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 |
|---|---|---|
| RMSE | 23.62453 | 6.96893 |

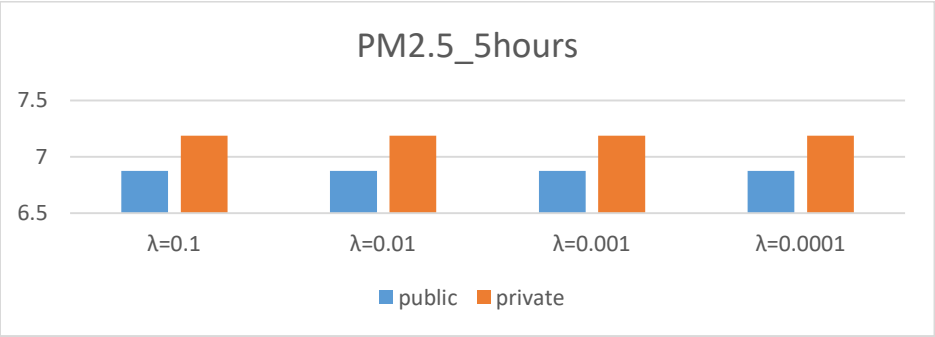
明顯看出只用 PM2.5 的準確率較高，推測是因為並非每一種汙染源皆會對 PM2.5 產生影響，過多的 features 反而造成干擾，導致訓練出來的 model 準確率不高。

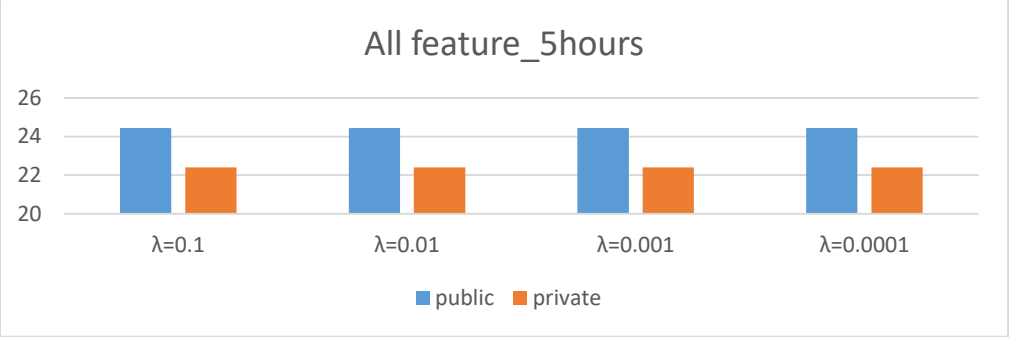
2. (1%)將 feature 從抽前 9 小時改成抽前 5 小時，討論其變化

| | All features | Only pm2.5 |
|---|---|---|
| RMSE | 23.44591 | 6.98697 |

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 $x^n$，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。

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