

111 年度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：111 年 11 月 19 日

第 1 頁，共 16 頁

單選題 50 題 (佔 100%)

D	1. 當有離群值 (Outlier) 時，下列哪一項處理方式最正確？ (A) 將其移除 (B) 不予理會 (C) 使用填補法 (Imputation) (D) 視分析目的而定
B	2. 關於遺缺值 (Missing Value) 的處理，下列敘述哪一項正確？ (A) 即使資料有遺失值也不應該被刪除，所有的資料都應該被完整保存下來 (B) 同一欄位 (column) 中太多資料都含有遺缺值時，不適合將這些資料全都刪除，應該找尋補齊遺缺值的方法 (C) 如果類別變量有遺缺值，則可以使用算術平均數填補該遺缺值 (D) 不論資料分布型態，皆可直接以資料中非缺失值的平均數或中位數來補齊缺失值
C	3. 下列哪一項「不」是資料前處理的步驟？ (A) 資料清理 (Data Cleaning) (B) 資料整合 (Data Integration) (C) 資料建模 (Data Modeling) (D) 資料變形 (Data Reshaping)

111 年度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：111 年 11 月 19 日

第 2 頁，共 16 頁

C	4. 下列哪一項是最「不」適合填補遺缺值（Missing Value）的方式？ (A) 熱卡填補法（Hot Deck Imputation） (B) 迴歸填補法（Regression Imputation） (C) 填補最大值 (D) 平均值填補（Mean/Mode Completer）
A	5. 若要檢查 E-mail 格式是否符合（如：qoo5205566@gmail.com），下列哪一種正則表達式（Regular Expression）可正確判斷？ (A) <code> /^[a-zA-Z0-9._%~]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,4})*\$ /</code> (B) <code> /^[a-zA-Z0-9._%~]+[a-zA-Z0-9.-]+\.[a-zA-Z]{2,4})*\$ /</code> (C) <code> /+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,4})*\$ /</code> (D) <code> /^[a-zA-Z0-9._%~]+@[a-zA-Z0-9.-]*\$ /</code>
D	6. 下列哪一項「不」是 SQL 中的聚合函數（Aggregate function）？ (A) count (B) sum (C) avg (D) cast
D	7. 下列哪一項「不」是非監督式學習的常見方法？ (A) 主成分分析（Principal Components Analysis） (B) 非計量多向度量尺法（Non-metric Multidimensional Scaling） (C) 對應分析（Correspondence Analysis） (D) 線性判別分析（Linear Discriminant Analysis）

111 年度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：111 年 11 月 19 日

第 3 頁，共 16 頁

B	<p>8. 下列哪一種圖形化方法，能直觀的觀察資料集中，不同變數之間的關聯性？</p> <p>(A) 勝率比 (Odds ratio)</p> <p>(B) 平行座標軸 (Parallel coordinates)</p> <p>(C) 目標投影追蹤 (Targeted projection pursuit)</p> <p>(D) 平均數全距管制圖 (X-bar range control chart)</p>
C	<p>9. 關於敘述統計的方法，下列哪一項錯誤？</p> <p>(A) 敘述統計是為了瞭解資料的頻率、趨勢、離散程度等特徵的一種方法</p> <p>(B) 觀察資料的分布、集中特性，如中位數、眾數、平均數、標準差也都是一種敘述統計方法</p> <p>(C) 集群分析 K-means Clustering 是一種將資料分類觀察的敘述統計方法</p> <p>(D) 圖示法，如：直方圖、餅圖、散點圖等，都是一種敘述統計方法</p>
D	<p>10. 下列是進行資料去識別化的動作，請問哪一個行為較「不」恰當？</p> <p>(A) 將只有唯一一個值的欄位刪去</p> <p>(B) 將性別的資料男和女轉為 0 和 1</p> <p>(C) 模擬某欄位的資料分布，重新佈署資料，以保有資料的原有特徵</p> <p>(D) 將性別欄位和年齡欄位隨機混合在一起</p>
A	<p>11. 下列哪一項是監督式 (Supervised) 的特徵工程 (Feature Engineering) 方法？</p> <p>(A) 線性判別分析 (Linear Discriminant Analysis)</p> <p>(B) 主成分分析 (Principal Component Analysis)</p> <p>(C) 潛在語意分析 (Latent Semantic Analysis)</p> <p>(D) 獨立成分分析 (Independent Component Analysis)</p>

