

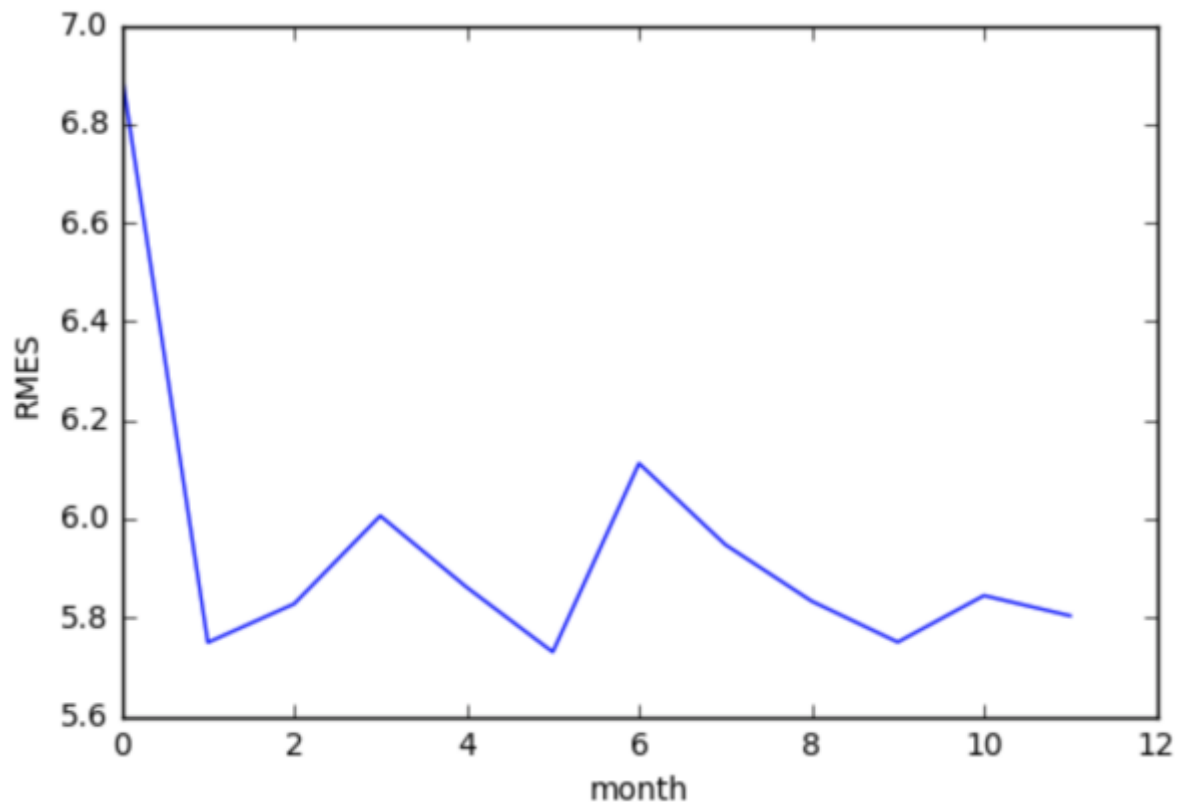
1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)

答：

抓每九小時的 data，NR 改成 0，轉成一個 (1, 162) 的矩陣，最後在前面加上一個 1 作為 bias 項，最終為一個 (1, 163) 的矩陣。

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：



只取一個月的資料取 feature 時，error 最大，之後 error 的值與資料的多寡沒有明顯關係，learning rate 與 iteration 次數都用一樣的數字。

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：

在實驗中比較 loss function $= L = \sum_{n=1}^N (y^n - w \cdot x^n)^2$ 及 $L_1 = \sum_{n=1}^N (y^n - w \cdot x^n - w_2 \cdot x^{2n})^2$ ， L_1 的 RMSE 為 6.88156521329， L 的 RMSE 為 5.8035002089，在這次的實驗中較高的複雜度沒有出現較好的結果。

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

答：

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 y^n ，模型參數為一向量 \mathbf{w} (此處忽略偏權值 b)，則線性回歸的損失函數(loss function)為 $\sum_{n=1}^N (\hat{y}^n - \mathbf{x}^n \cdot \mathbf{w})^2$ 。若將所有訓練資料的特徵值以矩陣 $\mathbf{X} = [\mathbf{x}^1 \mathbf{x}^2 \dots \mathbf{x}^N]$ 表示，所有訓練資料的標註以向量 $\mathbf{y} = [y^1 y^2 \dots y^N]^T$ 表示，請以 \mathbf{X} 和 \mathbf{y} 表示可以最小化損失函數的向量 \mathbf{w} 。

答：

$$\text{loss function} = L = \sum_{n=1}^N (y^n - \mathbf{w} \cdot \mathbf{x}^n)^2 = \sum_{n=1}^N (y^n)^2 + \sum_{n=1}^N 2\mathbf{w} \cdot \mathbf{x}^n y^n + \sum_{n=1}^N \mathbf{w}^2 (\mathbf{x}^n)^2$$

$$\frac{\partial L}{\partial \mathbf{w}} = 0 = 2 \sum_{n=1}^N \mathbf{x}^n y^n + 2\mathbf{w} \sum_{n=1}^N (\mathbf{x}^n)^2$$

$$\text{for min}(\mathbf{w}) \Rightarrow \frac{\partial L}{\partial \mathbf{w}} = 0$$

$$2 \sum_{n=1}^N \mathbf{x}^n y^n = -2\mathbf{w} \sum_{n=1}^N (\mathbf{x}^n)^2$$

$$\mathbf{X} \cdot \mathbf{Y} = \mathbf{w}(\mathbf{X}^T \cdot \mathbf{X})$$

$$\mathbf{w} = \mathbf{X} \cdot \mathbf{Y}(\mathbf{X}^T \cdot \mathbf{X})$$