

1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)

一開始我是照著和PM2.5相關係數高的feature選下來，後來發現相關係數太高的就等於選了兩個PM2.5，我猜測這樣對於準確度沒有幫助，因此我最後選擇了前9個小時的PM2.5, PM10和相關係數在中間的一些化學物質共118個feature。

(PM2.5, PM2.5 ^ 2, PM2.5 ^ 3, PM10, PM10 ^ 2, PM10 ^ 3, CO, O3, SO2, WD_HR, WIND_DIREC, WIND_SPEED, WS_HR)

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

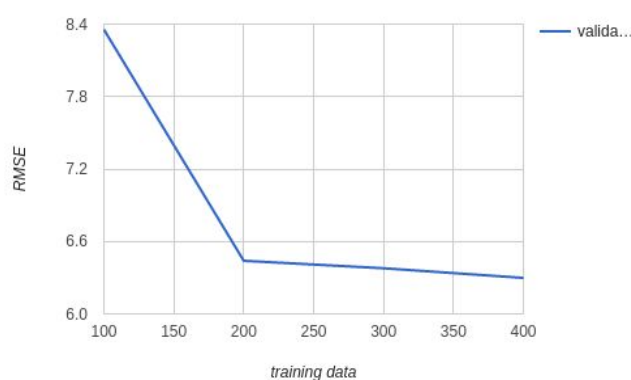
每個月的的前100, 200, 300, 400 hour作為

training data

每個月的後80hour作為validation data

Method: data adagrad, 10000 epoch

Feature: PM2.5, PM2.5 ^ 2, PM2.5 ^ 3, PM10, PM10 ^ 2, PM10 ^ 3, CO, O3, SO2, WD_HR, WIND_DIREC, WIND_SPEED, WS_HR (total: 118)



❖ training data增加有助於準確度的提升

3. 請比較不同複雜度的模型對於PM2.5預測準確率的影響

每個月的的前400 hour作為training data

每個月的後80hour作為validation data

Method: 最小平方法

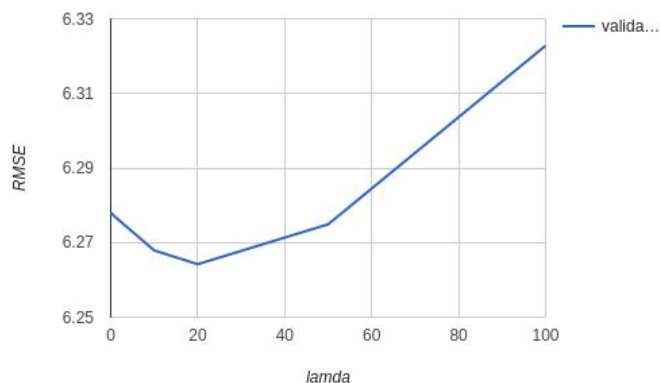
Feature: PM10 & PM2.5 的 1~X次方

X	1	2	3	4	5	6
training error	5.9126	5.8862	5.8329	5.8081	5.7659	5.7456
testing error	6.4400	6.4259	6.5151	6.5774	6.8033	6.8805

❖ 愈複雜的model對於training data的準確度愈高，但testing data可能會因為選過多不重要的feature倒致overfitting，因此要選一個適當的複雜度

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

每個月的前400 hour作為training data
每個月的後80hour作為validation data
Method: adagrad, 10000 epoch
Feature: PM2.5, PM2.5 ^ 2, PM2.5 ^ 3,
, PM10 ^ 2, PM10 ^ 3, CO, O3, SO2,
WD_HR, WIND_DIREC,
WIND_SPEED, WS_HR (total: 118)



❖ 適當的lamda可以增加與測的準確度，但是過高或太低會有反效果，Loss function PM10中的兩項要取得一個平衡

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 $X = [x^1 \ x^2 \ \dots \ x^N]$ 表示，所有訓練資料的標註以向量 $y = [y^1 \ y^2 \ \dots \ y^N]^T$ 表示，請以 X 和 y 表示可以最小化損失函數的向量 w 。

$$\begin{aligned} \text{let } Y - Xw &= 0 \\ Y &= Xw \\ X^T Y &= X^T X w \\ \Rightarrow w &= (X^T X)^{-1} X^T Y \end{aligned}$$