學號：B05901025 系級： 電機三 姓名：王鈺能

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

Features	Public	Private	Total
18*9+1	5.55975	7.10912	6.38163
9+1	5.78809	7.11616	6.486205

根據 Kaggle 分數，用 9*18+1 維 feature 的估測其準確度比僅用 9+1 維高，這是一個合理的結果，因為其他汙染源的確和 PM2.5 的值是有相關的，所以用所有污染源做估測，準確度的確應該比較高。

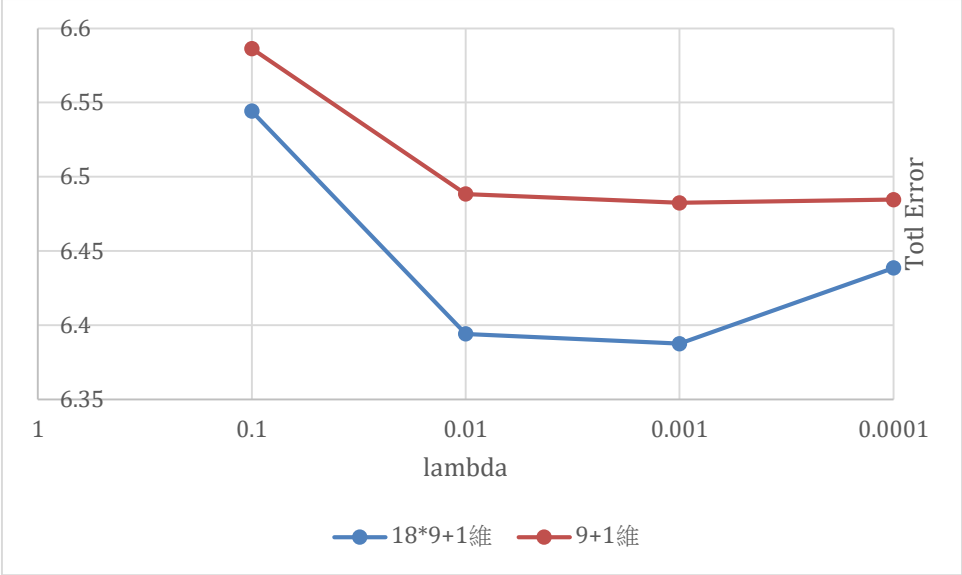
2. (1%)將 feature 從抽前 9 小時改成抽前 5 小時，討論其變化

Features	Public	Private	Total
18*5+1	5.97342	7.11563	6.569396
5+1	6.1732	7.11605	6.661327

和上一小題相比，Total Error 皆上升，因為資料量變少了，這是很合理的。但值得注意的一點是，用 18*5+1 維的 training 結果比 9+1 維還差，這可能代表估測 PM2.5 最主要的參數就是前幾小時的 PM2.5，其他參數相對影響較小。

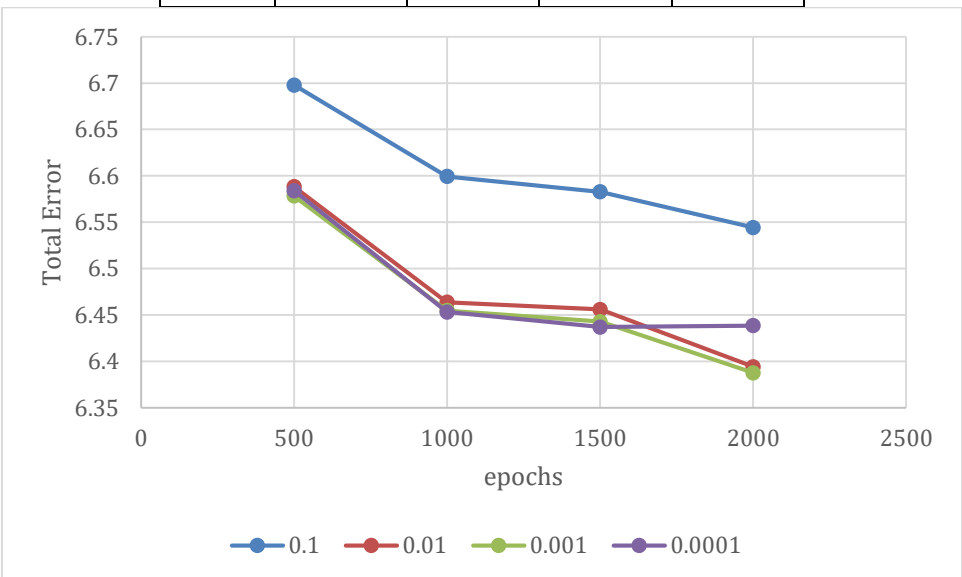
3. (1%)Regularization on all the weight with λ=0.1、0.01、0.001、0.0001，並作圖

	0.1	0.01	0.001	0.0001
18*9+1	6.544219	6.394096	6.387526	6.438566
9+1	6.586393	6.488347	6.482463	6.484592



在 λ=0.1 時的 Error 比其他三者大，可能是 regulation 太大導致 model 一開始就沒 trian 起來，λ 變小後 Error 有下降，但沒有比 λ=0 時好，我覺得是因為這題 model 太簡單，overfitting 的狀況不明顯，所以 regulation 沒有什麼效果。

epochs	500	1000	1500	2000
0.1	6.697858	6.599355	6.582838	6.544219
0.01	6.588283	6.463727	6.456016	6.394096
0.001	6.578298	6.45481	6.442883	6.387526
0.0001	6.583982	6.453157	6.436906	6.438566



進一步討論不同 λ 和 training epochs 的關係(batch size=32)，可以看出 λ=0.1 時無論在 epoch 多或少 Error 比較大，我認為理由同上，λ 大的話比較難 train 出好的 model，再者這題可能 overfit 情況不明顯，所以 λ 小表現較好。
(附註)Error 為 Kaggle 分數算出的 RMSE

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$\sum_{n=1}^{N} (y^n - x^n \cdot w)^2$ 。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) $(X^TX)^{-1}yX^T$

(C)