111 年度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：111 年 11 月 19 日

第 4 頁，共 16 頁

B	<p>12. 下列哪一種類型資料，適合使用資料增益（Information Gain, IG）進行特徵選取（Feature Selection）？</p> <p>(A) 擁有大量不同數值的資料特徵</p> <p>(B) 名目（Nominal）的資料特徵</p> <p>(C) 非離散化的數值特徵</p> <p>(D) 連續型的數值</p>
A	<p>13. 在使用線性模型時，下列哪一種方法用來將名目（Nominal）類型資料轉為實數（Real Number）類型資料，可以最公平的進行轉換而沒有對特定的可能值（Possible Value）造成偏差？</p> <p>(A) 直接將所有不同數值轉為單一維度的布林值，如將性別（男、女）轉為二維向量男 = (1,0) 及女 = (0,1)</p> <p>(B) 依照名目（Nominal）類型資料的數值資訊轉為相對應的實數值，如將體重（過重、一般、過輕）轉為 (1, 0, -1)</p> <p>(C) 直接將特徵值給予對應的實數值，如將天氣（晴、陰、雨）轉為 (0, 1, 2)</p> <p>(D) 依照特徵值給予範圍內隨機數值，如里程（遠、中、近）分別給予 100~1000（遠）、50~100（中）、0~50（近）的隨機數值</p>
C	<p>14. 在資料準備時，下列敘述哪一項錯誤？</p> <p>(A) 資料準備時，經過資料整合、清理、轉換、減少等步驟架構良好的資料</p> <p>(B) 資料整合包括蒐集資料、選擇資料、整合資料</p> <p>(C) 資料清理不包括減少變數數目、消除不一致、平衡偏斜資料</p> <p>(D) 資料轉換包括正規化資料、分散/整合資料、建構新屬性</p>

111 年度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：111 年 11 月 19 日

第 5 頁，共 16 頁

A	<p>15. 關於特徵（屬性）萃取（Feature Extraction）與轉換（Transformation），下列敘述哪一項正確？</p> <p>(A) 資料縮減泛指屬性挑選（Selection）與萃取（Extraction）</p> <p>(B) 屬性越多，表示後續建模有越多參數要調校，過度配適（Overfitting）的風險越低</p> <p>(C) 各屬性的量綱均一化屬於屬性萃取（Extraction）的工作</p> <p>(D) 主成分分析（Principal Component Analysis, PCA）是分佈偏斜屬性常用的轉換方法</p>
C	<p>16. 如附圖所示之程式碼，若使用 XPath 的語法要選擇在 plate 標籤層下的 apple，下列哪一項錯誤？</p> <pre><div class="table"> <bento></bento> <plate> <apple></apple> </plate> <apple></apple> </div></pre> <p>(A) //plate/apple</p> <p>(B) /div/plate/apple</p> <p>(C) /plate/apple</p> <p>(D) //apple[1]</p>

111 年度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：111 年 11 月 19 日

第 6 頁，共 16 頁

B	<p>17. 若要確保巨量資料運算平台之服務，不會因為單點毀損導致無法存取服務（Single Point of Failure, SPOF），我們會進行高可用性（High Availability, HA）的設計，關於 HA 的敘述，下列哪一項錯誤？</p> <p>(A) 服務層級協議（Service-Level Agreement）決定連續不中斷服務的程度，等級越高表示服務等級越高</p> <p>(B) Hadoop 上的 HDFS（Hadoop Distributed File System）的高可用性可透過配置 Active/Active 兩個 NameNodes 節點解決 SPOF 問題</p> <p>(C) 可以透過 JournalNode 的設計來儲存 HDFS（Hadoop Distributed File System）文件的紀錄，若發生 NameNode 損壞，新的 NameNode 可透過此紀錄恢復既有的文件紀錄</p> <p>(D) 具備有高可用性的架構下，發生 NameNode 損壞時，運行中的程式不受影響，仍會繼續完成工作</p>
B	<p>18. 如附圖所示，Hadoop 最基本架構包含下列哪些項目？</p> <div data-bbox="303 1265 550 1512"><p>(1) Spark</p><p>(2) HDFS</p><p>(3) MapReduce</p><p>(4) Hive</p></div> <p>(A) (1)和(2)</p> <p>(B) (2)和(3)</p> <p>(C) (3)和(4)</p> <p>(D) (1)(2)(3)(4)</p>

