學號：B04902105 系級： 資工三 姓名：戴培倫

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

[all features] 8.37718 + 7.11265

[pm2.5] 6.85937 + 5.92366

因為很多feature對於pm2.5的數值沒有什麼相關性，像是CO、NMHC 、RAINFALL、WS_HR，加入許多這種feature反而導致loss變大。

2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

[all features] 9.59264 + 8.85819

[pm2.5] 7.20036 + 7.17349

從9小時變5小時相當與只剩5/9的feature，對預測有幫助的feature也相對減少許多，因此相比於9小時算出的loss都多了不少。

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

[all features] [λ = 0] (training data 7.48369275836)

[λ = 0.1] 8.37718 + 7.11265 (training data 7.48369274534)

[λ = 0.01] 8.37718 + 7.11265 (training data 7.48369275778)

[λ = 0.001] 8.37718 + 7.11265 (training data 7.48369275817)

[λ = 0.0001] 8.37718 + 7.11265 (training data 7.48369275828)

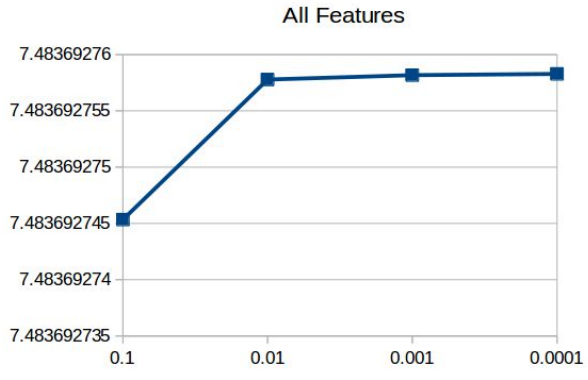
[pm2.5] [λ = 0] (training data 6.58098136257)

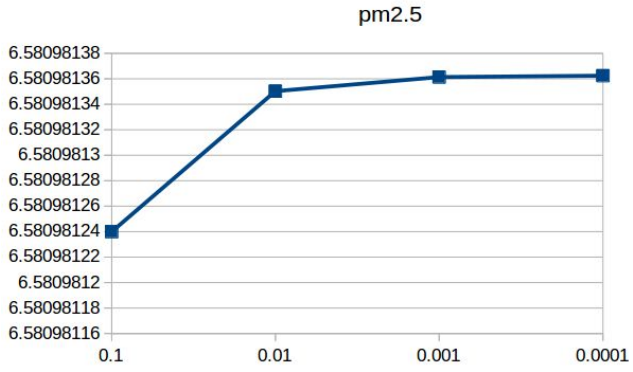
[λ = 0.1] 6.85937 + 5.92366 (training data 6.58098123994)

[λ = 0.01] 6.85937 + 5.92366 (training data 6.58098135031)

[λ = 0.001] 6.85937 + 5.92366 (training data 6.58098136135)

[λ = 0.0001] 6.85937 + 5.92366 (training data 6.58098136245)





在λ為0的情況下，我印出所有的weight，發現所有的weight的級數大多數皆為10^(-3) ~ 10^(-6)，加上λ為0.1, 0.01, 0.001，相乘後數值很小，因此在做regularization的時候沒有什麼差異性，training data的loss只差10^(-7)，而testing data則看不出差異。

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 $\sum(y^n - x^n \cdot w)^2$ 。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)

(a) $(X^TX)X^Ty$
(b) $(X^TX)^{-0}X^Ty$
(c) $(X^TX)^{-1}X^Ty$
(d) $(X^TX)^{-2}X^Ty$

$L(w) = \sum(y^n - x^n \bullet w)^2 = (Y - X \bullet W)^T (Y - X \bullet W)$

$d(L(w)) / dw = d(Y^TY - Y^TXW - W^TX^TY + W^TX^TXW) / dw$

$= 2X^TXW - 2X^TY = 0$

$\Rightarrow W = (X^TX)^{-1}X^TY$