

台灣電力公司 110 學年度大學及研究所獎學金甄選試題

類科：大數據與人工智慧

節次：第一節

科目：人工智慧、機率統計及計算機概論

注意
事項

1. 本試題共 6 頁，採雙面印刷，請注意正、背面試題。
2. 僅限使用簡易型計算器（不限廠牌、型號，功能以不超出 $+$ 、 $-$ 、 \times 、 \div 、 $\%$ 、 $\sqrt{\quad}$ 、MR、MC、MU、M+、M-、GT、TAX+、TAX-之運算為限；其他具有文數字編輯、發聲、振動、記憶儲存、內建程式、外接插卡、通訊或類似功能之計算工具一律禁止使用）。
3. 本試題為單選題共 50 題，每題各 2 分，共 100 分，須用 2B 鉛筆在專業科目答案卡畫記作答，於本試題、英文答案卡或其他紙張作答者不予計分。
4. 測驗式試題均為單選題，每題選項應有 4 個，以(A) (B) (C) (D)標示，請就各題選項中選出最適當者為答案；各題答對得該題所配分數，答錯不倒扣；畫記多於 1 個選項或未作答者，該題不予計分。
5. 考試結束前離場者，試題須隨答案卡繳回，俟該節考試結束後，始得至原試場索取。
6. 考試時間：與英文合併一節考試，共 120 分鐘。

1. 下列有關全球資訊網(World Wide Web)之敘述，何者正確？
(A) 即網際網路，透過統一資源辨識碼(Uniform Resource Identifier, URI)來定位資源
(B) 在西元 1949 年由提姆·柏內茲－李爵士(Sir Timothy Berners-Lee)發明
(C) 統一以超文本標記語言(HyperText Markup Language, HTML)作為傳輸資源之格式
(D) 不收取任何費用
2. 在關聯式資料庫 PostgreSQL 中，若要從資料表「exam」中查詢成績「score」最高之前 10 位學生姓名「name」，應使用下列哪種查詢語句？
(A) SELECT BY score GET name FROM exam FOR HIGHEST 10
(B) SELECT name FROM exam ORDER BY score DESC LIMIT 10
(C) SELECT name FROM HIGHEST score 10 IN exam
(D) SELECT BY HIGHEST score FROM exam FOR 10 THEN SHOW name
3. 下列有關 JavaScript 與 Java 關係之敘述，何者正確？
(A) 兩者為同一人所設計
(B) JavaScript 是 Java 的輕量化版本
(C) 兩者皆為編譯式程式語言
(D) 除了名稱以外兩者沒有多少關聯
4. 美國標準資訊交換碼(American Standard Code for Information Interchange, ASCII)定義了多少個字元？
(A) 26 (B) 100 (C) 128 (D) 256
5. 下列何者屬於電腦的輸入裝置？
(A) 顯示器 (B) 揚聲器 (C) 滑鼠 (D) 中央處理器
6. 下列何者屬於電腦的輸出裝置？
(A) 鍵盤 (B) 電源指示燈 (C) 滑鼠 (D) 中央處理器
7. 有關 HTTPS 與 HTTP 差異之敘述，何者正確？
(A) HTTPS 使用網際網路協定第六版(IPv6)以解決位址枯竭的問題
(B) HTTPS 使用了加密技術以提升安全性
(C) HTTPS 使用了類似多執行緒的技術來建立多個連線，以加快資料交換的速度
(D) HTTPS 是 HTTP 之舊稱

