學號：B04901136 系級： 電機三 姓名：張家銘

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

A.全部 9 小時內的污染源 feature 的一次項，RMSE：5.62719+7.44013=13.06732

B.全部 9 小時內 PM2.5 的一次項，RMSE：5.53562+7.46237=12.99799

兩種情況相較看起來， 全部 data：bias 較高、variance 較低

PM2.5：bias 較低、variance 較高

所以可以推測 9 個小時 PM2.5 資料是比較好的 model，但是因為 training data 比較少，所以要有 overfitting 的狀況，如果想要再改善此結果可以用 regulation。

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

A.全部 5 小時內的污染源 feature 的一次項，RMSE：5.44092+7.65925=13.10017

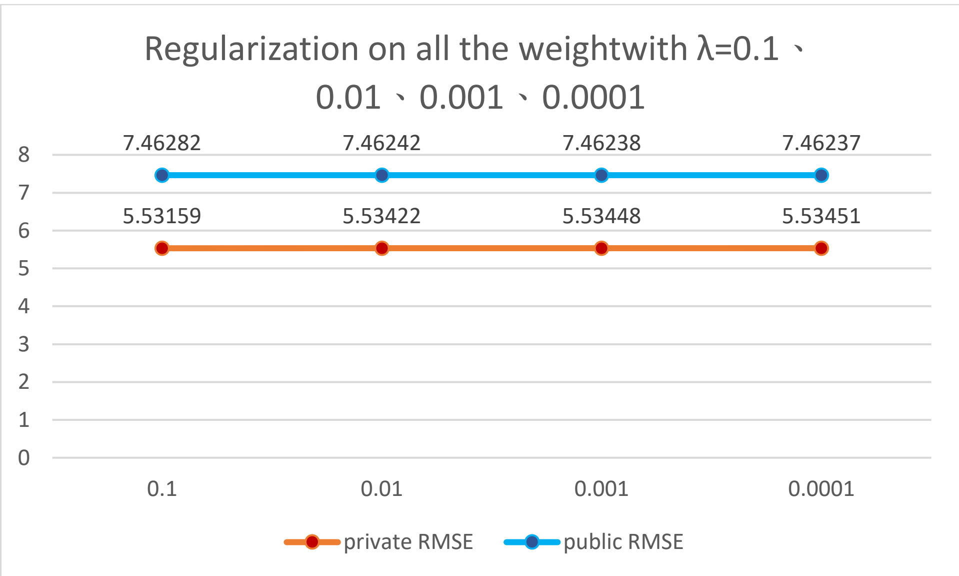
B.全部 5 小時內 PM2.5 的一次項，RMSE：5.79187+7.57904=13.37091

兩種情況相較看起來， 全部 data：bias 較低、variance 較高

PM2.5：bias 較高、variance 較低

所以可以推測 5 個小時全部 data 是比較好的 model，但是因為 training data 比較少，所以要有 overfitting 的狀況，如果想要再改善此結果可以用 regulation。

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



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)為一存量 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$\sum_{\square=1}^{\square} (\square^{\square} - \square^{\square} \cdot \square)^2$ 。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) $(X^TX)^{-2}X^Ty$

假設每筆訓練資料的 feature 數為 k

假設 d 是一個 shape = (N, 1)矩陣，

假設 u 是一個 shape = $(k, 1)$矩陣 $\rightarrow \mathrm{y}^n = x^n \times u + d$

以${x^n}_{c=i}$表示x^n第i個 colomn，

Suppose ${x^n}_{c=j} \cdot d \neq 0$ ，$exist\ u', d'$

$$s.t. x^n \cdot d' = 0\ , d = x^n \times \mathrm{u}' + \mathrm{d}'$$

$because\ x^n \cdot d' = 0$

$\rightarrow norm(d) = norm(x^n \times \mathrm{u}') + norm(d')$

$\rightarrow norm(d) > norm(d') \rightarrow d\ is\ not\ minimum$

$\rightarrow contradiction$

$\forall i\ {x^n}_{c=j} \cdot d = 0$

${x^n}^T(y^n) = {x^n}^T(x^n \times u + d) = {x^n}^T \times x^n \times u + {x^n}^T \times d$

$because\ \forall i\ {x^n}_{c=j} \cdot d = 0 \rightarrow {x^n}^T \times d = 0$

$\rightarrow {x^n}^T(y^n) = {x^n}^T \times x^n \times u$

$\rightarrow u = ({x^n}^T \times x^n)^{-1} \times {x^n}^T(y^n)$ 即為所求

Answer is C