學號：R05921069 系級： 電機碩一 姓名：黃武昱

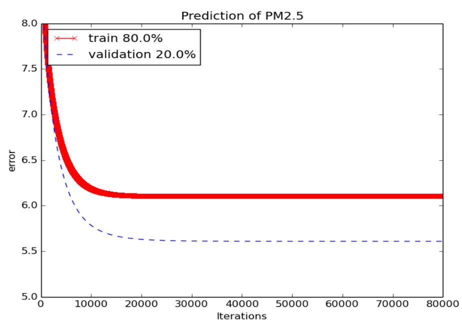
1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)

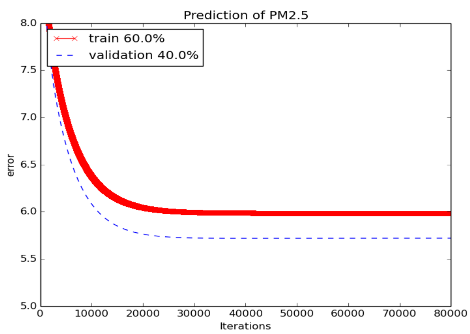
答：

我先是把一天的資料，18 個空氣汙染指標，0 到 23 小時為一組，之後再把剩下的 239 天依序排再第 1 天的後面，這樣會得到一(18,5760)大小的矩陣.

再依序把 18 個空氣汙染指標的前 9 小時與第 10 小時為 pm2.5 的第 10 小時為一筆資料，得到一個(1,163)的矩陣，之後再連續排列這 18 個空氣汙染指標的前 10 個小時為一組資料，最後會得到一個(5652,163)的矩陣，這個矩陣就是我們的輸入資料.

