(B)$ (D) $P(A B) = P(A B')$ ， B' 表示 B 之補集合
A	26. 在進行假設檢定的過程中，我們會先將結果分成兩種相反的決策：虛無假設（null hypothesis, H_0 ）和對立假設（alternative hypothesis, H_1 ）。假設擲銅板 100 次，出現正面 60 次，想檢定銅板是不公正的，導致出現正面機率較大，則對立假設 H_1 應為下列何者？ (A) 出現正面的機率 $> 1/2$ (B) 出現正面的機率 $= 1/2$ (C) 出現正面的機率 $< 1/2$ (D) 出現正面的機率 $= 60/100$
D	27. 下列何者「不」屬於連續型變數資料？ (A) 年齡 (B) 銷售金額 (C) 氣溫 (D) 性別
A	28. 盒鬚圖（box plot）利用資料的統計量，可粗略地看出資料是否具對稱性及分佈程度等，請問「不」包含下列何種統計量？ (A) 標準差 (B) 最小值 (C) 中位數 (D) 第一四分位數
B	29. 參考附圖，此為鳶尾花資料集中，三種花卉（setosa, versicolor, virginica）之萼片長度（sepal length）敘述統計資訊。請問下列敘述何者「不」正確？

109 年度初級巨量資料分析師能力鑑定試題

科目 2：資料處理與分析概論

考試日期：109 年 11 月 28 日

第 7 頁，共 12 頁

iris dataset		species		
		setosa	versicolor	virginica
sepal length (cm)	count	50	50	50
	mean	5.006	5.936	6.588
	std	0.35249	0.516171	0.63588
	min	4.3	4.9	4.9
	25%	4.8	5.6	6.225
	50%	5	5.9	6.5
	75%	5.2	6.3	6.9
	max	5.8	7	7.9
	max-min	1.5	2.1	3

(A) virginica 具有較大的萼片長度平均值

(B) 從 50 個 versicolor 樣本中隨機取出一個樣本為「X」，從 50 個 virginica 樣本中隨機取出一個樣本為「Y」，「X」之萼片長度一定會小於「Y」之萼片長度

(C) 在 50 個 setosa 樣本中，有 25%的樣本之萼片長度大於 5.2cm

(D) 此資料集之三種花卉中，virginica 的萼片長度分佈較為分散

C

30. 建構機器學習模型時，常對資料進行標準化 (standardization)。附圖為 Python 語言中透過 sklearn 對資料 scores 進行標準化之程式碼。請問下列何者對其標準化後資料 normalized_scores 的敘述「不」正確？

```
from sklearn import preprocessing
scores = np.array([134, 28, 56, 180, 301, 122, 84, 92])
normalized_scores = preprocessing.scale(scores) # 標準化(Standardization)
```

(A) normalized_scores 之平均值近乎於 0

(B) normalized_scores 之標準差近乎於 1

(C) normalized_scores 中的值將會介於 0 到 1 之間

(D) normalized_scores 符合標準常態分佈 (standard normal distribution)

