ML HW1 學號：R05H41011 系級： 統計碩二 姓名：薛仲廷

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

	Public	*Private*
全汙染源	7.53224	5.82931
PM2.5	7.43710	5.59743

由上表可知只取 PM2.5 的模型比較好，推斷可能是因為其他汙染源根所要預測的 PM2.5 沒什麼相關，造成 overfitting 的現象發生。

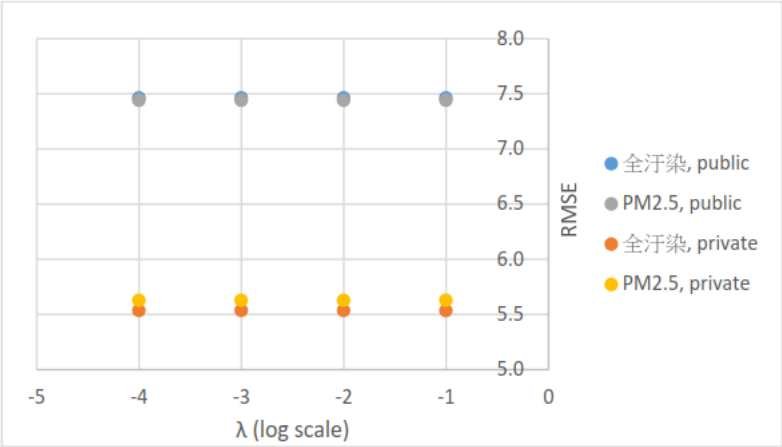
2. (1%)將 feature 從抽前 9 小時改成抽前 5 小時，討論其變化

		Public	*Private*
全汙染源	9hours:	7.53224	5.82931
	5hours:	7.23455	5.67242
PM2.5	9hours:	7.43710	5.59743
	5hours:	7.63411	5.67893

全汙染物時：9 小時的效果較佳，推斷為垃圾變數比較少產生的結果。

只取 PM2.5 時：5 小時的效果比較好，可能是單考慮一個變數時，取較多的資料對預測的結果也會比較好。

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



由上圖可知 regularization 在此範圍中效果並不明顯。

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 n=1Nyn-xnw2 。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)

(C)

y=X・w

X(transpose)・y = X(transpose)・X・w

Invertible(X(transpose)・X)・X(transpose)・y

= Invertible(X(transpose)・X)・X(transpose)・X・w

= w