111 年度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：111 年 11 月 19 日

第 7 頁，共 16 頁

D	<p>19.若要取得某一網頁的資料，下列哪一個方法最正確？</p> <p>(A) UPDATE</p> <p>(B) HEAD</p> <p>(C) PUT</p> <p>(D) GET</p>										
B	<p>20.如附圖所示，有一個 data 數列，請問經過 MapReduce 模型處理的結果，下列哪一項正確？</p> <p>data = [1,2,3,4,5]</p> <p>data.map(x => x * x).reduce(x, y => x + y)</p> <p>(A) 15</p> <p>(B) 55</p> <p>(C) 25</p> <p>(D) 49</p>										
A	<p>21.如附圖所示為某產品每月市場需求的機率分配，請問該產品每月需求量之期望值，下列哪一項正確？</p> <table><tr><td>需求個數</td><td>230</td><td>360</td><td>460</td><td>500</td></tr><tr><td>機率</td><td>0.2</td><td>0.35</td><td>0.25</td><td>0.20</td></tr></table> <p>(A) 387</p> <p>(B) 420</p> <p>(C) 299</p> <p>(D) 361</p>	需求個數	230	360	460	500	機率	0.2	0.35	0.25	0.20
需求個數	230	360	460	500							
機率	0.2	0.35	0.25	0.20							

111 年度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：111 年 11 月 19 日

第 8 頁，共 16 頁

D	<p>22. 若隨機變數 X 服從二項分配 (Binomial Distribution)，其每次試驗成功的機率為 0.4，試驗次數為 10 次，則 X 的期望值與變異數，下列哪一項正確？</p> <p>(A) 期望值為 2 與變異數為 1.2</p> <p>(B) 期望值為 2 與變異數為 2.4</p> <p>(C) 期望值為 4 與變異數為 1.2</p> <p>(D) 期望值為 4 與變異數為 2.4</p>
	<p>23. 某品牌燈泡平均壽命為 1200 小時，標準差為 36 小時。現購買此燈泡 36 隻，若 Z 為標準常態分配，請問此 36 支燈泡壽命至少為 1180 小時的機率，下列哪一項正確？</p> <p>(A) $\Pr(Z > -2.2)$</p> <p>(B) $\Pr(Z > -2.4)$</p> <p>(C) $\Pr(Z > -2.6)$</p> <p>(D) $\Pr(Z > -2.8)$</p> <p>答案修正</p> <p>本題為計算 36 支燈平均泡壽命，即</p> $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right),$ $P(\bar{X} \geq 1180) = P\left(Z \geq \frac{1180 - 1200}{36/\sqrt{36}}\right) = P(Z \geq -3.33)$

111 年度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：111 年 11 月 19 日

第 9 頁，共 16 頁

C	<p>24. 某公司所生產 10 公斤裝的糖果，其標準差為 0.4 公斤，欲估計母體平均數。若 Z 為標準常態分配，且 $\Pr(Z < -1.96) = 2.5\%$，在 95%信賴水準下，使估計誤差不超過 0.08 公斤，至少應抽多少包糖果來秤重？</p> <p>(A) 62 包</p> <p>(B) 75 包</p> <p>(C) 97 包</p> <p>(D) 116 包</p>
D	<p>25. 有一個隨機樣本數據為：1，3，5，9，11，13，下列敘述哪一項正確？</p> <p>(A) 全距為 5</p> <p>(B) 眾數為 7</p> <p>(C) 中位數為 5</p> <p>(D) 平均數為 7</p>
B	<p>26. 如附圖所示為 Python 語言，mydist 函數回傳是什麼結果？</p> <pre>def mydist(a,b): return abs(a[0]-b[0])+abs(a[1]-b[1]) print(mydist((1,1),(2,2)))</pre> <p>(A) 歐幾里得距離 (Euclidean distance)</p> <p>(B) 曼哈頓距離 (Manhattan distance)</p> <p>(C) 馬氏距離 (Mahalanobis distance)</p> <p>(D) 餘弦相似 (Cosine similarity)</p>