D

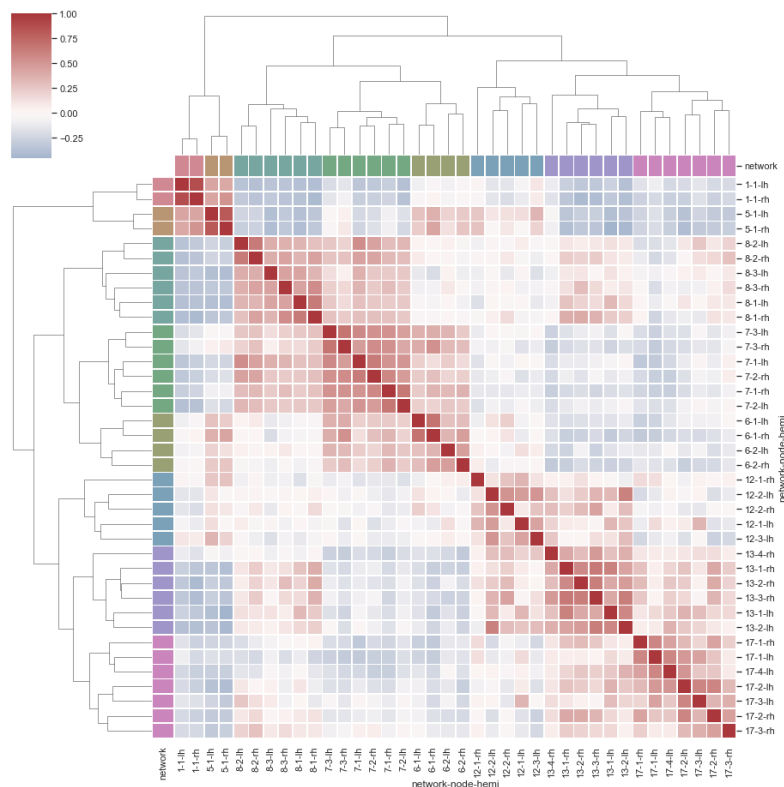
31. Python 語言中，執行 seaborn 模組 import seaborn as sns 的結果，使用下列何項函數可繪製附圖結果？

109 年度初級巨量資料分析師能力鑑定試題

科目 2：資料處理與分析概論

考試日期：109 年 11 月 28 日

第 8 頁，共 12 頁



- (A) sns.heatmap
- (B) sns.lmplot
- (C) sns.violinplot
- (D) sns.clustermap

- C 32. 關於空間密度集群算法 (Density-Based Spatial Clustering of Applications with Noise, DBSCAN)，下列敘述何者正確？
- (A) 需要預先聲明集群數量
 - (B) 無法分辨雜訊 (離群點)
 - (C) 根據資料點在特徵空間中的密度進行集群
 - (D) 儘管資料裡的分佈狀態差異很大，DBSCAN 仍能提供一個好的集群結果
- B 33. 關於常見的集群分析 (clustering analysis)，主要有階層式集群 (hierarchical clustering) 和分割式集群 (partitional clustering) 兩大類，下列敘述何者正確？
- (A) 需事先將觀察值分類，再依觀察值的屬性進行分析
 - (B) 通常以距離作為分類的依據，相對距離愈近，相似程度愈高
 - (C) 資料量夠大就不需人為定義分群數
 - (D) k 平均數集群分析法屬於階層式分群
- B 34. 關於機器學習 (machine learning)，下列敘述何者正確？
- (A) 非監督式學習 (unsupervised learning) 是具有人為標註的結果

109 年度初級巨量資料分析師能力鑑定試題

科目 2：資料處理與分析概論

考試日期：109 年 11 月 28 日

第 9 頁，共 12 頁

	(B) 階層式集群法 (hierarchical clustering) 是屬於非監督式學習 (C) 迴歸分析 (regression analysis) 是屬於非監督式學習 (D) 空間密度集群算法 (Density-Based Spatial Clustering of Applications with Noise, DBSCAN) 屬於分割式集群
D	35. 下列何者屬於監督式學習 (supervised learning) 演算法？ (A) 主成分分析 (Principal Component Analysis, PCA) (B) k 平均法 (k-means) (C) 集群分析 (clustering analysis) (D) 決策樹 (decision tree)
D	36. 關於生成式對抗網路 (Generative Adversarial Network, GAN)，下列敘述何者「不」正確？ (A) 由生成網路 (generator) 與判別網路 (discriminator) 組成 (B) 生成網路的輸出要盡量模仿訓練資料中的真實樣本 (C) 判別網路的輸入為真實樣本或生成網路的輸出，判別網路的作用是将生成網路的輸出從真實樣本中儘可能分辨出來 (D) 生成網路與判別網路相互對抗、不斷調整，最終目的是生成網路可以產生逼真的樣本、而判別網路仍可以區別出來自生成網路的輸出或是真實資料
B	37. 下列何種集群方法可以解決資料中有離群值及類別屬性的問題？ (A) k 平均數 (k-means) (B) k 代表點 (k-medoids) (C) k 近鄰 (k nearest neighbor) (D) k 奇異值分解 (k-singular value decomposition)
D	38. 關於探索式資料繪圖，下列敘述何者正確？ (A) 直方圖之 X 軸資料是間斷不連續的 (B) 長條圖不適合用於類別型資料分析 (C) 分位數圖 (QQ Plot) 無法用於常態分佈視覺化檢驗 (D) 盒鬚圖中的盒子下上 (或左右) 邊界分別為第一與第三四分位數
D	39. 關於 k 平均數 (k-means) 集群法，下列敘述何者正確？ (A) 可以處理類別型資料 (B) 不同的起始群集中心，產生的分群結果皆相同 (C) 穩定性高，對異常值或極端值不敏感 (D) 不適合非球形、數據密度變化大或有離群數據的集群問題
D	40. 關於非監督式學習 (unsupervised learning)，下列敘述何者正確？ (A) 需要事先以人力標記資料 (B) 迴歸分析 (regression analysis) 是屬於非監督式學習

