系級： 機械碩 姓名：王俞程

請實做以下兩種不同 feature 的模型，回答第 (1) ~ (3) 題：

(1) 抽全部 9 小時內的污染源 feature 當作一次項(加 bias)

(2) 抽全部 9 小時內 pm2.5 的一次項當作 feature(加 bias)

備註：

a. NR 請皆設為 0，其他的數值不要做任何更動

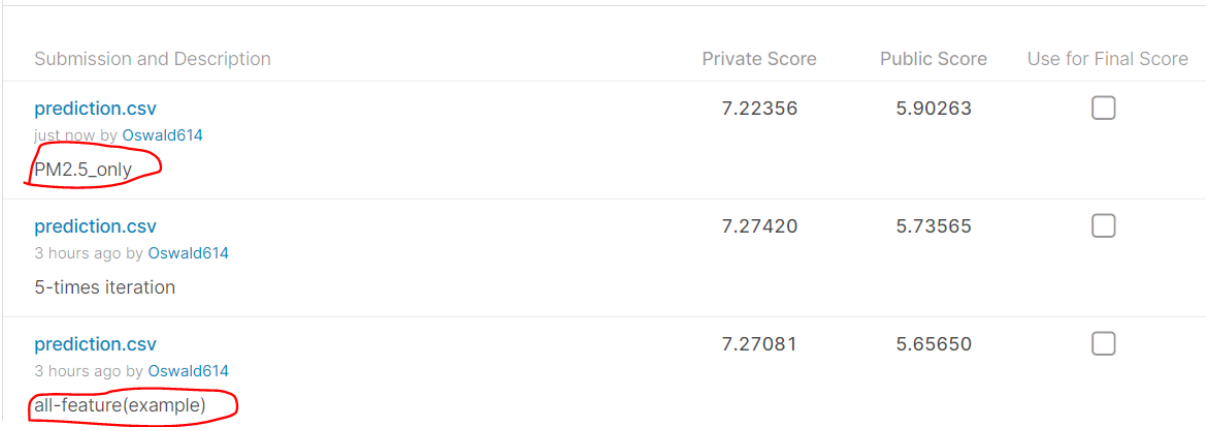
b. 所有 advanced 的 gradient descent 技術(如: adam, adagrad 等) 都是可以用的

c. 第 1-3 題請都以題目給訂的兩種 model 來回答

d. 同學可以先把 model 訓練好，kaggle 死線之後便可以無限上傳。

e. 根據助教時間的公式表示，(1) 代表 p = 9x18+1 而(2) 代表 p = 9*1+1

1. (2%)記錄誤差值 (RMSE)(根據 kaggle public+private 分數)，討論兩種 feature 的影響

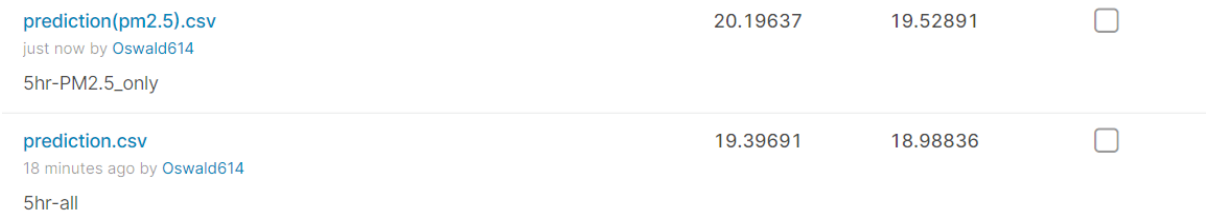


Submission and Description	Private Score	Public Score	Use for Final Score
prediction.csv just now by Oswald614 PM2.5_only	7.22356	5.90263	☐
prediction.csv 3 hours ago by Oswald614 5-times iteration	7.27420	5.73565	☐
prediction.csv 3 hours ago by Oswald614 all-feature(example)	7.27081	5.65650	☐