8. 下列有關副檔名之敘述，何者正確？
- (A) 在微軟的 NTFS 檔案系統上，副檔名最多只能有 3 個字元
 - (B) 在檔案總管中，透過修改副檔名，可以改變檔案內容之格式
 - (C) 微軟首先採用以副檔名作為辨識檔案格式之方式
 - (D) 不論在 GNU / Linux 或 Windows 系統上，應用程式都可以無視副檔名，並依照與副檔名不同之檔案格式來讀取檔案
9. 當輸入資料數量越來越多時，擁有下列哪種時間複雜度之演算法會有最差速度？
- (A) $O(1)$
 - (B) $O(n)$
 - (C) $O(2^n)$
 - (D) $O(n^3)$
10. 在 GNU / Linux 系統上，系統管理員帳戶之名稱為何者？
- (A) administrator
 - (B) admin
 - (C) root
 - (D) sudo
11. 在 GNU / Linux 與 Windows 的 PowerShell 上，名為「min」的使用者將如何進入家目錄？
- (A) go home
 - (B) `cd ~min`
 - (C) `cd ~`
 - (D) `cd /`
12. 在 GNU / Linux 系統上，下列何者為絕對路徑？
- (A) `./ documents / note`
 - (B) `./ documents / note`
 - (C) `/ documents / note`
 - (D) `documents / note`
13. 下列有關傳輸控制協定 (Transmission Control Protocol, TCP) 與使用者資料報協定 (User Datagram Protocol, UDP) 之敘述，何者有誤？
- (A) TCP 透過序號來確保封包之順序
 - (B) TCP 透過握手來確保封包之完整性
 - (C) 因 UDP 設計有檢查碼，故較 TCP 更能確保完整性
 - (D) 因 UDP 沒有握手，故速度較 TCP 更快
14. 下列有關 double 與 decimal 差異之敘述，何者有誤？
- (A) decimal 在儲存十進制資料時，有更高的準確度 (accuracy)
 - (B) double 在運算速度上具有優勢
 - (C) decimal 在儲存空間上具有優勢
 - (D) double 在各程式語言中有較高支援度
15. 下列何種方法無助於防止電腦受到病毒侵害？
- (A) 安裝防毒軟體
 - (B) 正確設置防火牆
 - (C) 培養良好的上網習慣，不隨意下載來路不明的檔案
 - (D) 將電腦放在無菌環境
16. 下列有關中間人攻擊 (Man-In-The-Middle attack, MITM) 之敘述，何者正確？
- (A) 中間人駭入資料庫取得加密資料的密碼並竊取資訊
 - (B) 中間人針對一通訊，攔截兩通訊端發出之訊息，並替換成假訊息後再傳送對方
 - (C) 中間人加入群組對話後傳送釣魚網站連結給群組成員
 - (D) 中間人收到郵件後，將含有惡意軟體之檔案夾帶至附件中，並轉寄其他人
17. 自由軟體基金會 (Free Software Foundation) 對於自由軟體 (free software, libre software) 的定義，不包含下列何者？
- (A) 不論目的為何，任何人都有使用該自由軟體的權利
 - (B) 任何人都有重新散布該自由軟體的自由
 - (C) 任何人都可以取得並研究該自由軟體的原始碼
 - (D) 任何人都可以取得該自由軟體的著作權，意即該自由軟體已屬於公共領域

18. 在命令列介面(Command-Line Interface, CLI)上, 若要中斷正在執行的程式, 通常可以使用下列哪種組合按鍵?
- (A) CTRL+C (B) SHIFT+C (C) CTRL+I (D) SHIFT+I
19. 下列何者不是跨站請求偽造(Cross-Site Request Forgery, CSRF)成功之關鍵?
- (A) 拜訪者必須在短期內登入受害網站, 並認證尚未失效
(B) 攻擊者必須建立一個包含惡意網址之網站, 或將惡意網址嵌入到其他論壇上
(C) 受害網站不能使用同步權杖模式(Synchronizer Token Pattern)進行驗證
(D) 拜訪者的瀏覽器不能禁止執行 JavaScript 程式
20. 下列有關數位簽章之敘述, 何者正確?
- (A) 即電子簽章 (B) 具有不可否認性
(C) 沒有法律效力, 但可作為參考 (D) 必須有觸控板進行輔助
21. 下列有關 JavaScript 之敘述, 何者有誤?
- (A) 為原型導向(prototype-based)程式語言 (B) 為強型別(strong typed)程式語言
(C) 為動態型別(dynamic typed)程式語言 (D) 以上皆是
22. 下列有關開放式系統互聯模型(Open System Interconnection model, OSI)之敘述, 何者正確?
- (A) 網際網路協定(Internet Protocol, IP)屬於網路層(Network Layer)
(B) 傳輸控制協定(Transmission Control Protocol, TCP)屬於傳輸層(Transport Layer)
(C) 超文本傳輸協定(HyperText Transport Protocol, HTTP)屬於應用層(Application Layer)
(D) 以上皆是
23. 在 Python 中, 先宣告「arr = [0, 1, 2, 3, 4, 5, 6]», 則「arr[1:5:2]」會回傳下列何者?
- (A) [0, 2, 4] (B) [1, 3, 5] (C) [1, 3] (D) [0]
24. 下列何者是不可能被利用於旁路攻擊(side-channel attack)的訊息?
- (A) 執行時間 (B) 電磁洩漏 (C) 功率消耗 (D) 以上皆是
25. 若訓練時使用數據集之全部特徵, 模型在訓練集上準確率為 100 %, 驗證集上準確率為 70 %, 將出現下列何種問題?
- (A) 配適不足(underfitting) (B) 過度配適(overfitting)
(C) 配適良好 (D) 以上皆非
26. 下列有關資料敘述與摘要統計之內容敘述, 何者有誤?
- (A) 資料抽樣常見的有: 簡單隨機抽樣、系統抽樣、分層隨機抽樣
(B) 進行資料計算與圖表製作, 例如: 次數分配表、直方圖
(C) 衡量資料集中趨勢的統計量, 例如: 平均數、中位數、眾數
(D) 比較兩筆資料的分散程度, 例如: 相關係數
27. 下列有關變異係數(Coefficient of Variation, CV)與標準分數(Z-Score)之敘述, 何者最不正確?
- (A) 變異係數(Coefficient of Variation, CV)為無單位數值, 故適合對兩組不同單位資料的分散程度進行比較
(B) 標準分數(Z-Score)是將原來不同評分尺度的分數, 轉換成具有同一評分尺度的分數, 以利彼此比較和運算之用
(C) 變異係數(Coefficient of Variation, CV)適合對兩組單位相同, 但平均數(Mean)相差很大的資料進行比較
(D) 標準分數(Z-Score)係將原始資料映射到[0,1]區間

