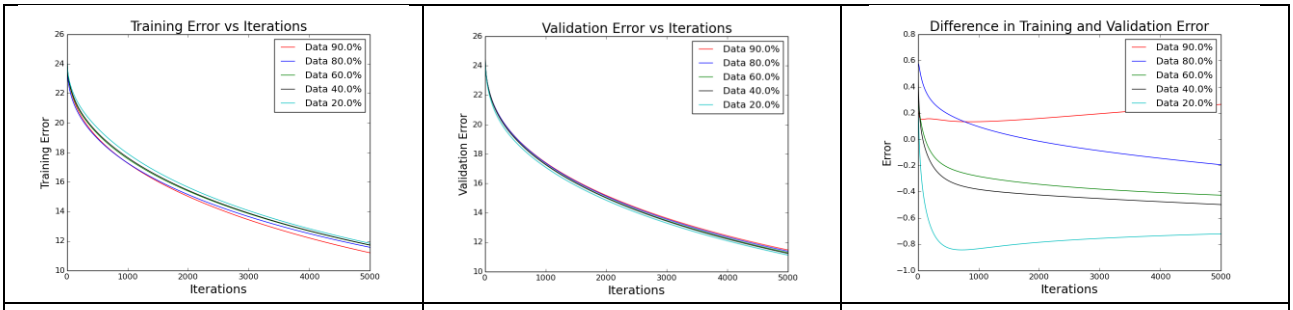
學號：R05921086 系級： 電機四 姓名：邱名彥

All errors are RMSE

1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)

We use the first 9 hours and features PM10, PM2.5, Rainfall, Wind_Direct, and Wind_Speed. Choosing the entire set of 9 hours should allow the model to accurately predict the 10th hour. We choose these 5 features according to the Kaggle Score. We test our model by adding in a single feature at a time. We keep the feature in our model if the kaggle score improves and we remove the feature if our score decreases to get a final result of 5.72.

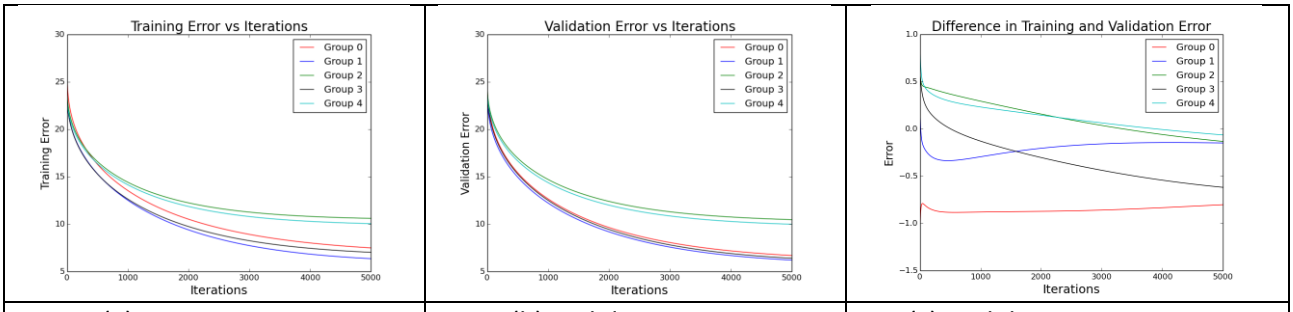
2.請作圖比較不同訓練資料量對於 PM2.5 預測準確率的影響



(a) Training Set Losses (b) Validation Set Losses (c) Validation – Training Loss

We create 5 models and reduce the size of the training set in each training cycle. We use train our model using 90%, 80%, 60%, 40%, and 20% of the total training data. We use the unused data as extra data for the validation set. It can be observed that with larger training data, the training error decreases. The validation error still seems to be relatively unaffected. This shows that having more data is better to predict more accurate PM2.5 values.

3. 請比較不同複雜度的模型對於 PM2.5 預測準確率的影響

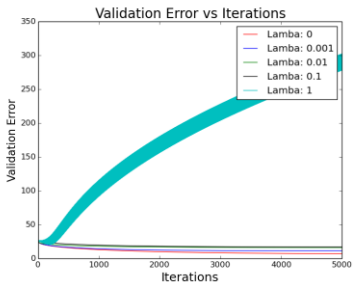
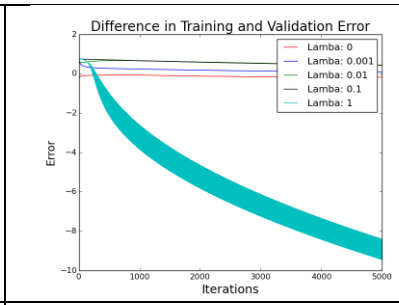


(a) Training Set Losses (b) Validation Set Losses (c) Validation – Training Loss

We create 5 test groups and run our training model to get the above graphs. We can see that using different features can significantly affect our PM2.5 accuracy prediction. Therefore it is important to select features are related to PM2.5 and remove features that are not correlated in order to have better accuracy

Group #	Selected Features
0	PM2.5
1	All Features
2	PM2.5, Rainfall, SO2, Wind_Speed
3	Amb_Temp, PM10, PM2.5, Rainfall, Wind_Direct, Wind_Speed
4	Amb_Temp, NO, NO2, NOx, O3, PM10, PM2.5, Rainfall, SO2, WD_HR, Wind_Direct, Wind_Speed, WS_HR

4. 請討論正規化(regularization)對於 PM2.5 預測準確率的影響

		
(a) Training Set Losses	(b) Validation Set Loses	(c) Validation – Training Loss

Regularization is a technique to attempt to solve the problem of overfitting by using the equations shown below

$$L(X,Y)+\lambda L(X,Y)$$

The λ term adjusts model complexity so that the trained PM2.5 model can accurately predict values by smoothing the magnitude of our loss function. Smoothing the loss function

We run 5 models with varying lambas on Group 3 from Q3 and get the above graph. We show that adding regularization may or may not improve our accuracy.

5. 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 n=1Nyn-wxn2 。若將所有訓練資料的特徵值以矩陣 X = [x^1 x^2 … x^N] 表示，所有訓練資料的標註以向量 y = [y^1 y^2 … y^N]^{T}表示，請以 X 和 y 表示可以最小化損失函數的向量 w 。

答：

loss function:$\sum_{n=1}^{N}(y^n - w * x^n)^2$

Let $\sum_{n=1}^{N}(y^n - w * x^n)^2 = 0$

$=\sum_{n=1}^{N} y^n = \sum_{n=1}^{N} w * x^n$

$$=\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = w * \begin{bmatrix} (x^1)^T \\ \vdots \\ (x^n)^T \end{bmatrix}$$

Let $A = \begin{bmatrix} (x^1)^T \\ \vdots \\ (x^n)^T \end{bmatrix}, B = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, X = w$

$AX = B$ if A is not a square matrix

Then we multiply A^Tin to get $A^T AX = A^T B$ and multiply $(A^T A)^{-1}$to get $X = (A^T A)^{-1} A^T B$

Type equation here.

We get $W = \begin{bmatrix} [x^1 & \dots & x^N] * \begin{bmatrix} (x^1)^T \\ \vdots \\ (x^{nT} \end{bmatrix} \end{bmatrix} * [x^1 \quad \dots \quad x^N] * \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = [X * X^T]^{-1} XY$