可以看出在 public 的部分，使用全部汙染源的結果較佳，然而在 private 的部分，則是僅用 PM2.5 的結果較佳。

推測是因為 18 項污染源中，不是每一項皆會與 PM2.5 相關，故在使用 18 項污染源上，會影響預測值。

2. (1%)將 feature 從抽前 9 小時改成抽前 5 小時，討論其變化



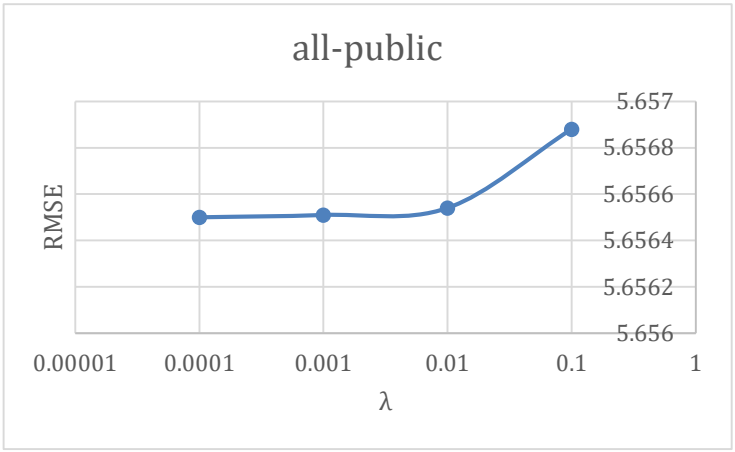
prediction(pm2.5).csv just now by Oswald614 5hr-PM2.5_only	20.19637	19.52891	☐
prediction.csv 18 minutes ago by Oswald614 5hr-all	19.39691	18.98836	☐

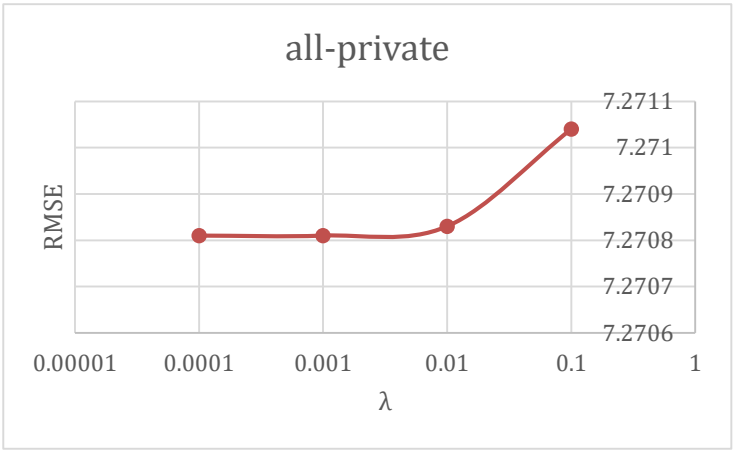
由 9 小時改成 5 小時後，我們可以發現誤差變的非常大。推測是因為原先 9 小時涵蓋 1/3 天，而 5 小時大概 1/5。較無法反映長時間的觀察。

若有足夠測試資料，將會嘗試以抽 24 小時來比較。

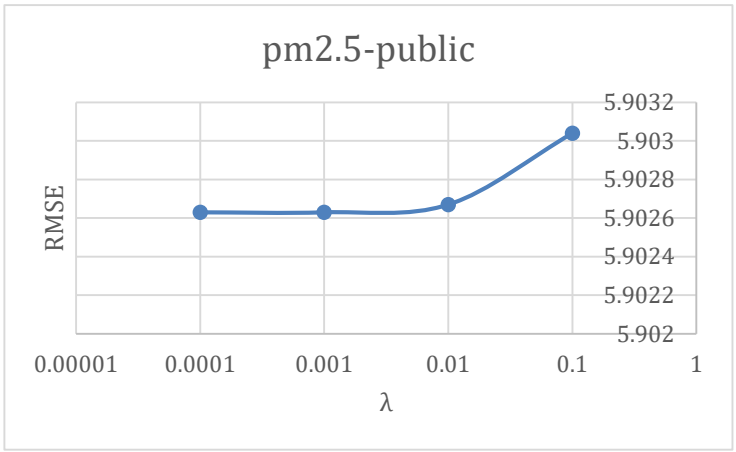
3. (1%)Regularization on all the weight with λ=0.1、0.01、0.001、0.0001，並作圖

