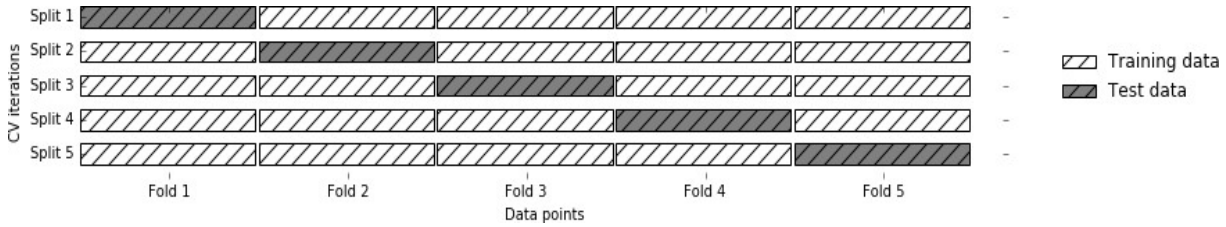
NYCU 2021FALL 5334
DME Machine Learning - Home Work 4

1). 編寫一個 Python 程序以讀取 HW4 數據文件 *Write a Python program to read HW4 data file* ("hw4_boston.csv")。HW4 數據文件每行是 1 個數據集(*Every line contains 1 dataset*). 每個數據集(i.e. 每行 Every line)包含 104 個 features (input data) and 1 個 target data (price). 每個數據都用逗號分隔. 整體 506 數據集. *Each data is separated by a comma and there are 506 datasets.*

2). 讀取整體 506 數據集, *Read 506 datasets.* Use 5 folds for cross evaluation. Test contains 101 個數據集. Training contains 405 個數據集.



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 train/test scores which shows your 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). *Submit your Python code in E3, and the scores of training and test datasets for 5 folds (jpg file). Sample result is shown below.* 提交你的 Python 代碼, 以及 5 folds 的訓練和測試數據集的分數。 例子顯示在此處。

```
========== RESTART: C:\Python38\ml\hw4.py ==========
Ridge0.15, Fold 5, train/test scores: 0.92/0.78
Ridge0.15, Fold 4, train/test scores: 0.91/0.87
Ridge0.15, Fold 3, train/test scores: 0.92/0.90
Ridge0.15, Fold 2, train/test scores: 0.93/0.82
Ridge0.15, Fold 1, train/test scores: 0.91/0.79
>>>
```

5). 估計所需時間 *Estimate Time*: 2-8 小時 *hours*

6). 截止時間 *Due*: 在 2021 年 11 月 19 日上課之前(*before the 11/19 class*)提交您的 python 程序 ("yourID_name_Ridge_HW4.py") 和 print-screen 圖 ("yourID_name_Ridge_alpha.jpg"). The zip file name is "yourID_name_HW4.zip".