學號：B04901060 系級： 電機三 姓名：黃文璁

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

9 小時	public	private
全部汙染源	7.46237	5.53562
PM2.5	7.44013	5.62719

答： 9 小時全部汙染源的 public RMSE 比只抽 PM2.5 的 RMSE 來得高（約高 0.02）。
9 小時全部汙染源的 private RMSE 比只抽 PM2.5 的 RMSE 來得低（約低 0.1）。
由於 public、private 沒有絕對高低，兩種 feature 就本題來說不容易分出好壞。

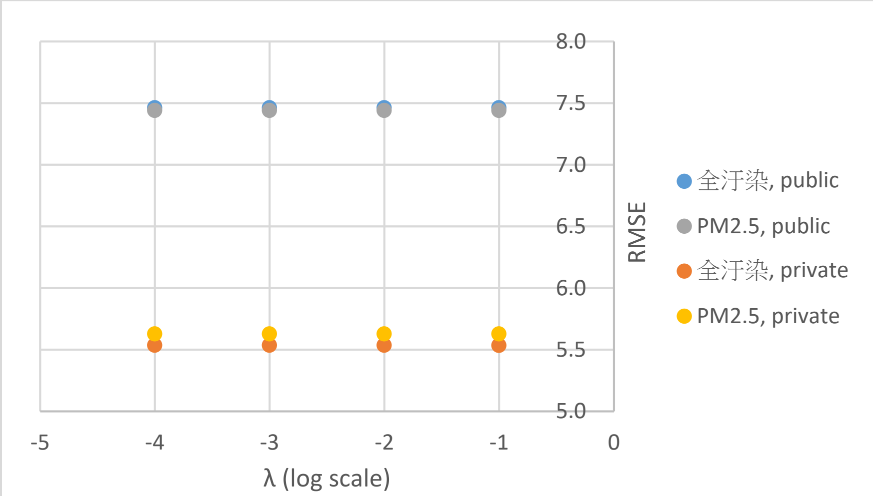
2. (1%)將 feature 從抽前 9 小時改成抽前 5 小時，討論其變化

5 小時	public	private
全部汙染源	7.65925	5.44092
PM2.5	7.57904	5.79187

答： 5 小時全部汙染源的 public RMSE 比只抽 PM2.5 的 RMSE 來得高（約高 0.08）
5 小時全部汙染源的 private RMSE 比只抽 PM2.5 的 RMSE 來得低（約低 0.35）
綜合 public 和 private 分數來看，抽 5 小時全部汙染源的結果較好。
另外值得注意的是，抽 5 小時全部汙染源在 private set 的結果要比抽 9 小時來得好。

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

Regularization	λ	0.0001	0.001	0.01	0.1
全部汙染源	public	7.46237	7.46236	7.46233	7.46198
	private	5.53562	5.53561	5.53553	5.53477
PM2.5	public	7.44013	7.44013	7.44013	7.44012
	private	5.62719	5.62719	5.62719	5.62720



答： 在本題指定的 λ 範圍中，regularization 的效果不大（如上圖）。

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

答： (c)

Proof:

Using matrix calculus:

$L(w,X,y) = \sum_{n=1}^{N}(\mathrm{y}^n - \mathrm{x}^n \cdot \mathrm{w})^2 = (y - X\mathrm{w})^T(y - X\mathrm{w})$

$\frac{\partial}{\partial w}L(w,x,y) = \frac{\partial}{\partial w}\left[(y - X\mathrm{w})^T(y - X\mathrm{w})\right] = -2(y - Xw)^T X = 0$

$(y - Xw)^T X = y^T X - (Xw)^T X = y^T X - w^T X^T X = 0$

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

$(X^T X)^T w = X^T y,\ (X^T X)^T = X^T X$

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