學號：B04701232 系級： 工管三 姓名：陳柔安

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

	only PM2.5	all feature
Testing RMSE	6.5962445333	6.49556276496

用所有的污染源作為 feature 的誤差值較單純只用 PM2.5 的誤差值低，兩者大約差 0.1，可以從中推測預測 PM2.5 時與前 9 小時的 PM2.5 值十分有關聯，但若同時也抽取其他的 feature 來輔助，能夠有更好的預測結果。

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

	only PM2.5	all feature
Testing RMSE（9hr）	6.5962445333	6.49556276496
Testing RMSE（5hr）	6.74452277777	6.60684664643

只抽取前五小時的預測結果較抽取前九小時的預測結果差，無論是第一或第二個 model，但雖然少了將近一半的 feature 仍能預估出誤差值相近的結果，顯示其實主要影響這個 model 的是前五小時的數據，再往前推四小時可能相關性沒那麼大，只是做些微的調整。

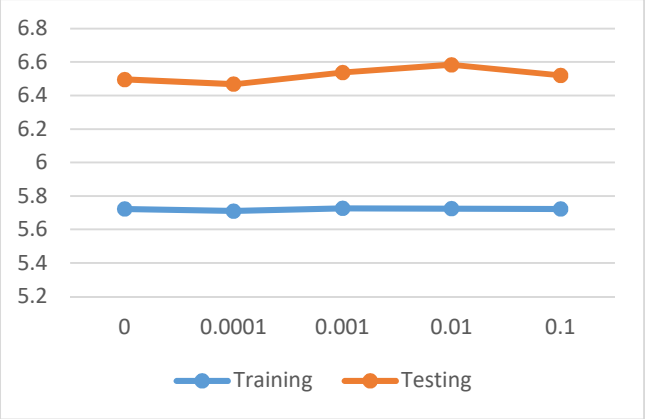
3. (1%)Regularization on all the weight with λ=0.1、0.01、0.001、0.0001，並作圖

only PM2.5

	λ=0	λ=0.0001	λ=0.001	λ=0.01	λ=0.1
Training	6.1230215221	6.1230215292	6.12302159321	6.1230222333	6.12302863401
Testing	6.5962445333	6.59624453279	6.59624452824	6.59624448265	6.59624402717

all feature

	λ=0	λ=0.0001	λ=0.001	λ=0.01	λ=0.1
Training	5.72248673445	5.71030948349	5.72637500644	5.72432516496	5.72258413769
Testing	6.49556276496	6.46825910985	6.53813361448	6.58391777488	6.52054012777



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$\sum_{n=1}^{N}(\hat{y}^n - w \cdot x^n)^2$。若將所有訓練資料的特徵值以矩陣 $X = [x^1\ x^2 \ldots x^N]^T$ 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 為 invertible)

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