學號：R06922152 系級： 資工碩一 姓名：袁晟峻

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 | Only pm2.5 |
| public | 24.67863 | 6.85360 |
| private | 22.51487 | 7.08239 |

明顯看出只用PM2.5的準確率較高，推測是因為並非每一種汙染源皆會對PM2.5產生影響，過多的features反而造成干擾，導致訓練出來的model準確率不高。

2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

|  |  |  |
| --- | --- | --- |
|  | All features | Only pm2.5 |
| public | 24.44748 | 6.86278 |
| private | 22.39961 | 7.10901 |

5小時的All features 結果比起9小時的準確率些微上升，可能可以歸因於少了許多干擾的feature，但在只有PM2.5的情況下準確率卻下降了，可能是減少了重要feature造成的結果。

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

改變λ對最後訓練的結果影響不大，推測可能是因為原本的model並沒有overfitting的情況，甚至可能有些underfitting，所以改變λ並不太影響訓練結果

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 xn，其標註(label)為一存量 yn，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 。若將所有訓練資料的特徵值以矩陣 X = [x1 x2 … xN]T 表示，所有訓練資料的標註以向量 y = [y1 y2 … yN]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。

1. (XTX)XTy
2. (XTX)-0XTy
3. (XTX)-1XTy
4. (XTX)-2XTy

