

人工智慧：機器學習

Python 3 範例試卷

【認證說明與注意事項】

- 一、本項考試為操作題，所需總時間為 60 分鐘，時間結束前需完成所有考試動作。成績計算滿分為 100 分，合格分數為 70 分。
- 二、操作題為三大題十二小題，第一大題至第二大題每題 30 分，第三大題 40 分，總計 100 分。
- 三、請按照題目要求撰寫程式，再將求取之答案輸入填答視窗中。
- 四、填充題，答案請依題目指示填入數值，若小數點位數換算錯誤或是填答不完整而錯誤者均不予計分。選擇題作答時請填入選項代號 A、B、C、D，勿直接填入答案數值，否則不予計分。
- 五、術科試題內未要求修改之設定值，以原始設定為準，不需另設。
- 六、各題之程式檔必須依題目指示儲存於 C:\ANS.CSF\各指定資料夾備查，測驗結束前必須自行存檔，並關閉開發環境，檔案名稱錯誤或未符合存檔規定及未自行存檔者，均不予計分。
- 七、注意全型、半型字元、英文字母大小寫、小數點的位數是否與題目的要求相同。
- 八、試卷內 0 為阿拉伯數字，O 為英文字母，作答時請先確認。所有滑鼠左右鍵位之訂定，以右手操作方式為準，操作者請自行對應鍵位。
- 九、有問題請舉手發問，切勿私下交談。

壹、操作題 100%(第一題至第二題每題 30 分、第三題 40 分)

一、紅酒種類預測

(一)、題目說明：

1. 請至 C:\ANS.CSF\ML01 資料夾開啟 **MLD01.py**，依下列題意進行組合及改寫，再將求取之答案依序輸入填答視窗中。
2. 檔案請另存新檔為 **MLA01.py**，作答完成請儲存所有檔案至 C:\ANS.CSF\指定資料夾內。
3. 請注意，資料夾或程式碼中所提供的檔案路徑，不可進行變動。

(二)、設計說明：

1. 請使用決策樹分類 (Decision Tree Classifier)，撰寫程式，讀取 **wine.csv**，這個資料集統計紅酒的品質資料，此資料集包含下列欄位：
欄位 0 (Target)：紅酒的分類 (總共分為 3 類，分別為 1~3)
欄位 1-13 (Data)：各種紅酒中各項化學成分檢驗結果，包含如：酒精、蘋果酸、鎂、黃酮、顏色強度、色澤…等等。
2. 請將 75% 的資料做為訓練資料集，25% 的資料用於測試資料集，`random_state=5`。
3. 請用 Data 來預測 Target 的值。
4. 請使用 `round` 函數計算至小數點第二位，印出所訓練分類器的準確度 **【xx.xx】%**，並預測分類。

(三)、請依序回答下列問題：

1. 請填入分類器的準確度 Accuracy of the classifier 為多少% (不需填入%，計算至小數點後第二位)？
2. 輸入資料 [1.51, 1.73, 1.98, 20.15, 85, 2.2, 1.92, .32, 1.48, 2.94, 1, 3.57, 172]，請填入預測分類的選項？
(A) 第 3 類 (B) 無法分類 (C) 第 2 類 (D) 第 1 類
3. 輸入資料 [14.23, 1.71, 2.43, 15.6, 127, 2.8, 3.06, .28, 2.29, 5.64, 1.04, 3.92, 1065]，請填入預測分類的選項？
(A) 第 3 類 (B) 無法分類 (C) 第 1 類 (D) 第 2 類
4. 輸入資料 [13.71, 5.65, 2.45, 20.5, 95, 1.68, .61, .52, 1.06, 7.7, .64, 1.74, 720]，請填入預測分類的選項？
(A) 第 2 類 (B) 第 3 類 (C) 無法分類 (D) 第 1 類

二、資料分群處理

(一)、題目說明：

1. 請至 C:\ANS.CSF\ML02 資料夾開啟 **MLD02.py**，依下列題意進行組合及改寫，再將求取之答案依序輸入填答視窗中。
2. 檔案請另存新檔為 **MLA02.py**，作答完成請儲存所有檔案至 C:\ANS.CSF\指定資料夾內。
3. 請注意，資料夾或程式碼中所提供的檔案路徑，不可進行變動。

(二)、設計說明：

1. 請撰寫一程式，利用 `sklearn.datasets.samples_generator` 裡的 `make_blobs` 函式產生出資料。總樣本數 200，樣本中心數 4，資料集的標準差為 0.50，隨機種子為 0。
2. 請利用 K-means 演算法分群 (K-means Clustering) 來做分群。
3. 使用集群內誤差平方和 (`kmeans.inertia_`) 來判斷分群數量為何值是比較恰當的，值大於 90 以上的納入計算。其中 K-means 內的參數 (分群中心) 初始化為「k-means++」，K-means 演算法的隨機運作次數為「15」，隨機產生中心的隨機序列 (`random state=0`)，最大迭代次數為 200。

(三)、請依序回答下列問題：

5. 請使用集群內誤差平方和 (`kmeans.inertia_`) 來判斷分群數量，需取值大於 90 以上者納入計算，可分為幾群？
6. 承上題，只取集群內誤差平方和分數大於 90 以上的數據，請依據計算後之數據，填入正確的選項 (選項內兩組數據需完全相同)？
(A) [1749.6046, 94.0224] (B) [1794.0460, 304.9484] (C) [1774.5460, 736.9635] (D) [736.9635, 901.0224]
7. 請輸入分群後最小中心點 X 的位置 (四捨五入取至小數點後第四位)？
8. 請輸入分群後最大中心點 Y 的位置 (四捨五入取至小數點後第四位)？

三、鳶尾花品種預測

(一)、題目說明：

1. 請至 C:\ANS.CSF\ML03 資料夾開啟 **MLD03.py**，依下列題意進行組合及改寫，再將求取之答案依序輸入填答視窗中。
2. 檔案請另存新檔為 **MLA03.py**，作答完成請儲存所有檔案至 C:\ANS.CSF\指定資料夾內。
3. 請注意，資料夾或程式碼中所提供的檔案路徑，不可進行變動。

(二)、設計說明：

1. 請撰寫一程式，讀取 sklearn.datasets 中的鳶尾花 (iris) 資料集，此一資料集包含三種不同的鳶尾花品種 (setosa、versicolor、virginica)，儲存為 150*4 的 numpy.ndarray，其欄位如下：
花萼長度 (petal length)
花萼寬度 (petal width)
花瓣長度 (sepal length)
花瓣寬度 (sepal width)
2. 請建立一個 KNN 機器學習模型，用鳶尾花資料集中的四種特性來預測鳶尾花品種，請將測試資料比例設定為 40% (其餘為訓練資料)，random_state=1。
3. 依據給定的 data 預測此鳶尾花品種。
4. 列印測試資料正確率。

(三)、請依序回答下列問題：

9. 請填入測試資料正確率 (四捨五入取至小數點後第四位) ?
10. 給定 data 為 [5, 2.9, 1, 0.2]，請填入鳶尾花品種代號？
(A) virginica (B) setosa (C) none (D) versicolor
11. 給定 data 為 [5.7, 2.8, 4.5, 1.2]，請填入鳶尾花品種代號？
(A) none (B) virginica (C) versicolor (D) setosa
12. 給定 data 為 [7.7, 3.8, 6.7, 2.1]，請填入鳶尾花品種代號？
(A) setosa (B) virginica (C) none (D) versicolor