學號：B04901080 系級： 電機三 姓名：戴靖軒

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	Both (RMS)
All feature	7.46237	5.53562	6.57001
PM 2.5	7.44013	5.62719	6.59624

Private 全部汙染源優於只取 PM2.5，Public 則相反，整體而言全取較佳。

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

	Public	Private	Both (RMS)
All feature	7.66119	5.44024	6.64417
PM 2.5	7.57651	5.79427	6.74452

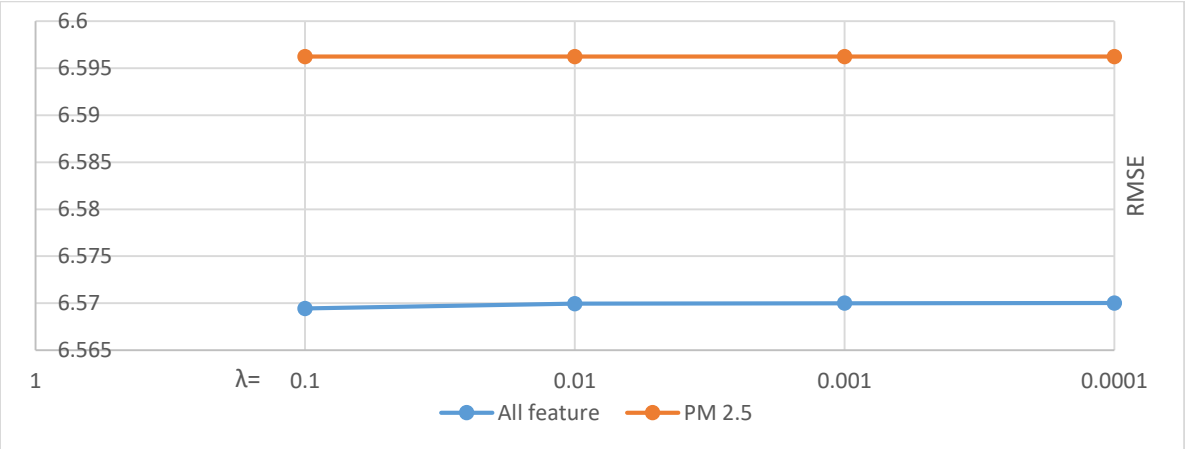
Private 全部汙染源優於只取 PM2.5，Public 則相反，整體而言全取較佳。

除了全取的 Private 變小，其餘誤差 5 小時相較於 9 小時皆變大。

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

All feature	Public	Private	Both (RMS)
λ=0.1	7.46198	5.53477	6.56943
λ=0.01	7.46233	5.53553	6.56995
λ=0.001	7.46236	5.53561	6.57000
λ=0.0001	7.46237	5.53562	6.57001
PM 2.5	Public	Private	Both (RMS)
λ=0.1	7.44012	5.62720	6.59624
λ=0.01	7.44013	5.62719	6.59624
λ=0.001	7.44013	5.62719	6.59624
λ=0.0001	7.44013	5.62719	6.59624

在此範圍 λ 影響不大(見下圖)。



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$\sum_{n=1}^{N}(y^n - x^n \cdot w)^2$。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)

(c) $(X^TX)^{-1}X^Ty$

$L(\mathrm{w}) = \|y - Xw\|^2 = (y - Xw)^T(y - Xw) = y^Ty - w^TX^Ty - y^TXw + w^TX^TXw$
$(w^TX^Ty)^T = y^TXw$ *為* 1×1 *的矩陣*
$L(\mathrm{w}) = y^Ty - 2w^TX^Ty + w^TX^TXw$
對w微分並令它為零
$-X^Ty + (X^TX)w = 0$
$w = (X^TX)^{-1}X^Ty$

Source:
https://en.wikipedia.org/wiki/Linear_least_squares_(mathematics)#Derivation_directly_in_terms_of_matrices