學號：B05502145 系級： 電機三 姓名：林禹丞

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 當作一次項(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

c. 第 1-3 題請都以題目給訂的兩種 model 來回答

d. 同學可以先把 model 訓練好，kaggle 死線之後便可以無限上傳。

e. 根據助教時間的公式表示，(1) 代表 p = 9x18+1 而(2) 代表 p = 9*1+1

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

Features	Train	Private	Public	Avg
All	5.354287	7.14983	5.6127	6.381265
Only PM2.5	7.33989	7.58096	6.21171	6.896335

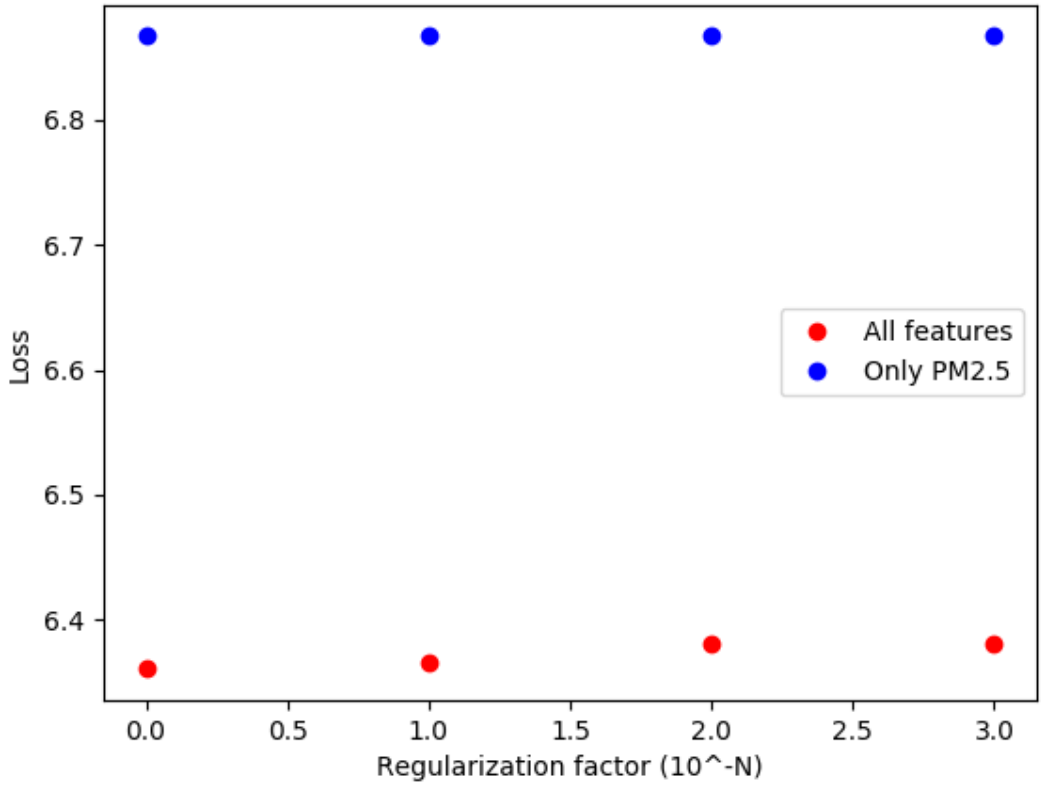
在(1)的模型下，我們有取 9 小時內的 18 種 features 再加上 bias，所以總共有 18*9+1=163 個參數。而在(2)的模型下只有 9 小時內的 PM2.5 加上 bias，總共有 9+1=10 個參數。由結果來看我們可以發現(2)在 Train 上的 Loss 有 7.33989，比(1)的 Train loss 大的許多，這是因為(2)的參數太少所以產生 underfitting 的結果。因此，自然的這樣的模型在 Test 上也不會有好的結果。

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

Features	Private	Public	Avg	Train	Hours
All	7.14983	5.6127	6.381265	5.354287	9
All	7.13936	5.89763	6.518495	5.49231	5
Only PM2.5	7.58096	6.21171	6.896335	7.33989	9
Only PM2.5	7.59987	6.36204	6.980955	7.347675	5

在 5 小時的情況下，(1)模型有 5*18+1=91 個參數，(2)模型有 5+1=6 個參數。雖然只取 5 小時 training set 的 data 數量會變多，可能可以把 model train 的更好，但是 data 增加的數量只有 2~3%，影響不大。因此，類似的，根據 1.的討論，我們可以發現參數太少也會有 underfitting 的結果。

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$\sum_{n=1}^{N} (y^n - w \cdot x^n)^2$ 。若將所有訓練資料的特徵值以矩陣 $X = [x^1\ x^2 \dots x^N]^T$ 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 ？請選出正確答案。(其中 X^TX 為 invertible)

(a) $(X^TX)X^Ty$
(b) $(X^TX)yX^T$
(c) $(X^TX)^{-1}X^Ty$
(d) $(X^TX)^{-1}yX^T$

答案：(c)