學號：B05901068 系級：電機三 姓名：蕭如芸

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

	(1) 18 種污染源	(2) PM2.5
Public	5.63401	5.90263
Private	7.21528	7.22356
RMSE	6.47311	6.59624

取 18 種污染源在 public 和 private 都比只取 PM2.5 的誤差小，可見 PM2.5 以外的污染源可以提供預測 PM2.5 的資訊，不同污染源之間有相關性。

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

	(1) 18 種污染源	(2) PM2.5
Public	5.98170	6.22732
Private	7.16701	7.22552
RMSE	6.60101	6.74491

(1) 18 種污染源：在 public dataset 上，取前 9 小時的誤差比取前 5 小時的誤差小；但在 private dataset 上，取前 5 小時的誤差比較小。推測前 5 小時的資料可能已經提供足夠的資訊預測下一個小時的 PM2.5。

(2) PM2.5：在 public dataset 上，取前 9 小時的誤差比取前 5 小時的誤差小；在 private dataset 上，取前 9 小時的誤差略小於取前 5 小時的誤差。前 5 小時的資

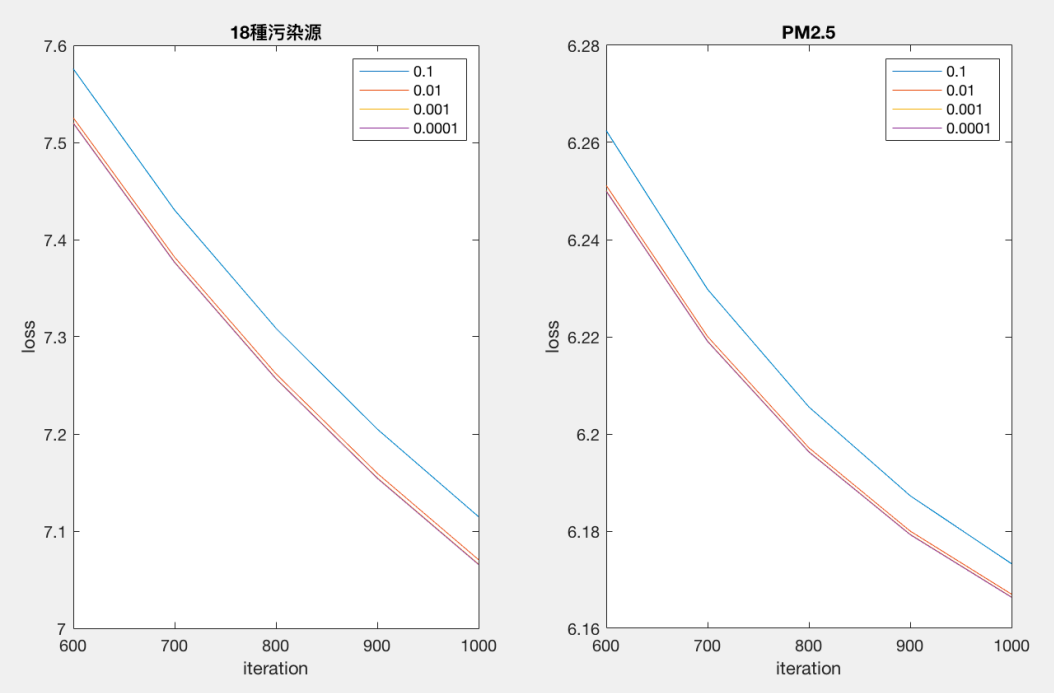
料可以提供一定的資訊，但由於只取 PM2.5 作為 feature，取更多小時的資訊可以得到更小的誤差。

若比較 1. 2. 題四種 feature 在 private dataset 上的誤差，結果如下：

18 種污染源, 5 小時 < 18 種污染源, 9 小時 < PM2.5, 9 小時 < PM2.5, 5 小時

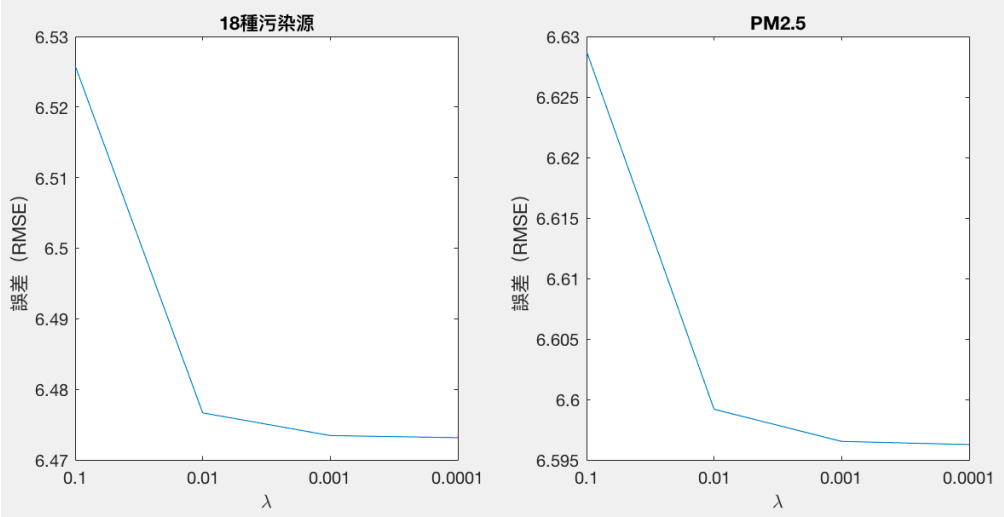
3. (1%)Regularization on all the weight with λ=0.1、0.01、0.001、0.0001，並作圖

為了看出不同 λ 值在 training 過程中 loss 的變化，取 iteration 600 至 1000 之間的 loss 作圖，此時 weight 尚未收斂，較能看出不同 λ 值對 loss 的影響。λ 越大，training 過程中的 loss 越大。



根據 kaggle 上 public 和 private 的分數，計算不同 λ 值在 testing data 上的誤差值（RMSE）。

λ	(1) 18 種污染源	(2) PM2.5
0.1	6.52587	6.62879
0.01	6.47667	6.59919
0.001	6.47345	6.59653
0.0001	6.47315	6.59627



不論是取 18 種污染源或是只取 PM2.5 作為 feature，λ 越小，testing 時的誤差越小。和第 1. 題的結果比較，沒有 regularization 的結果更好。在我的實驗結果中，regularization 沒有改進預測結果，可能原因是我的模型只有使用一次項，不太會發生 overfit 的現象。

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 $X = [x^1\ x^2 \cdots x^N]^T$ 表示，所有訓練資料的標註以向量 $y = [y^1\ y^2 \cdots y^N]^T$ 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)

$$L = (y - X \cdot w)^2$$

$$\frac{\partial L}{\partial w} = 2 \cdot X^T \cdot (y - X \cdot w) = 0$$

$$X^T y - X^T X w = 0$$

$$X^T y = X^T X w$$

$$w = (X^T X)^{-1} X^T y$$