學號：R05546022 系級： 工工所碩二 姓名：謝立成

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□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□ 0，其他的數□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□ (RMSE)(根據 kaggle public+private 分數)，討論兩種 feature 的影響

9hr	All feature	only pm2.5
Private	7.4663	7.4401
public	5.3011	5.6272
RMSE	6.4748	6.5962

We can see that although only PM2.5 can train the public data lower, the private data is better when all feature is taken into account.

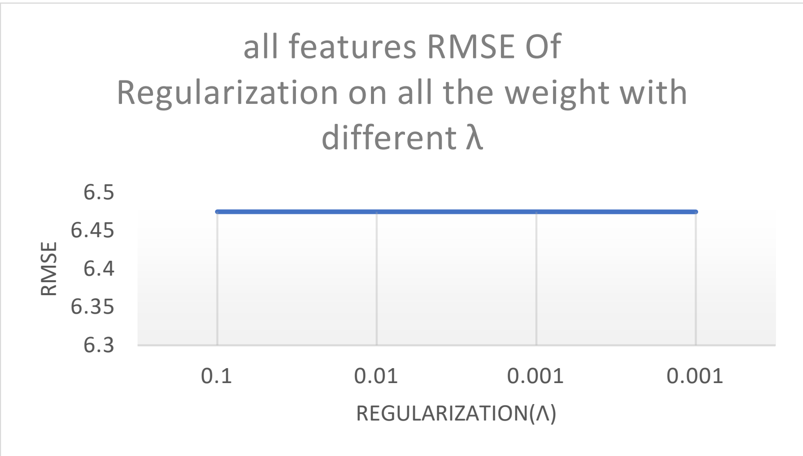
2. (1%)將 feature 從抽前 9 小時改成抽前 5 小時，討論其變化

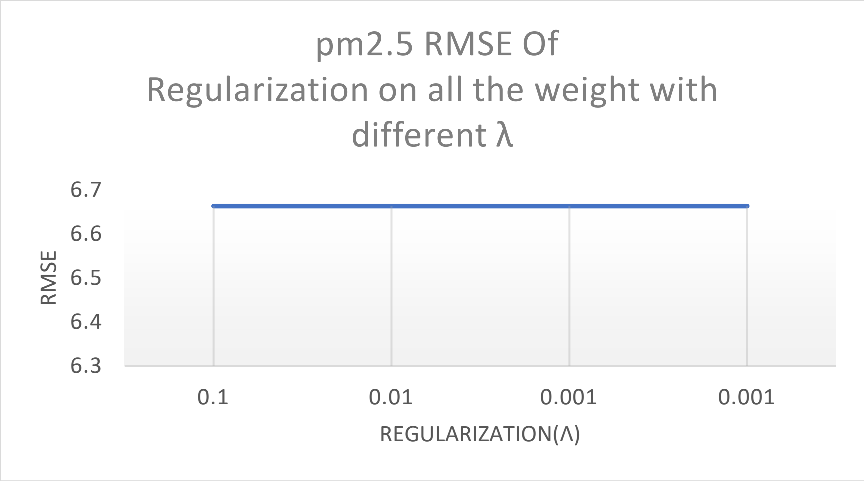
5hr	All feature	only pm2.5
Private	7.7375	7.579
public	5.3782	5.7919
RMSE	6.6631	6.7449

Same result as above, but is poorer than that of 9 hour training

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

It appears to be little difference when λ is different





	Regularization(λ)	*public*	*private*	*RMSE*
all features	0.1	7.4663	5.301	6.4748
	0.01	7.4663	5.301	6.4748
	0.001	7.4663	5.301	6.4748
	0.001	7.4663	5.3011	6.4748
only pm2.5	0.1	7.7375	5.3782	6.6631
	0.01	7.7375	5.3782	6.6631
	0.001	7.7375	5.3782	6.6631
	0.001	7.7375	5.3782	6.6631

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)□一向量 x^n，其標註(label)□一存量 y^n，模型參數□一向量 w (此處忽略偏權□ b)，則線性回歸的損失函數(loss function)□$\sum_{□=1}^{□}(□^{□} - □^{□} \cdot □)^2$ 。若將所有訓練資料的特徵□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 □ invertible)

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

ANS : (c)

$$\mathrm{L} = \sum_{1}^{\mathrm{N}} \left(y - w^{\mathrm{T}} x\right)^{\mathrm{T}} \left(y - w^{\mathrm{T}} x\right)$$

Minimize loss：

$$\frac{d\mathrm{L}}{d\mathrm{W}} = 2 \times (x^{\mathrm{T}} y - x^{\mathrm{T}} x w) = 0$$

$$\rightarrow \hat{w} = \left(x^{\mathrm{T}} x\right)^{-1} x^{\mathrm{T}} y$$