學號：R05522621 系級： 機械碩二 姓名：李哲銘

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 的一次項(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

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

feature (9 hours)	kaggle public	kaggle private	RMSE
All	7.48225	5.28980	6.479431
PM2.5	7.44013	5.62719	6.596241

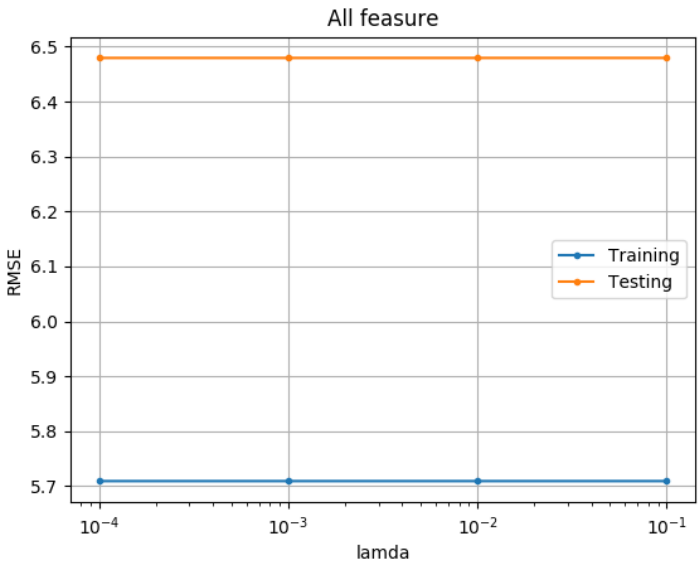
上表之結果可發現兩者之結果差異並不大，feature 數量對一次項 model 影響並不大，但在抽全部汙染源的表現似乎較好，由此推估在一次項 model 裡應有其他汙染源會影響第十小時的 PM2.5 結果。

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

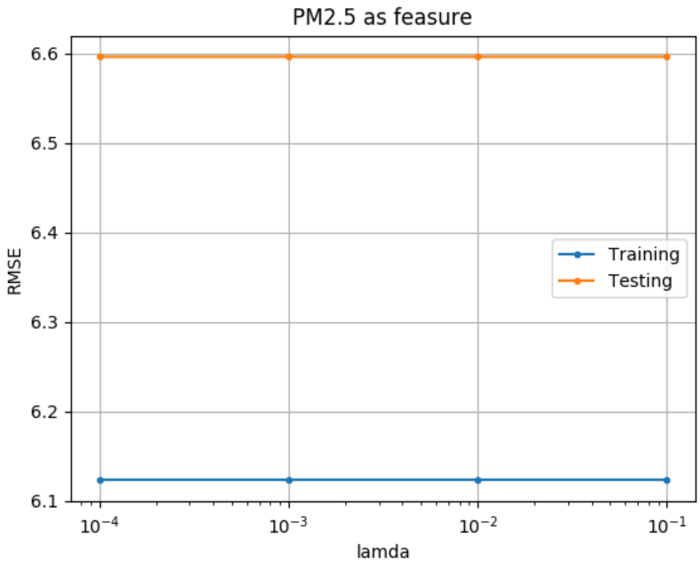
feature (5 hours)	kaggle public	kaggle private	RMSE
All	7.66479	5.32895	6.601012
PM2.5	7.57904	5.79187	6.744909

與第一題之結果比較可知，不管是抽全部汙染源還是只抽 PM2.5，只取連續 5 小時之誤差皆較大，如果是抽全部汙染源可以發現誤差變得更大，因此我認為前九小時皆會影響第十小時之結果。

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



λ=0.1	λ=0.01	λ=0.001	λ=0.0001
5.709383(train)	5.709381	5.709381	5.709381
6.479429(test)	6.479431	6.479431	6.479431



λ=0.1	λ=0.01	λ=0.001	λ=0.0001
6.123022(train)	6.123022	6.123022	6.123022
6.596240(test)	6.596241	6.596241	6.596241

與第一題比較可發現沒有 Regularization 與 Regularization 後之結果改變非常微小，且 λ=0.1 與 λ=0.0001 之差距約只差小數點後五位或六位，我認為應是 λ 不大，因此對訓練出來模型的影響並不大。

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 ？請寫下算式並選出正確答案。(其中 X^TX 為 invertible)

(a) $(X^TX)X^Ty$
(b) $(X^TX)^{-0}X^Ty$
(c) $(X^TX)^{-1}X^Ty$
(d) $(X^TX)^{-2}X^Ty$

正確答案為(c) $(X^TX)^{-1}X^Ty$，令 $L = \sum_{n=1}^{N}(y^n - x^n \cdot w)^2$ ，將損失函數對 w 偏微分等於 0 得

$$\frac{\partial L}{\partial w} = -2\sum_{n=1}^{N}(y^n - x^n \cdot w)\,x^n = [0 \cdots 0]^T$$ ，化簡後得 $$\begin{pmatrix} y^1 - x^1 \cdot w \\ \vdots \\ y^N - x^N \cdot w \end{pmatrix}^T \begin{pmatrix} x^1 \\ \vdots \\ x^N \end{pmatrix} = (y - Xw)^T X = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}^T$$ ，

將 X 乘進去得 $(X^T y - X^T Xw)^T = [0 \cdots 0]^T$ ，移項得 $X^T y = X^T Xw$ ，因 X^TX 為 invertible，可得 $w=(X^TX)^{-1}X^Ty$，因此(c)即為所求。