學號：R05944019 系級：網媒所碩一 姓名：張嘉豪

1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)

答：(1) 取連續 9 小時的 PM2.5

train_x = [x1, x2, x3, x4, x5, x6, x7, x8, x9]

(2) 取連續 9 小時的 PM2.5 做一維以及二維的 feature

train_x = [x1, x1^2, x2, x2^2, x3, x3^2, … , x9, x9^2]

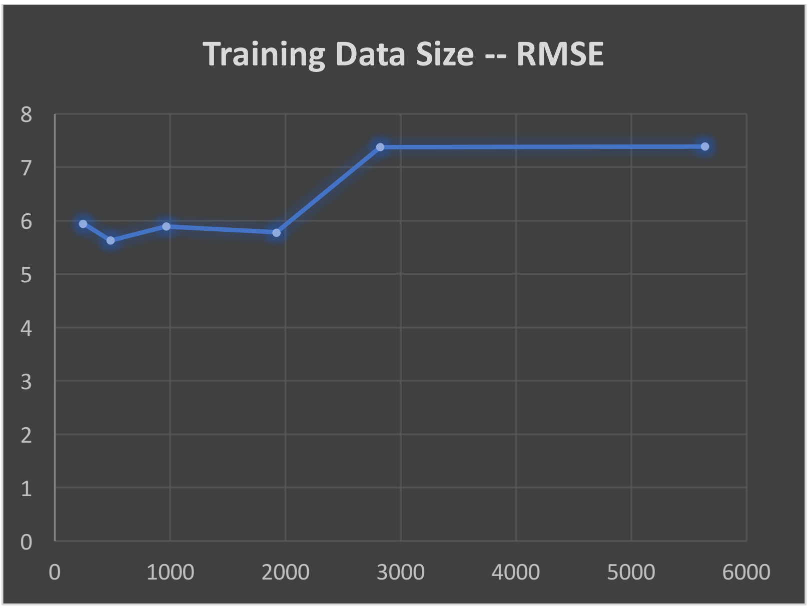
(3) 取連續 9 小時的所有空氣汙染指標 (18 種)

train_x = [x1, x2, x3, … , x161, x162]

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：

當 training data 為 480 的時候 RMSE 為最小值，隨著 training data 數量的上升 RMSE 會跟上升。



Training Data	RMSE
240	5.94909
480	5.62657
960	5.89209
1920	5.77999
2820	7.37733
5640	7.38749

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：

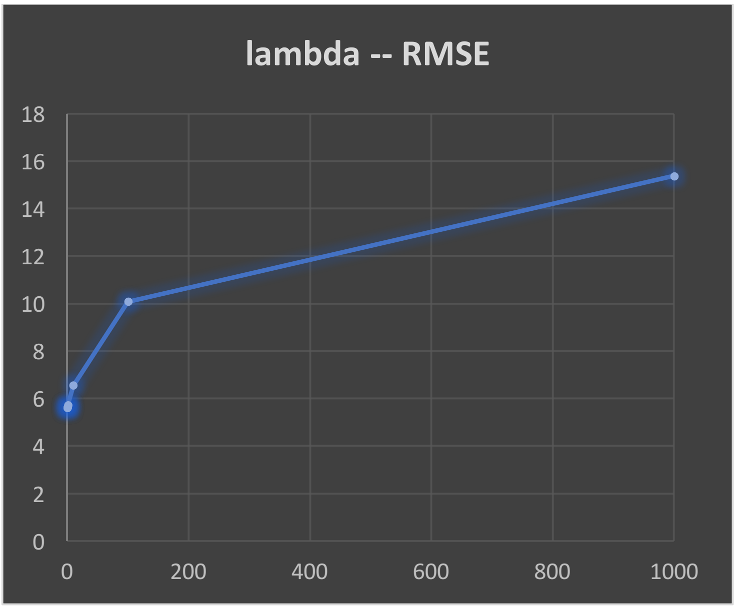
用一維加上二維的 PM2.5 當作 feature 效果並沒有比較好，若將全部的空氣污染指標當作 feature 的話會 overfitting，可以看得出來 RMSE 急速上升。

模型	RMSE
PM2.5 一維	5.62657
PM2.5 一維＋二維	5.75601
所有空氣污染指標	7.66866

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

答：

lambda=0.1 的時候達到最低點，接著隨著 lambda 越來越大，其預測結果的 RMSE 也越來越大。



lambda	RMSE
0	5.62657
0.01	5.62657
0.1	5.61768
1	5.71766
10	6.54535
100	10.07307
1000	15.36961

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 $\mathbf{x}^n$，其標註(label)為一存量 $\mathbf{y}^n$，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$\sum_{n=1}^{N}(y^n - w \cdot x^n)^2$ 。若將所有訓練資料的特徵值以矩陣 $\mathbf{X} = [\mathbf{x}^1\ x^2 \dots x^N]$ 表示，所有訓練資料的標註以向量 $\mathbf{y} = [\mathbf{y}^1\ y^2 \dots y^N]^T$ 表示，請以 **X** 和 **y** 表示可以最小化損失函數的向量 **w** 。

答：

$$\frac{\partial L}{\partial w} = 2 \cdot (\mathrm{y} - \mathrm{w} \cdot \mathrm{x}) \cdot \mathrm{x} = 0 \quad \rightarrow \quad x^T \cdot (\mathrm{y} - \mathrm{w} \cdot \mathrm{x}) = 0 \quad \rightarrow \quad x^T \cdot y - x^T \cdot x \cdot w = 0$$

$$\rightarrow \quad x^T \cdot y = x^T \cdot x \cdot w \quad \rightarrow \quad \mathrm{w} = (x^T \cdot x)^{-1} \cdot x^T \cdot y$$