DME Machine Learning - Home Work 4

1)Write a Python program to read HW4 data file ("*hw4_boston.csv*"). Every line contains 1 dataset, and each dataset contains 104 features and 1 target (housing price). There are 506 datasets.
編寫一個 Python 程序以讀取 HW4 數據文件 "*hw4_boston.csv*"。HW4 數據文件每行是 1 個數據集. 每個數據集包含 104 個 features and 1 個 target (housing price). 整體 506 數據集.

2) Read 506 datasets. Randomly pick 86 datasets as the test set, and use the remaining 420 datasets for 5-fold cross validation training.
讀取整體 506 數據集, 隨機選取 86 個數據集作為測試集,使用剩餘的 420 個數據集進行 5 折交叉驗證訓練
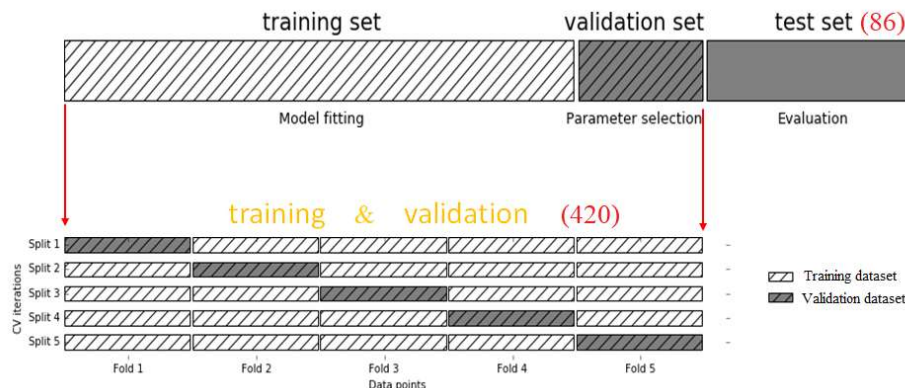
3). Use the *Ridge Regression* algorithm and choose your "alpha" value for the *Ridge Regression*. Your goal is to find an "alpha" value to have your 5-fold cross validation model neither overfitting nor underfitting. You may use the average score of that from the 5 folds.
您的目標是找到一個 "alpha" 值以顯示您的 5 折交叉驗證訓練模型既不過度擬合也不欠擬合。 您可以用 5 folds 中的訓練/測試分數來獲得平均分數。

4) Use your model of the 5-fold cross validation and run the test set (86 datasets). Print the score of your test set.
使用 5 折交叉驗證的模型並運行測試集（86 個數據集）。 打印測試集的分數。

5). Submit your Python code in E3, and the scores of 5 folds training and validation (420 data), and test set (86) score (jpg file sample). 提交你的 Python 代碼, 5 folds 的訓練/驗證的分數, 和測試數據集(86)的分數。 例子顯示。

```
(506, 104)
(420, 104)
(86, 104)
Ridge (alpha 1) Boston, fold 0, Train/Test score: 0.86/0.88
Ridge (alpha 1) Boston, fold 1, Train/Test score: 0.86/0.82
Ridge (alpha 1) Boston, fold 2, Train/Test score: 0.87/0.85
Ridge (alpha 1) Boston, fold 3, Train/Test score: 0.89/0.71
Ridge (alpha 1) Boston, fold 4, Train/Test score: 0.86/0.88

Ridge (alpha 1) Boston, 5-fold Train/Test average score: 0.87/0.83

Ridge (alpha 1) Boston, 5-fold/verify score: 0.87/0.78
```

6). Estimate Time: 4-8 hours   估計所需時間: 4-8 小時

7). Submission Due: before the 11/18/2022 class   截止時間：在 2022 年 11 月 18 日上課之前
 Submit to E3 （" yourID_name_Ridge_HW4.py"） and print-screen 圖 （" yourID_name_Ridge_alpha.jpg"）.