

1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)

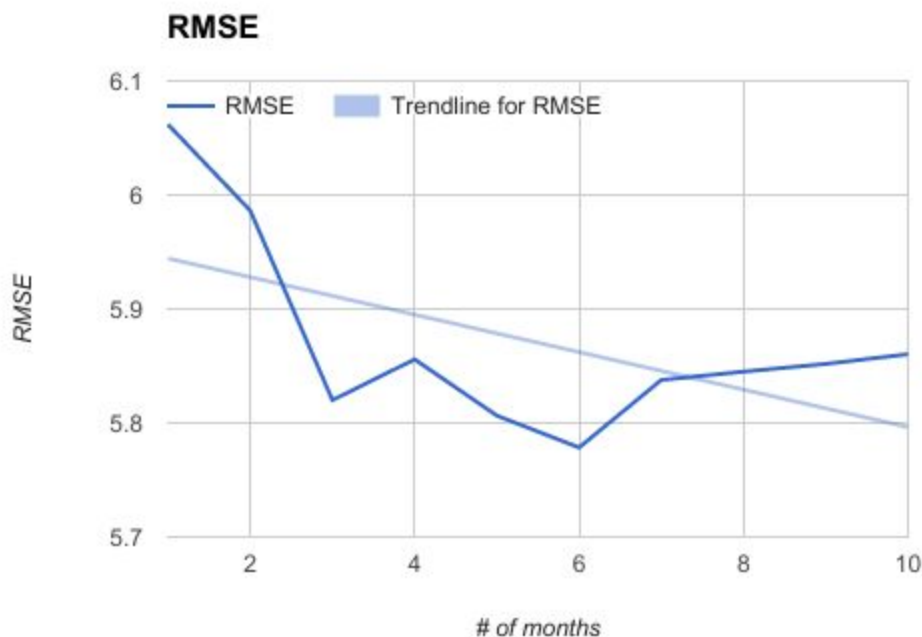
答：

- 1) Google 可能對 PM2.5 影響較大的因素
- 2) 透過 trial and error 嘗試各種 feature 組合
- 3) 使用 feature scaling 來加速 gradient descent
- 4) 從 training data 中切出 validation set 評估每種 feature 組合的表現 (但跟 kaggle 的結果不一定成正相關)

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：

- 實作 Adagrad
- 以 PM2.5 第一個小時到第九個小時的一次方為 feature
- 使用最後兩個月為 validation set, 剩下的調整不同月數量做 training



# of months	1	2	3	4
RMSE	6.06187	5.98636	5.82006	5.85581
# of months	5	6	7	8
RMSE	5.80644	5.77845	5.83774	5.84504
# of months	9	10		
RMSE	5.85181	5.86034		

由圖可清楚看出 training 使用的資料量越多，趨勢上很明顯是越準確的。

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：

Features	RMSE
PM2.5-9 ¹	6.59418
PM2.5-1 ¹ ~ PM2.5-9 ¹	5.86034
PM2.5-1 ¹ ~ PM2.5-9 ¹ PM2.5-1 ² ~ PM2.5-9 ²	5.89758
PM2.5-1 ¹ ~ PM2.5-9 ¹ PM2.5-1 ² ~ PM2.5-9 ² PM2.5-1 ³ ~ PM2.5-9 ³	5.93192
PM2.5-1 ¹ ~ PM2.5-9 ¹ PM10-1 ¹ ~ PM10-9 ¹	5.78952
PM2.5-1 ¹ ~ PM2.5-9 ¹ PM10-1 ¹ ~ PM10-9 ¹ RAINFALL-1 ¹ ~ RAINFALL-9 ¹	5.79466
PM2.5-1 ¹ ~ PM2.5-9 ¹ PM10-1 ¹ ~ PM10-9 ¹ RAINFALL-1 ¹ ~ RAINFALL-9 ¹ O3-1 ¹ ~ O3-9 ¹	5.75993

PM2.5-1 ¹ ~ PM2.5-9 ¹ PM10-1 ¹ ~ PM10-9 ¹ RAINFALL-1 ¹ ~ RAINFALL-9 ¹ O3-1 ¹ ~ O3-9 ¹ WD_HR-1 ¹ ~ WD_HR-9 ¹ WIND_SPEED-1 ¹ ~ WIND_SPEED-9 ¹ WS_HR-1 ¹ ~ WS_HR-9 ¹	5.78996
ALL-1 ¹ ~ ALL-9 ¹	5.72339

P.S. 每個 feature 以 "指標名 - 第幾小時 ^ 次方" 來表示

實驗各種組合後發現：

- feature 越多結果不一定越好
- feature 的次方越高結果幾乎會變差

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

答：

λ	Training RMSE	Validating RMSE
1	6.18341	5.84922
10	6.37210	5.89194
100	9.78734	8.31825
1000	13.89033	11.61689
10000	14.65590	12.24565

實驗後發現在這裡 regularization 的作用不是很大，不管 λ 怎麼調整。

推測原因可能是因為我用的 feature 都是一次式，原本曲線就已經蠻 smooth 了，因此 regularization 原本的初衷在這裡的影響便很小。

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 。

答：

我們知道最佳解會出現在微分等於 0 的位置，也就是：

$$(y - X \cdot w)^T \cdot X = 0$$

因此，我們可以推出：

$$\begin{aligned} X^T X w &= X^T y \\ w &= (X^T X)^{-1} X^T y \end{aligned}$$