111 年度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：111 年 11 月 19 日

第 10 頁，共 16 頁

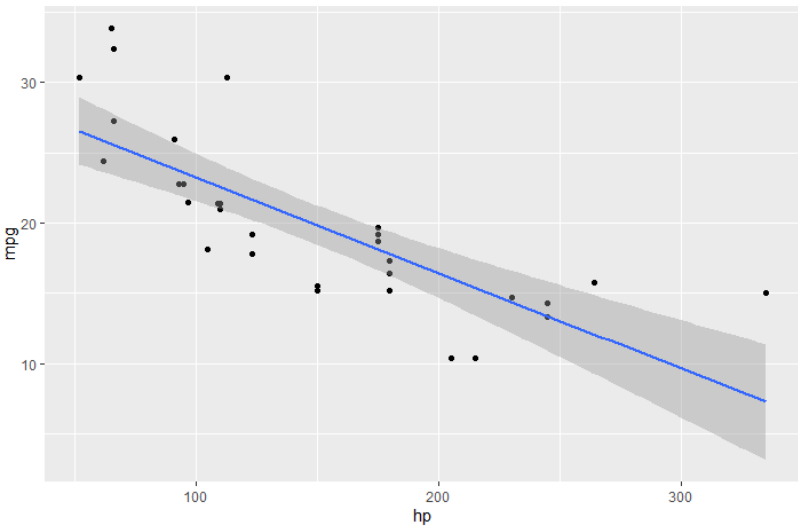
B	<p>27. 某公司生產的燈泡有 10% 的不良率。此公司為了品質對每一個燈泡做檢驗，將其分類為「通過」或「不通過」。若檢驗員有 5% 的機會分類錯誤，則下列哪一項是被分類為「不通過」的百分比？</p> <p>(A) 12%</p> <p>(B) 14%</p> <p>(C) 16%</p> <p>(D) 18%</p>
B	<p>28. 關於關聯法則的敘述，下列哪一項錯誤？</p> <p>(A) 為找出所有頻繁項目集與找出頻繁項目集中具有強關聯規則的規則</p> <p>(B) 從數量低的集合開始，當發現該集合不是頻繁的，則它的母集反而需要考慮</p> <p>(C) FP-growth 算法比 Apriori 算法更有效率</p> <p>(D) 當資料集很大時，Apriori 算法需要不斷掃描資料集造成運行效率很低</p>
D	<p>29. 關於線性迴歸與邏輯斯迴歸，下列敘述哪一項錯誤？</p> <p>(A) 線性迴歸的應變數連續型，邏輯斯迴歸應變數是類別型</p> <p>(B) 線性迴歸的應變數和自變數之間的關係假設是線性相關的，邏輯斯迴歸的應變數和自變數是非線性的</p> <p>(C) 線性迴歸應變數 y 的觀察值是服從常態分布的；邏輯斯迴歸應變數 y 是服從二項分布 0 和 1 或者多項分布的</p> <p>(D) 關於係數估計，線性迴歸採用最大似然法（Maximum Likelihood, ML），邏輯斯迴歸採用最小平方法（Least Square, LS）</p>
D	<p>30. 下列哪一個觀念錯誤？</p> <p>(A) 計算全國失業率時，為了降低誤差，全面調查的結果最為精準</p> <p>(B) 計算全國失業率時，以抽樣方法可以減少計算成本</p>

111 年度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：111 年 11 月 19 日

第 11 頁，共 16 頁

	<p>(C) 計算全國失業率時，以抽樣方法會有計算誤差</p> <p>(D) 計算全國失業率時，以台北市的失業率直接推算其他五個直轄市（新北、桃園、台中、台南、高雄）的失業人數，是精確的估計方法</p>
D	<p>31.關於 PCA 主成分分析與 SVD 奇異值分解的比較，下列敘述哪一項正確？</p> <p>(A) PCA 較 SVD 更一般化</p> <p>(B) SVD 提供資料矩陣只有橫列的基底</p> <p>(C) PCA 提供資料矩陣之縱行與橫列的基底</p> <p>(D) SVD 較 PCA 更一般化</p>
D	<p>32.如附圖所示為 R 語言，執行 ggplot2 套件視覺化分析，下列敘述哪一項正確？</p>  <p>(A) hp 與 mpg 變數呈現正相關</p> <p>(B) 使用以下函數可繪製題目之結果</p> <pre>ggplot(mtcars, aes(hp, mpg)) + geom_point()</pre> <p>(C) hp 與 mpg 變數無明顯相關性</p> <p>(D) 使用以下函數可繪製題目之結果</p>