109 年度初級巨量資料分析師能力鑑定試題

科目 2：資料處理與分析概論

考試日期：109 年 11 月 28 日

第 10 頁，共 12 頁

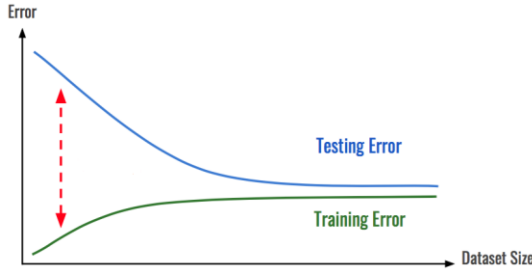
	<p>(C) 無法發現變數間或觀測值間的子群體</p> <p>(D) 集群分析為非監督式學習的方法之一</p>
C	<p>41. 參考附圖，R 語言中，執行 fpc 套件的 dbscan 函數，進行空間密度集群算法（Density-Based Spatial Clustering of Applications with Noise, DBSCAN）分析，下列敘述何者正確？</p> <pre>> library(fpc) > df <- iris[, -5] > df.dbscan <- dbscan(df, eps=0.45, MinPts=5) > df.dbscan</pre> <pre>dbscan Pts=150 MinPts=5 eps=0.45 0 1 2 border 24 4 13 seed 0 44 65 total 24 48 78</pre> <p>(A) MinPts 表示可到達區域的最小點個數，本例為 0.45</p> <p>(B) eps 表示可達區域定義鄰域的距離，本例為 5</p> <p>(C) db.dbscan 物件最右側集群編號 2 的種子點（seed）個數為 65</p> <p>(D) DBSCAN 最終集群結果分為三個緊密的群</p>
B	<p>42. 關於類別相關指標，下列敘述何者「不」正確？</p> <p>(A) 普遍率（prevalence）與偵測普遍率（detection prevalence）都必大於偵測率（detection）</p> <p>(B) F 衡量其實是在等權值的假設下，偵測率（detection）與召回率（recall）兩者的調和平均數</p> <p>(C) 特異性（specificity）是陰性事件被正確預測出來的比例</p> <p>(D) 1 減去真陽性（True Positive, TP）後即為假陰性（False Negative, FN），它是陽性事件中被錯誤預測（假陰性）的比例</p>
D	<p>43. 評估分類模型優劣時須先計算混淆矩陣（confusion matrix），關於 2x2 混淆矩陣，下列敘述何者「不」正確？</p> <p>(A) 行列交叉得到真陽性（True Positive, TP）、真陰性（True Negative, TN）、偽陽性（False Positive, FP）、偽陰性（False Negative, FN）四種情況</p> <p>(B) 橫列可以表示預測的結果，分別是預測為陰性與陽性</p> <p>(C) 預測為陽性的個數如果增加，則陰性的預測數自然會減少</p> <p>(D) 陽性與陰性的預測數量是獨立非相依的</p>
B	<p>44. 下列何者「不」是資料降維的方法？</p> <p>(A) 主成分分析（Principal Component Analysis, PCA）</p> <p>(B) 單熱編碼（one-hot encoding）</p> <p>(C) 局部線性嵌入（locally linear embedding）</p> <p>(D) 等距特徵映射（Isometric feature mapping, Isomap）</p>
C	<p>45. 參考附圖，將機器學習的訓練錯誤（training error）與測試錯誤（testing</p>