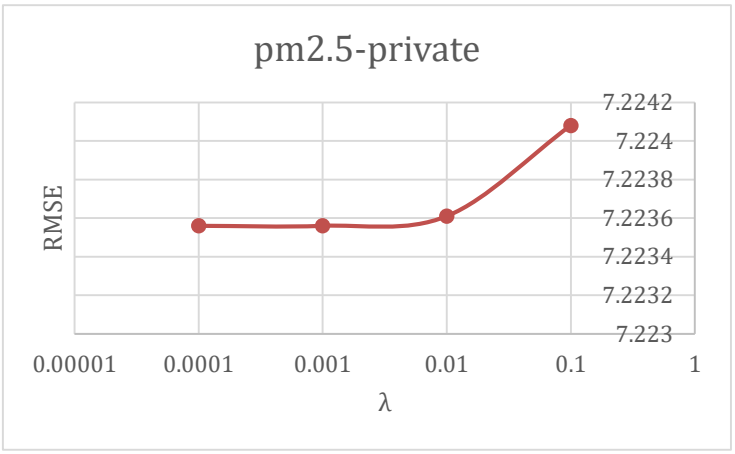
All-feature



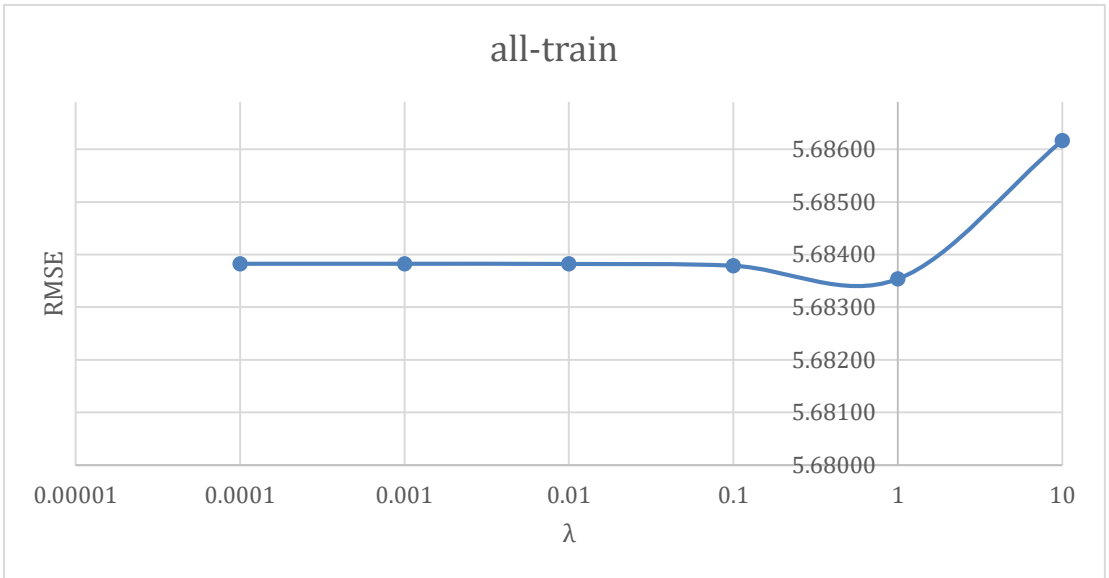


Pm2.5 only





training data



4. (1%)在線性回歸問題中，假設有 N 筆訓練資料，每筆訓練資料的特徵 (feature) 為一向量 x^n，其標註(label)為一純量 y^n，模型參數為一向量 w (此處忽略偏權值 b)，則線性回歸的損失函數(loss function)為$\sum_{□=1}$ $(□^□ - □^□ \cdot □)^2$ 。若將所有訓練資料的特徵值以矩陣 $X = [x^1\ x^2 \dots x^N]^T$ 表示，所有訓練資料的標註以向量 $y = [y^1\ y^2 \dots y^N]^T$ 表示，請問如何以 X 和 y 表示可以最小化損失函數的向量 w ？請選出正確答案。(其中 X^TX 為 invertible)

(a) $(X^TX)X^Ty$
(b) $(X^TX)yX^T$
(c) $(X^TX)^{-1}X^Ty$
(d) $(X^TX)^{-1}yX^T$

C
因 Y-X·W=0 ，故 W = $X^{-1}Y$ → C

最佳成績

Submission and Description	Private Score	Public Score	Use for Final Sco
prediction.csv a month ago by Oswald614 all_ransa_1std	7.02875	5.55573	☐
prediction.csv	7.07700	5.52450	☐