學號：B06901063 系級： 電機二 姓名：黃士豪

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

| feature 類型 | 全部污染源 | 只有 PM2.5 的一次項 |
|---|---|---|
| public 分數 | **5.52052** | **5.79810** |
| private 分數 | **7.07749** | **7.10913** |

選用全部污染源的誤差較小，推測是因為 feature 較多，不會受到單獨資料的影響，且 PM2.5 濃度也非一獨立事件。

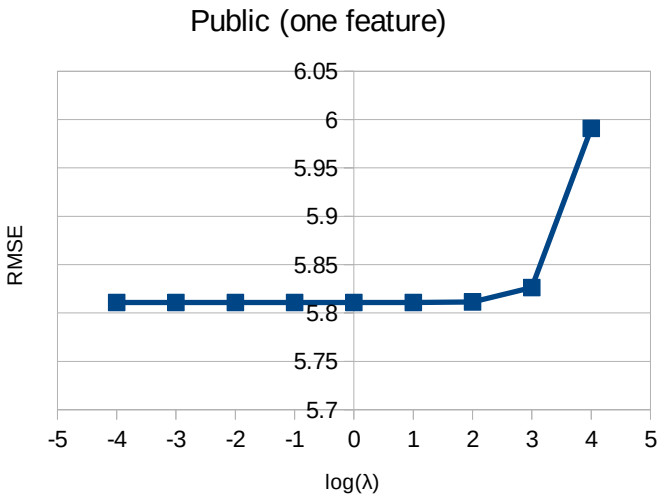
2. (1%)將 feature 從抽前 9 小時改成抽前 5 小時，討論其變化

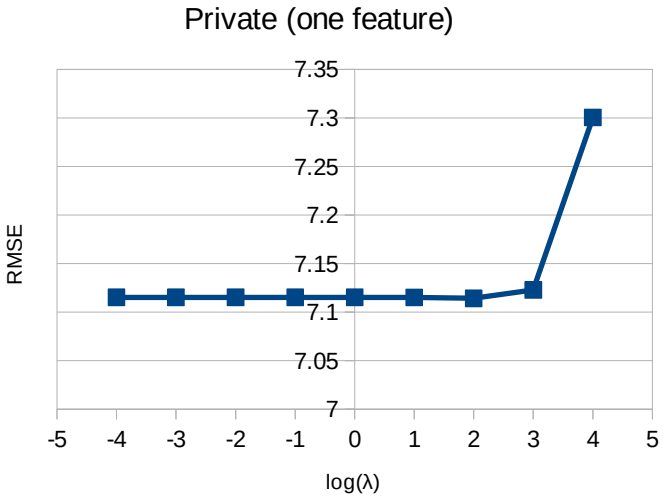
| 分數(public/private) | 全部污染源 | 只有 PM2.5 的一次項 |
|---|---|---|
| 9 小時 | **5.52052 / 7.07749** | **5.79810 / 7.10913** |
| 5 小時 | **5.93168 / 7.06710** | **6.18407 / 7.12622** |

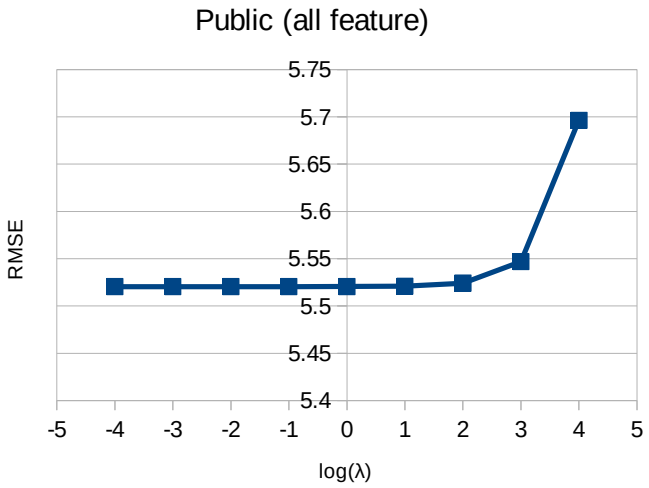
抽取前 9 小時得出的結果普遍較好，推測是因為有更多的資訊去進行 model 建立，使得預測較為準確。

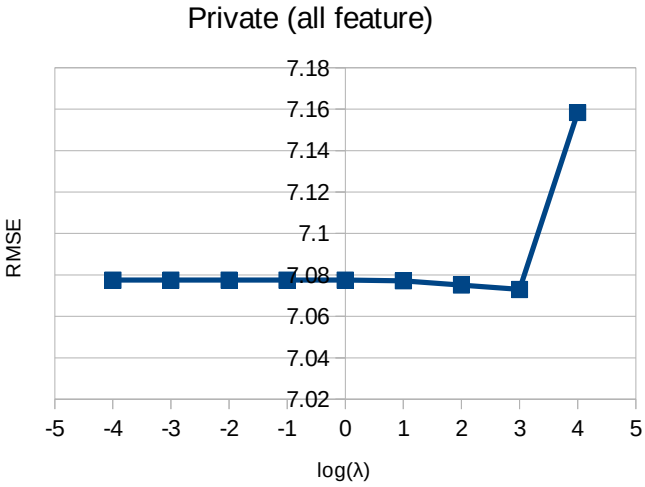
3. (1%)Regularization on all the weight with λ=0.1、0.01、0.001、0.0001，並作圖

| 分數(public/private) | 全部污染源 | 只有 PM2.5 的一次項 |
|---|---|---|
| λ = 10000 | **5.69632 / 7.15834** | **5.99082 / 7.30048** |
| λ = 1000 | **5.54694 / 7.07299** | **5.82638 / 7.12283** |
| λ = 100 | **5.52418 / 7.07512** | **5.8114 / 7.1142** |
| λ = 10 | **5.5209 / 7.07717** | **5.81083 / 7.11517** |
| λ = 1 | **5.52056 / 7.07746** | **5.8108 / 7.1153** |
| λ = 0.1 | **5.52052 / 7.07749** | **5.8108 / 7.11532** |
| λ = 0.01 | **5.52052 / 7.07749** | **5.8108 / 7.11532** |
| λ = 0.001 | **5.52052 / 7.07749** | **5.8108 / 7.11532** |
| λ = 0.0001 | **5.52052 / 7.07749** | **5.8108 / 7.11532** |




由上可知，λ過小時對結果並沒有顯著的影響，但當λ慢慢增大，雖然 public 部份的測資誤差會加大，private 部份卻會減少。當達到一定大小後，便必定造成誤差增加。

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 $\sum_{n=1}^{N}\left(y^n - x^n \cdot w\right)^2$ 。若將所有訓練資料的特徵值以矩陣 X = $[x^1\ x^2\ \ldots\ x^N]^T$ 表示，所有訓練資料的標註以向量 y = $[y^1\ y^2\ \ldots\ y^N]^T$ 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) $(X^TX)^{-1}yX^T$

Ans : C