學號：R06945003 系級： 生醫電資碩一 姓名：林鈺盛

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

(1) public: 7.46583 private: 5.43321

(2) public: 7.44013 private: 5.62719

綜合 public 以及 private 兩個分數，第一種 model 也就是取全部污染源當作 feature 有比較準確的預測率，不過其實兩者的 error 差距不大，因此很難去討論兩種 model 的影響，不過可以知道的是取所有污染源當作 feature 可以在 training data 上面有較小的 error，但同時比較有 overfitting 的風險。

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

(1) public: 7.66969 private: 5.40540

(2) public: 7.57904 private: 5.79187

本題也是第一種 model（全部污染源）的情況有較準確的預測，而其中比較驚訝的是 private 的部分居然有 5.40540 的誤差，這也是我本次作業中最好的結果，可惜的是我並沒有選擇此結果。將九小時改成五小時，可以發現在 public 的部分誤差稍稍增加，private 的部分整體來說誤差也是增加的，因此我認為本次作業抽前九小時應該會比前五小時能夠得到更正確的預測結果。

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

本題結果如下表所示，因為作圖出來結果為六條水平線，看不出趨勢變化，在這邊就先不作圖。在 iteration 設為 100 萬次的情況下，調整λ的值對於結果的影響非常非常小。

total_feature			
lambda	training	public	private
0.0001	5.68439296	**7.46583**	**5.43321**
0.001	5.68439336	**7.46583**	**5.43321**
0.01	5.68439744	**7.46583**	**5.43321**
0.1	5.68443818	**7.46583**	**5.43321**
on_PM2.5			
lambda	training	public	private
0.0001	6.12306172	**7.44013**	**5.62719**
0.001	6.12342349	**7.44013**	**5.62719**
0.01	6.12704	**7.44013**	**5.62719**
0.1	6.16308844	**7.44013**	**5.62719**

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 $\mathbf{x}^n$，其標註(label)為一存量 $\mathbf{y}^n$，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$\sum_{n=1}^{N}(y^n - w \cdot x^n)^2$。若將所有訓練資料的特徵值以矩陣 $\mathbf{X} = [\mathbf{x}^1\ x^2 \ldots x^N]^T$ 表示，所有訓練資料的標註以向量 $\mathbf{y} = [\mathbf{y}^1\ y^2 \ldots y^N]^T$ 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 $\mathbf{X}^T\mathbf{X}$ 為 invertible)

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

$$L = \sum_{n=1}^{N}(x_n^T w - y_n)^2$$
$$L = ||Xw - y||^2$$
$$\nabla L = 2(X^TXw - X^Ty) = 0$$

因此可以得到

$$w = (X^TX)^{-1}X^Ty$$

故本題答案為(c)