111 年度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：111 年 11 月 19 日

第 12 頁，共 16 頁

	<pre>ggplot(mtcars, aes(hp, mpg)) + geom_point() + geom_smooth(method = "lm")</pre>
A	<p>33. 若有兩個向量 $A = \langle 2, 0, 0 \rangle$，$B = \langle 2, 2, 1 \rangle$；請問這兩個向量之間的餘弦（Cosine）相似度最接近下列哪一個數字？</p> <p>(A) 0.67</p> <p>(B) 0.33</p> <p>(C) 1</p> <p>(D) 0.5</p>
D	<p>34. 關於主成分分析（Principal Component Analysis, PCA）屬性萃取的主要用途，下列哪一項正確？</p> <p>(A) 萃取出重要的主要主成分後，可以長條圖視覺化多變量資料</p> <p>(B) 可將低度相關的預測變數矩陣 X，轉換成相關且量多的潛在變數集合</p> <p>(C) 可將最攸關的訊息與無關的雜訊結合</p> <p>(D) 可將問題領域中的數個變數，組合成單一或數個具訊息力的特徵變數</p>
B	<p>35. 下列哪一項集群方法，可以解決資料中有離群值及類別屬性的問題？</p> <p>(A) K 平均法（K-means）</p> <p>(B) K 代表點（K-medoids）</p> <p>(C) K 近鄰（K nearest neighbor）</p> <p>(D) K 奇異值分解（K-singular value decomposition）</p>
B	<p>36. 在 R 語言中使用 arules 套件，下列哪一個指令可顯示關聯規則的支持度（Support）、信賴度（Confidence）與增益（Lift）結果？</p> <p>(A) view</p> <p>(B) quality</p>

111 年度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：111 年 11 月 19 日

第 13 頁，共 16 頁

	(C) exam (D) show
C	37. 關於探索式資料分析，下列哪一個是常用來觀察極端值的分佈？ (A) 雷達圖 (Radar Chart) (B) 泡泡圖 (Bubble Chart) (C) 盒鬚圖 (Box Plot) (D) 桑基圖 (Sankey Diagram)
D	38. 關於探索式資料分析，下列哪一個最常用來呈現時間資料趨勢的概念？ (A) 圓餅圖 (Pie Chart) (B) 三元平衡圖 (Ternary Plots) (C) 直方圖 (Histogram) (D) 折線圖 (Line Chart)
C	39. 關於非監督式學習 (Unsupervised Learning)，下列敘述哪一項錯誤？ (A) 在資料集內之變數中，沒有預測目標 (B) 資料劃分為集群，可理解各集群的特性 (C) 迴歸分析和 K-means 兩種演算法常被用來進行資料的分類 (D) 變數間的關係與資料的樣式 (Pattern)，找出資料間分佈的趨勢
D	40. 關於 K 平均法 (K-means)，下列敘述哪一項錯誤？ (A) 需給定 K 群，並隨機挑選 K 個點作為群集中心，亦可給定既有的點來挑選 (B) 分群的好壞在於組內變異 (within) 的值要小，表示同一群裡面的點離散程度小、較密集 (C) 亦可用亂度或是密度來測量取代距離，規律的數據結構下，亂度值是低

111 年度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：111 年 11 月 19 日

第 14 頁，共 16 頁

	<p>的，而隨機性的數據結構的亂度值則是高的</p> <p>(D) 分群後的結果，部分的點可能會分配到多個不同的群</p>
D	<p>41. 關於常用的決策樹（Decision Tree）演算法，下列哪一項錯誤？</p> <p>(A) ID3（Iterative Dichotomiser 3）</p> <p>(B) C4.5</p> <p>(C) CART（Classification and Regression Tree）</p> <p>(D) OLSR（Ordinary Least Squares Regression）</p>
B	<p>42. 關於線性迴歸模型的敘述，下列哪一項錯誤？</p> <p>(A) 線性迴歸方程式，其資料分布之趨勢線不一定是直線</p> <p>(B) 資料共線性（Multicollinearity）問題，不會影響線性迴歸模型的優化</p> <p>(C) 若某資料為標準常態分佈（Standard Normal Distribution），則其標準差（Standard Deviation）為 1</p> <p>(D) 在訓練模型時，需要注意是否出現過度配適（Overfitting）的情形</p>
B	<p>43. 在簡單線性迴歸模型（Linear Regression）當中，斜率的估計值代表下列哪一種意義？</p> <p>(A) 觀察值的預測值</p> <p>(B) 當自變數變動一單位時，依變數的平均變動估計值</p> <p>(C) 當自變數為 0 時，依變數的估計值</p> <p>(D) 當自變數為 0 時，依變數的平均估計</p>
D	<p>44. 監督式學習（Supervised Learning）「不」包含下列哪一項？</p> <p>(A) 分類（Classification）</p> <p>(B) 推估（Estimation）</p> <p>(C) 預測（Prediction）</p>

