學號：r06942095 系級： 電信碩一 姓名：劉翔瑜

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) private+public=5.49085+7.81118=13.30

(2) private+public=5.62719+7.44013=13.06

可知只取 pm2.5 的值來 train 得的 model，其預測可能會更接近實際情況，推測是由於(1)中有許多的 feature 跟 pm2.5 的變化毫無關係，故將這些 feature 考慮進去得出的 model 誤差會加大

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

改成抽前 5 小時，則:

(1) private+public=5.37664+7.72036=13.097 (down

(2) private+public=5.7955+7.57119=13.366 (up

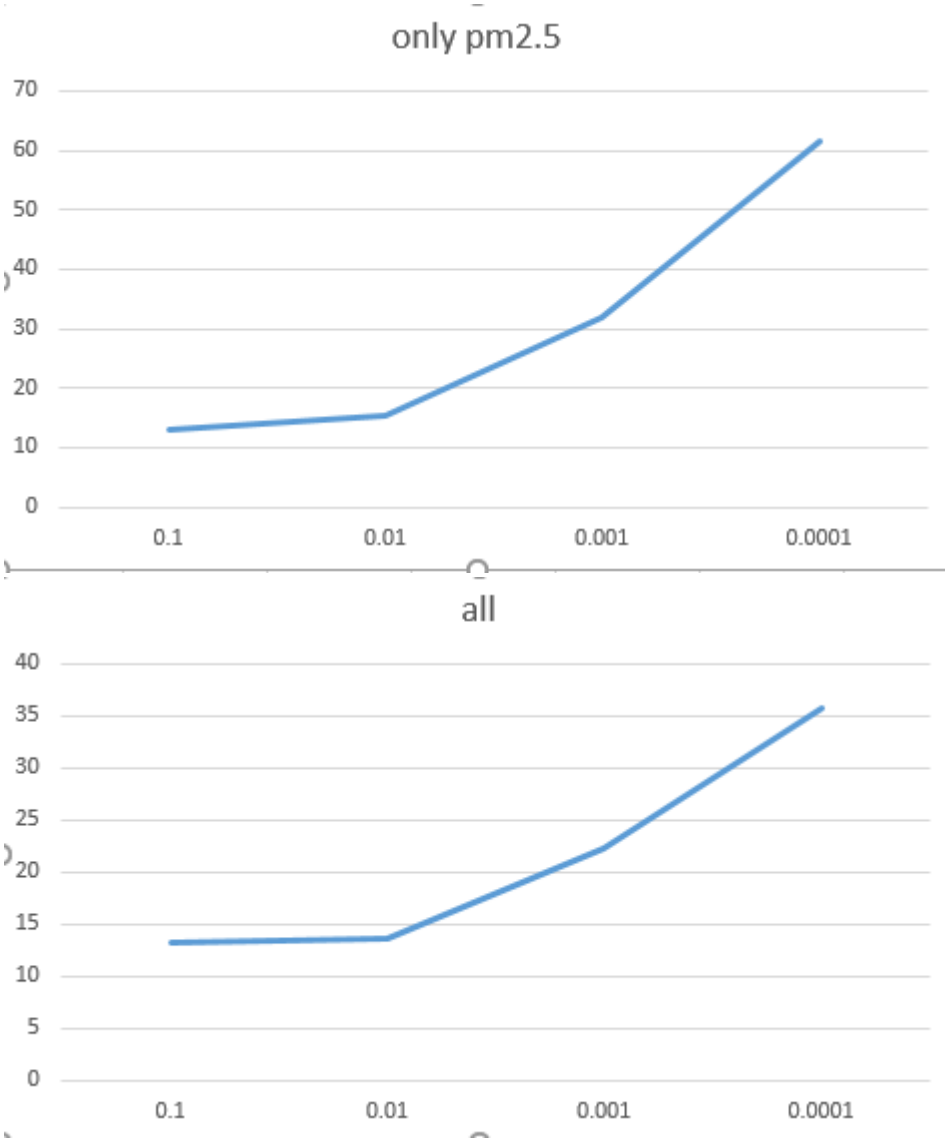
由(2)可推測出，前 9 個小時的 pm2.5 可能皆與第 10 小時的 pm2.5 有關係，故只取 5 小時，預測較不準

比較(1).(2)之表現，可知取 all 比只取 pm2.5 好，推測是未來之 pm2.5 可能不只與 pm2.5 有關係，也許 pm10 等等之值也有相關

若只取 5 小時，(1)之表現也變好了，猜測是若取 9 小時，考慮的變數太多，造成些許 overfitting 的情況

綜合以上結果，我們可以觀察(1),(2)中 model 對各個 feature 給的 weight，推測那些 feature 是我們所需要的，再利用這些 feature 來 train，也許可以得出更接近現實的預測 model

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



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

A: (c)