

1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)

答：

我在一開始所使用的 feature 為前九天的所有 value，並且使用基本的 linear model 如 Equation1 所示。然而 error 一直卡在 6.2 左右無法下降，後來改只採用前九天的 PM2.5 作為 feature 後，錯誤率下降至 5.8。之後將模型改成二次方後錯誤率又下降到 5.7 左右，因此在最後作業繳交上所採用的 feature 為前九天之 PM2.5 value，並且採用二次方模型訓練，如 Equation2。由此結果可得知其部分參數對於預測結果會造成 noises。

$$y = b + wx \quad \text{Equation 1}$$

$$y = b + wx + ax^2 \quad \text{Equation 2}$$

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：

我在同一個模型以及參數的情況下測試了三種資料量，每天分別取樣 2/4/7/13 筆資料，總 training data 分別為 480/960/3120，雖然說 training data 越多筆照理來說 variance 以及 error rate 會越來越小，然而就結果來觀察反而是在 480 筆以及 1680 筆 training data 時表現較好，如 Figure 1 以及 Table 1 所示。

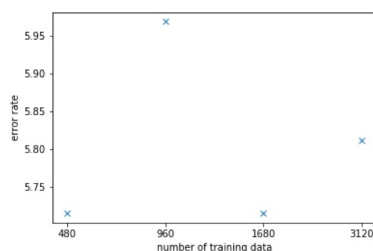


Figure 1

Training data 數量	Error rate
480	5.71475
960	5.96937
1680	5.71475
3120	5.81091

Table 1

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：

我實際上傳了在採用相同參數的兩種模型，分別為一次方以及二次方，如 Equation1,2 所示。由結果可得知在二次方的模型中所得到的結果相較於一次方而言是較為出色的，一次方的 error rate 再嘗試了許多次後得到的結果最終卡在

error rate = 5.8 左右，然而二次方的模型 error rate 可下降至 5.7 左右，如 Table2 所示。

模型	Error rate
$y = b + wx$	5.81664
$y = b + wx + ax^2$	5.71475

Table 2

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

答：

在本次作業中一樣採用相同的參數上傳了四種 regularization，分別為：0/0.1/10/1000 發現其結果設定在 0.1 以及 10 是最好的，當設定到 1000 可發現因為 function 過於平滑造成錯誤率反而上升了，如 Table3 所示。

λ	Error rate
0	5.71475
0.1	5.71475
10	5.71475
1000	5.76339

Table 3

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

答：

$$w = (X^T X)^{-1} X^T y$$