28. 在文字探勘技術中，為了使文字資料轉換為電腦可識別之數值資料，資料科學家建立二維結構，其中屬性為字典中所有字詞，而屬性數量為字典中詞彙數量。每篇文章經過斷詞後，將於此結構中建立一筆紀錄，判斷每個字詞在各文章中是否出現。請問此技術運用下列何種屬性萃取(feature extraction)方法？
- (A)單熱編碼(one-hot encoding) (B)分級裝箱(binning)
(C)四捨五入(rounding) (D)對數轉換(log transformation)
29. 下列有關迴歸模型之敘述，何者有誤？
- (A)可用來解釋資料間的因果關係
(B)利用自變數來預測依變數未來可能產生之值
(C)視其函數之型態可分為線性與非線性
(D)根據自變數個數可分為簡單迴歸分析(simple regression analysis)及複迴歸分析(multiple regression analysis)
30. 下列有關求解線性迴歸($Y = \beta_0 X_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m$)之敘述，何者有誤？
- (A)簡單線性迴歸，只包括一個自變量和一個因變量，可使用最小平方法求其解
(B)簡單線性迴歸解得模型是否有效，需計算相關係數 r 以確認 X 對 Y 有顯著影響，並呈現密切的線性相關
(C)多元線性迴歸之自變數間的相關程度，可以高於自變數與因變數間的相關程度
(D)多元迴歸公式之結構等於一個只有輸入和輸出層的神經網路，可用隨機梯度下降法(Stochastic gradient descent, SGD)求出 β 的最佳解
31. 下列何者不是當訓練資料有遺缺資料(missing data)，某些樣本缺少某些維度資料時之處理方式？
- (A)將遺缺資料的樣本移除 (B)以原始資料的平均值補上遺缺資料
(C)將遺缺資料用「合理」的數字補上 (D)將資料分成 n 組交叉檢驗
32. 下列何者不是用來評估模型的驗證指標(validation index)？
- (A)均方誤差(Mean Squared Error, MSE)
(B)混淆矩陣(confusion matrix)與敏感度(sensitivity)
(C)ROC 曲線與曲線下方面積(Area Under Curve, AUC)
(D)特徵值(eigenvalue)
33. 下列何種集群法(clustering)是利用兩兩樣本(或群)間的距離與樹狀結構，將資料進行分群，一開始會將所有資料視為一個完整群體，在迭代過程中不斷分裂為較小群體，直到所有資料都成為單獨個體？
- (A)階層式集群 (B)密度集群 (C) k-means 集群 (D) k-medoids 集群
34. 下列何者不用以處理連續值之預測問題？
- (A)簡單線性迴歸(simple linear regression) (B)多元迴歸分析(multiple regression analysis)
(C)支援向量迴歸(support vector regression) (D)羅吉斯迴歸(logistic regression)
35. 下列有關類神經網路(artificial neural networks)之敘述，何者有誤？
- (A)類神經網路至少包括投入層與隱藏層，投入層之各節點接收資料表中自變數，隱藏層則試圖預測因變數
(B)除了投入層之外，其餘各層神經元節點均需設置活化函數(activation function)
(C)最簡單的類神經網路是感知機(perceptron)，它是一種線性分類模型
(D)近年來分段線性活化函數漸受歡迎，如整流線性單元(Rectified Linear Unit, ReLU)和硬式雙曲正切函數(hard hyperbolic tangent function)大範圍地取代了 S 型函數(Sigmoid function)與雙曲正切函數(hyperbolic tangent function)，係因兩者更適合訓練多層神經網路

36.下列何者不是資料前處理的步驟？

- (A)資料清理(Cleaning) (B)資料操弄(Manipulation)
(C)資料建模(Modeling) (D)資料變形(Reshaping)

37.下列有關資料處理過程(資料採集、資料採礦、資料分析、資料清理)之順序，何者正確？

- (A)資料採集>資料分析>資料清理>資料採礦
(B)資料採集>資料清理>資料分析>資料採礦
(C)資料採集>資料清理>資料採礦>資料分析
(D)資料清理>資料採集>資料採礦>資料分析