109 年度初級巨量資料分析師能力鑑定試題

科目 2：資料處理與分析概論

考試日期：109 年 11 月 28 日

第 11 頁，共 12 頁

	<p>error) 對資料集的大小作圖，紅線位置模型的表現為何？</p>  <p>(A) 模擬退火 (simulated annealing) (B) 梯度下降 (gradient descent) (C) 過度配適 (overfitting) (D) 降維 (dimension reduction)</p>				
A	<p>46. 實務上常見各類樣本分佈差距大的不平衡學習 (imbalanced learning) 情況，關於不平衡學習的處理方式，下列何者「不」正確？</p> <p>(A) 以過度抽樣 (over sampling) 或降低抽樣 (down sampling) 解決，此種方法可避免模型過度配適或遺失多數樣本中的重要訊息 (B) 運用正負樣本的懲罰權重來解決，若分析建模的算法支援樣本權重設定，此方法是簡單有效的解決途徑 (C) 以薈萃式學習 (ensemble learning) 集成模型解決，形成模型預測能力良好的森林 (D) 進行屬性挑選 (feature selection) 以解決類別不平衡問題，透過縱行的操弄來提高模型績效</p>				
D	<p>47. 關於附圖的混淆矩陣 (橫列為真實值，縱行為預測值)，請問其正確率與失誤率各是多少？</p> <table border="1" data-bbox="383 1355 566 1489"> <tr> <td>25</td><td>25</td></tr> <tr> <td>11</td><td>29</td></tr> </table> <p>(A) 40%，60% (B) 44%，56% (C) 55%，45% (D) 60%，40%</p>	25	25	11	29
25	25				
11	29				
D	<p>48. 機器學習模型中，關於模型的偏差 (bias) 與變異 (variance)，下列敘述何者正確？</p> <p>(A) 高偏差代表模型過於複雜 (B) 高變異代表模型過於簡單 (C) 模型訓練的目標為低偏差與高變異 (D) 偏差與變異之間存在抵換 (trade-off) 關係</p>				
B	<p>49. 附圖為某一垃圾郵件預測模型之混淆矩陣 (confusion matrix)。為評價</p>				

109 年度初級巨量資料分析師能力鑑定試題

科目 2：資料處理與分析概論

考試日期：109 年 11 月 28 日

第 12 頁，共 12 頁

此模型，我們除了透過預測正確率 (accuracy) 做評估外，更精準的會透過「召回率 (recall)」和「精確度 (precision)」進行評估。請問下列何者為此混淆矩陣之「召回率 (recall)」和「精確度 (precision)」？

	垃圾郵件	非垃圾郵件	總計
預測為「垃圾郵件」	30 (TP)	20 (FP)	50
預測為「非垃圾郵件」	50 (FN)	80 (TN)	130
總計	80	100	180

- (A) recall: 0.167、precision: 0.111
(B) recall: 0.375、precision: 0.6
(C) recall: 0.384、precision: 0.4
(D) recall: 0.444、precision: 0.556

- D 50. 機器學習已應用在各行各業之領域，依照應用問題與其場景的不同，必須對數據與問題有良好的理解才能進行合理的建模工作。若我們希望以監督式學習建立未來房價預測模型，下列何者較「不」適合？
- (A) 線性迴歸模型 (linear regression)
(B) 多元線性迴歸模型 (multiple regression)
(C) 神經網路模型 (neural network)
(D) k 平均數模型 (k-means)