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





答： 由上圖可知，如果 train data 取 80%的數量，跟取 60%的數量來比較，這兩個的 error 相對是很接近的，

但是，在 validation 取 80%數量的 error 會比取 60%數量的 error 來的小很多

3. 請比較不同複雜度的模型對於 PM2.5 預測準確率的影響

答：

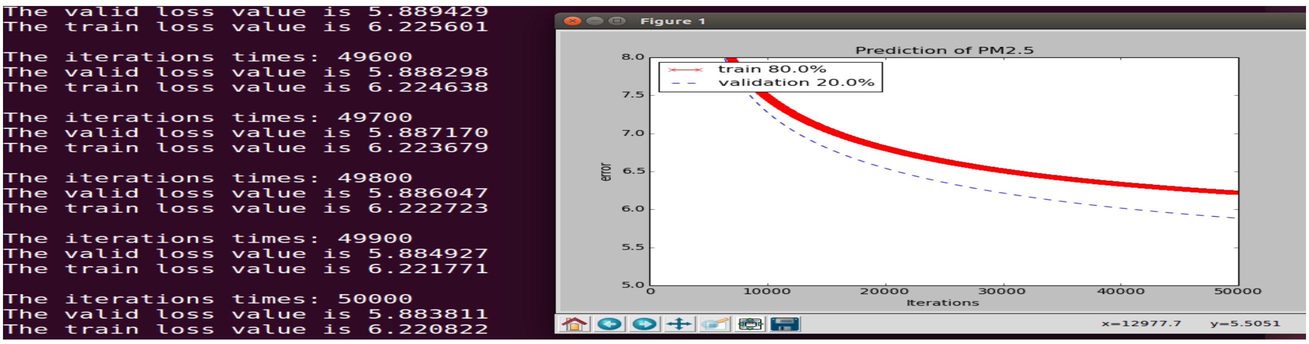


圖 1

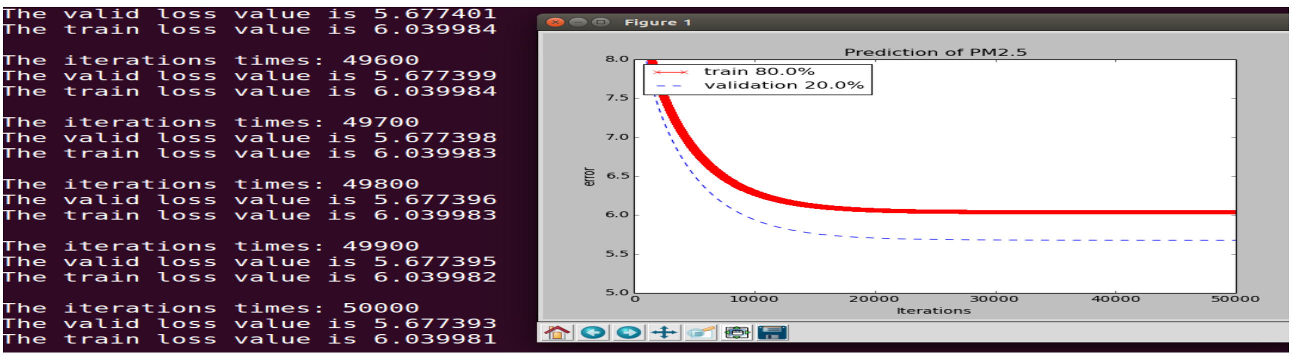


圖 2

抽取的 feature 越多，模型越複雜，那麼其特性是越接近非線性的曲線。

在 learning regression 中，模型越複雜，準確率也會愈來愈低.

例如：圖 1 中，我抽取 18 個空氣汙染指標的前 9 小時來預測資料跟圖 2 只抽取 pm2.5 那筆空氣汙染指標的 1~9 小時來預測資料；相比之下，只抽取 pm2.5 那筆空氣汙染指標的 1~9 小時的 PM2.5 預測準確率會高於抽取 18 個空氣汙染指標的前 9 小時，因為抽取 18 個空氣汙染指標的前 9 小時的模型複雜度遠高於只抽取 pm2.5 那筆空氣汙染指標的 1~9 小時的模型.

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

答：

Regularization 即是調整 learning rate，可以減低 overfiting 的發生。

如果沒有 regularization，在 overfiting 的情況下，那麼在 training 過程中，training error 會變得小，但是 validation error 卻會增大，即對 PM2.5 的預測準確率會變差。

但是，如果加入 regularization 後，雖然能降低 overfiting 的發生，與讓 training error 和 validation error 的數值相近，卻會讓 error 的數值上升.

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 $X = [x_1\ x_2\ \ldots\ x_N]$ 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 : $\sum_{n=1}^{N}(y^n - w\cdot x^n)^2$

令 $\sum_{n=1}^{N}(y^n - wx^n)^2 = 0$

$\Rightarrow \sum_{n=1}^{N} y^n = \sum_{n=1}^{N} wx^n$

$$\Rightarrow \begin{bmatrix} y_1\\ y_2\\ \vdots\\ y_N\end{bmatrix} = w\cdot\begin{bmatrix}(x^1)^T\\ (x^2)^T\\ \vdots\\ (x^N)^T\end{bmatrix}$$

令 $B = \begin{bmatrix} y_1\\ y_2\\ y_N\end{bmatrix}$, $A = \begin{bmatrix}(x^1)^T\\ (x^2)^T\\ (x^N)^T\end{bmatrix}$, $x = w$

$\Rightarrow Ax = B$ ，若 A 為非方陣

$\Rightarrow$ 同乘 A^T , $A^TAx = A^TB$

$\Rightarrow$ 同乘 $(A^TA)^{-1}$, 得 $x = (A^TA)^{-1}A^TB$

$$\text{故得 } w = \left([x^1\ x^2 \cdots x^N]\begin{bmatrix}(x^1)^T\\ (x^2)^T\\ \vdots\\ (x^N)^T\end{bmatrix}\right)^{-1}\cdot[x^1\ x^2 \cdots x^N]\begin{bmatrix} y_1\\ y_2\\ \vdots\\ y_N\end{bmatrix} = (x\cdot x^T)^{-1}xy$$