學號：B04902021 系級： 資工三 姓名：陳弘梵

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)

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

	public test loss	private test loss	public + private
所有汙染源一次項	7.46718	5.40110	6.516542732
PM2.5 一次項	7.45924	5.63440	6.610095488

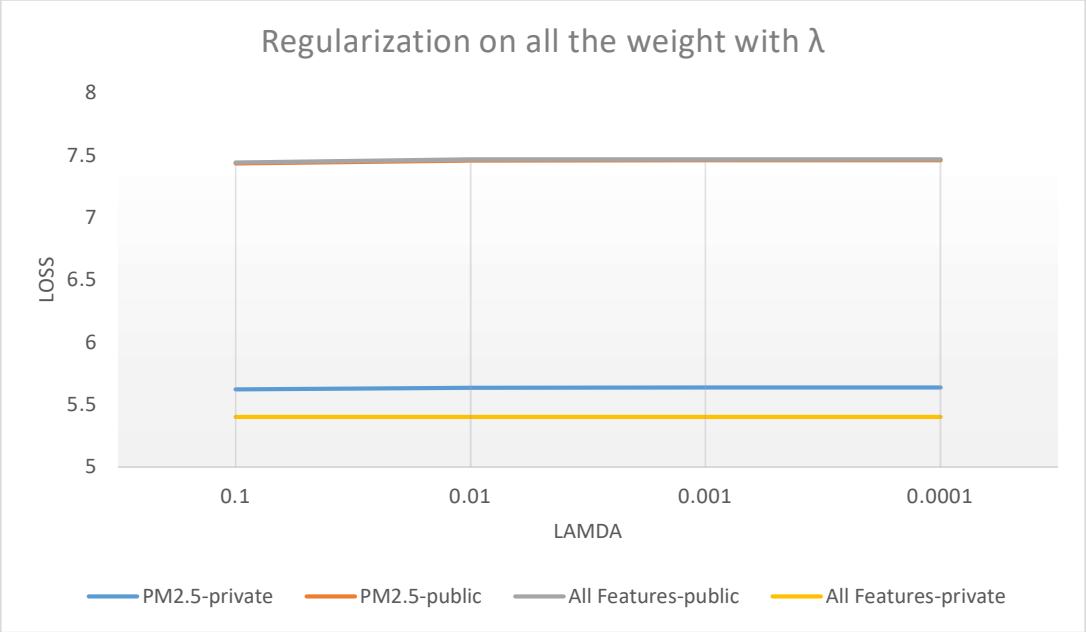
由我寫的程式跑出來，所有汙染源採用的 training loss 為 5.68786，較僅使用 PM2.5 的 6.12363 來的小，有可能是 iteration 的大小設置或是到達極值的判斷寫法造成。我判斷結束是檢查現在新的 loss 是否仍比上次的小，iteration 則是設為 10 萬。也有採用 minimum normalization。推估是因為抽取的 feature 較多，與 PM2.5 之間的關係較為密切，僅僅以其他時段的 PM2.5 預測缺少了許多真正的因子。

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

	public test loss	private test loss
所有汙染源一次項	7.65465	5.39988
PM2.5 一次項	7.61130	5.80136

我的程式中，採取所有汙染源的結果仍然表現比較好，整體而言相較於抽取 9 小時的表現較差，應該是參考的連續數據較少導致與剩下的資料關聯較低。由此推估：採用汙染源項目越多、連續抽取小時數做為 feature 越多，所能得到的成功越接近真實。

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



在圖中所有汙染物項目與 PM2.5 的 Public 線幾乎重合，與第一題所列數據接近。從圖可以看出 Regularization 採用的這四個 lamda 值之間，改動對此訓練影響不大。

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 $X = [x^1\ x^2 \dots x^N]^T$ 表示，所有訓練資料的標註以向量 $y = [y^1\ y^2 \dots y^N]^T$ 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) $(X^TX)^{-2}X^Ty$

$$loss = (Y - w^TX)(Y - w^TX)^T = \sum \begin{pmatrix} (y^1 - wx^1)^2 \\ (y^2 - wx^2)^2 \\ \dots \\ (y^N - wx^N)^2 \end{pmatrix}$$

找極值，微分=0:

$$\begin{pmatrix} x^1(y^1 - wx^1) \\ x^2(y^2 - wx^2) \\ \dots \\ x^N(y^N - wx^N) \end{pmatrix} = \begin{pmatrix} x^1y^1 - wx^{1^2} \\ x^2y^2 - wx^{2^2} \\ \dots \\ x^Ny^N - wx^{N^2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \dots \\ 0 \end{pmatrix} \Rightarrow X^TY - wX^TX = 0$$

$$\Rightarrow w = (X^TX)^{-1}X^TY$$

ANS: (C)