111 年度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：111 年 11 月 19 日

第 15 頁，共 16 頁

	(D) 分群 (Clustering)													
D	<p>45. 機器學習模型中，關於模型的偏差 (bias) 與變異 (variance)，下列敘述哪一項正確？</p> <p>(A) 高偏差代表模型過於複雜</p> <p>(B) 高變異代表模型過於簡單</p> <p>(C) 模型訓練的目標為低偏差與高變異</p> <p>(D) 偏差與變異之間存在抵換 (trade-off) 關係</p>													
A	<p>46. 下列哪一項「不」屬於監督式學習 (Supervised Learning) ？</p> <p>(A) K 平均法 (K-Means)</p> <p>(B) 決策樹 (Decision Tree)</p> <p>(C) 支援向量機 (Support Vector Machine, SVM)</p> <p>(D) 邏輯迴歸 (Logistic Regression)</p>													
D	<p>47. 如附圖所示，關於二元分類 (binary classification)，若一分類模型產生之混淆矩陣 (confusion matrix)，該模型之精確度 (precision) 為下列哪一項？</p> <table><tr><th colspan="2" rowspan="2"></th><th colspan="2">正確答案</th></tr><tr><th>True</th><th>False</th></tr><tr><th rowspan="2">預測結果</th><th>True</th><td>8</td><td>3</td></tr><tr><th>False</th><td>12</td><td>11</td></tr></table> <p>(A) 3 / 11</p> <p>(B) 8 / 20</p> <p>(C) 19 / 34</p> <p>(D) 8 / 11</p>			正確答案		True	False	預測結果	True	8	3	False	12	11
				正確答案										
		True	False											
預測結果	True	8	3											
	False	12	11											

111 年度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：111 年 11 月 19 日

第 16 頁，共 16 頁

A	<p>48. 使用線性模型（Linear Model）方法時，針對稀疏資料（Sparse data），最適合使用下列哪一種處理方式？</p> <p>(A) 核方法（Kernel Method）</p> <p>(B) 過採樣（Over Sampling）</p> <p>(C) 降採樣（Down Sampling）</p> <p>(D) 交叉驗證（Cross Validation）</p>																														
A	<p>49. 如附圖所示，針對同一份資料建立的四種複迴歸模型，根據各種模型之指標資訊，請問下列哪一個為最佳模型？</p> <table><tr><th>模型編號</th><th>模型</th><th>AIC</th><th>BIC</th><th>C_p</th><th>R²</th></tr><tr><td>模型 1</td><td>$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$</td><td>-55</td><td>50</td><td>3</td><td>0.8</td></tr><tr><td>模型 2</td><td>$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$</td><td>-55</td><td>50</td><td>4</td><td>0.8</td></tr><tr><td>模型 3</td><td>$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$</td><td>-30</td><td>60</td><td>3</td><td>0.8</td></tr><tr><td>模型 4</td><td>$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$</td><td>10</td><td>100</td><td>3</td><td>0.8</td></tr></table> <p>AIC 為赤池信息量準則（Akaike Information Criterion）； BIC 為貝葉斯信息準則（Bayesian Information Criterion）； C_p 為馬洛克斯 C_p（Mallows' Cp）； R² 為判定係數（coefficient of determination）</p> <p>(A) 模型 1</p> <p>(B) 模型 2</p> <p>(C) 模型 3</p> <p>(D) 模型 4</p>	模型編號	模型	AIC	BIC	C _p	R ²	模型 1	$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$	-55	50	3	0.8	模型 2	$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$	-55	50	4	0.8	模型 3	$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$	-30	60	3	0.8	模型 4	$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$	10	100	3	0.8
模型編號	模型	AIC	BIC	C _p	R ²																										
模型 1	$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$	-55	50	3	0.8																										
模型 2	$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$	-55	50	4	0.8																										
模型 3	$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$	-30	60	3	0.8																										
模型 4	$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$	10	100	3	0.8																										
D	<p>50. 關於監督式學習（Supervised Learning）常用的演算法，下列敘述哪一項錯誤？</p> <p>(A) Linear Regression 可以建立自變數與應變數的關係</p> <p>(B) LASSO 是迴歸演算法的延伸，對於精簡的模型有較高的獎勵，並懲罰較複雜的模型</p> <p>(C) 隨機森林是多棵決策樹的集成模型</p> <p>(D) SVM 在特定條件下，可以將資料維度映射到低維度上，藉此找到較佳的分類方式</p>																														