38.某家廠商想瞭解影響民眾用電量之可能因素，廠商相信氣溫、風速、是否為假日應會影響民眾用電量。廠商蒐集 30 天用電量資料，算出最小平方迴歸式如下：

$\hat{y} = 250 + 4.5X_1 - 3.0X_2 + 130X_3 + 0.5X_1X_2$ ，其中 X_1 =氣溫， X_2 =風速， $X_3=1$ 為假日、 $X_3=0$ 為非假日。依此迴歸估計式，在某一週末假日之氣溫是 28 度且風速是每小時 12 公里，則估計當天用電量為幾度？

- (A) 638 (B) 656 (C) 583 (D) 684

39.假設今年魷魚遊戲總共吸引 456 人參與，其年齡分布如下表(每組人數比例不含上限年齡)：請問參與者之平均年齡為幾歲(即算術平均數)？(四捨五入至小數點第 2 位)

- (A) 35.04 (B) 37.17 (C) 39.06 (D) 41.23

年齡組	參與人數
0-10	0
10-20	17
20-30	52
30-40	153
40-50	198
50-60	35
60-70	1

40.因應加碼券抽獎活動，小明於網路上勾選「國旅券」，若這週國旅券將抽出 2 組雙位數號碼，對應身份證字號末 2 碼，抽籤方式採 sample without replacement 機制，假設 10 碼身份證之首碼字母有 22 種可能，剩餘 9 碼之每一碼皆包含 0-9 不同數字可能，且每組身份證的中獎機率一樣，則小明中獎機率為下列何者？(四捨五入至小數點第 4 位)

- (A) 0.0201 (B) 0.0239 (C) 0.0328 (D) 0.0209

41.承第 40 題，若抽籤方式改為 sample with replacement 機制，並且加開 1 組 3 位數號碼，則小明的中獎機率為下列何者？(四捨五入至小數點第 4 位)

- (A) 0.0201 (B) 0.0239 (C) 0.0328 (D) 0.0209

42.彩蛋裡有 50 支白籤、5 支紅籤，自彩蛋中抽出一籤。假設有一隨機變數 X ，抽中白籤時， $X = 0$ ，抽中紅籤時， $X = 1$ ，請問 X 之變異數為下列何者？

- (A) 0.04 (B) 0.07 (C) 0.09 (D) 0.1

43.有一枚公正硬幣，請問投擲 3 次至少出現 2 次正面的機率為下列何者？

- (A) 0.1 (B) 0.25 (C) 0.5 (D) 0.75

44. 小明在遊樂場中夾娃娃機，每次 10 元。假設每次夾中的機率固定為 1 成，小明是個好勝心強的人，不贏不歸。請問小明平均要花多少錢才會成功？
(A) 10 (B) 100 (C) 20 (D) 120
45. 假設某一 COVID-19 篩檢方法結果統計，呈陰性反應者佔 80%，呈陽性反應者佔 20%，且呈陰性反應者當中有 10% 為確診者，呈陽性反應者當中有 90% 為確診者，請問確診者被篩檢出陽性的機率為下列何者？(敏感性)(四捨五入至小數點第 2 位)
(A) 0.69 (B) 0.72 (C) 0.53 (D) 0.42
46. 承第 45 題，請問非確診者被篩檢出陽性的機率為下列何者？(偽陽性)(四捨五入至小數點第 2 位)
(A) 0.02 (B) 0.03 (C) 0.04 (D) 0.05
47. 已知 $E[2X+4] = 8$ ， $E[(X+1)^2] = 20$ ，請問 $\text{Var}(X+1)$ 為下列何者？
(A) 8 (B) 9 (C) 10 (D) 11
48. 假設每日確診人數服從期望值為 15 人之卜瓦松分布，請問在一周時間內，確診人數的期望值與變異量分別為下列何者？
(A) 期望值：105、變異量：105 (B) 期望值：105、變異量：15
(C) 期望值：15、變異量：105 (D) 期望值：15、變異量：15
49. 當資料串列通訊採用 Modbus RTU 協定時，假設每個字元含有 8-bit 資料、1 個 start bit、1 個 parity check bit 及 1 個 stop bit，傳輸速率是 9600 bit/s，每次發送訊息前必須先等待 3.5 個字元，則接收端最快接收到下個訊息的時間為下列何者？
(A) 1 ms (B) 2 ms (C) 3 ms (D) 4 ms
50. 在迴歸分析應用中，當分析函式中除了自變數、依變數外尚存在類別變項，最適合的分析應採用下列何者？
(A) 簡單迴歸 (B) 複迴歸 (C) 多元迴歸 (D) 多變量迴歸