

[繳交期限]: 2019/6/21 (Fri) 23:59 (注意: 遲交如同未繳交, 都以零分計算)


([建議]: 無論是否能完成所有問題需求, 請務必於期限內, 上傳期末報告電子檔!)

[繳交方式]: 請將期末報告電子檔(ipynb)上傳至 LMS 教學網, 檔案名稱如下:

“MLwP_final_第OO組.ipynb” (例如: MLwP_final_第一組.ipynb)

[報告格式]:

1. 報告內容一開始, 請註明各組組別和組員們的“系所 / 年班 / 學號 / 姓名”。

2. 其餘內容及格式, 請參考下列文件: 

(i) “Scikit-Learn_Workshop_2-SL-Classification_Models.ipynb” (課程教材)

(ii) Manav Sehgal, “Titanic Data Science Solutions”

<https://www.kaggle.com/startupsci/titanic-data-science-solutions>

(iii) Claire Longo, “Digits Recognizer : Neural Network Approach”

<https://www.kaggle.com/statinstilettos/neural-network-approach>

[期末報告問題]: 100%

請將 Breast Cancer Wisconsin (Original) Dataset - UCI : [wisc_bc_data.csv](#) 特徵資料部份先進行正規化 (normalization); 之後, 再分別利用下列演算法建立預測模型:

- (1) Nearest Neighbors [kNN, $k = \sqrt{\text{number of training data}}$]
- (2) Naive Bayes [GaussianNB]
- (3) SVM
- (4) Logistic Regression
- (5) Random Forests
- (6) Neural Networks [MLPClassifier]

其中, 上述各演算法的結果, 必須包括下列輸出:

- (a) Accuracy score for [80% training data & 20% testing data]
- (b) Confusion Matrix for [80% training data & 20% testing data]
- (c) Cross-Validation Results for [cv = 5]
- (d) Leave-One-Out cross-validation result

[加分題]: (Optional)

請將 [wisc_bc_data.csv](#) 特徵資料部份進行 z-score 標準化 (standardization); 之後, 再依上述過程解題。

< NOTE > :

如欲繳交 “加分題” 解答, 請註明並附加在上傳檔案之後!