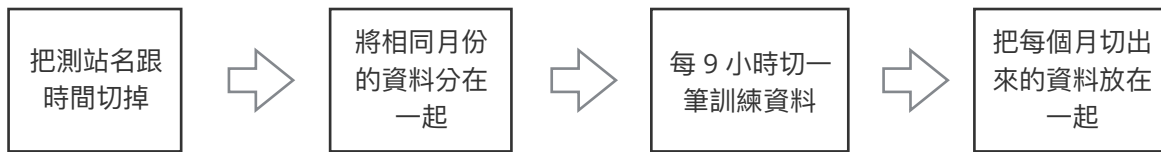


學號：B03901039 系級：電機三 姓名：童寬

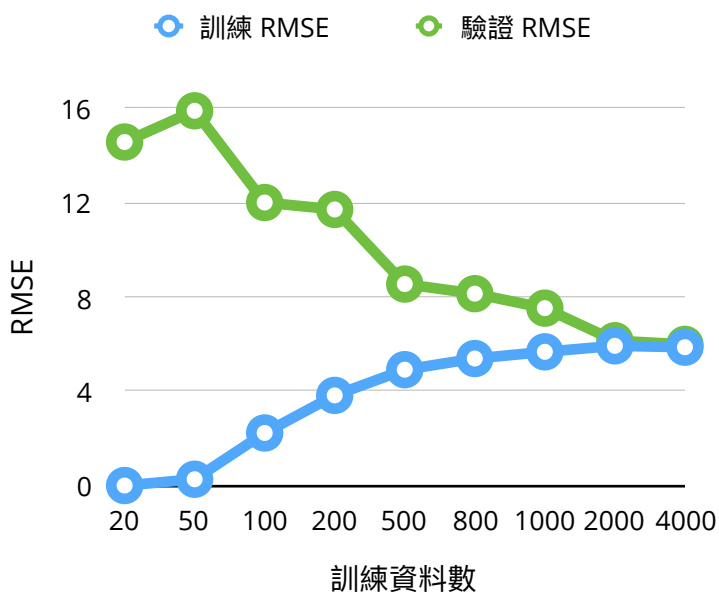
1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)

答：



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：



實驗設計：

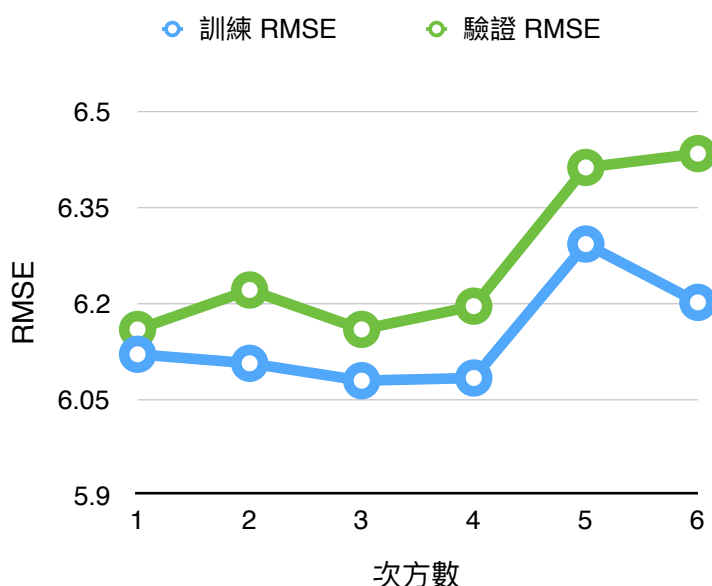
- 取 CO, O3, PM2.5, WIND_DIREC, WIND_SPEED 為特徵
- 驗證資料量是訓練資料量的 20%
- 有用 stochastic 的方法，batch 數量是訓練資料量的 2%
- Iteration 設為 1000
- Learning rate 有用 adagrad 調整

結果分析：

- 訓練資料數少的時候，很容易就能很好的 fit 在訓練資料上。但在面對未知的資料，準確率就會很差。
- 資料多的時候，訓練資料不會 fit 這麼完美，但準確率在未知數據上較高。

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：



實驗設計：

- 取 PM2.5 為特徵
- 增加 PM2.5 的不同次方項，作為複雜度的變化
- Iteration 會跑到驗證誤差出現明顯極小值
- 其他設計跟前一題一樣

結果分析：

- 隨著次方數增加，一開始訓練的誤差是往下降的，應該是因為複雜的模型可以把訓練資料 fit 的更好。但到了 4 次方誤差卻開始上升，這應該是因為模型變得不好訓練，沒辦法找到 global minimum。

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

答：

Lambda	訓練 RMSE	驗證 RMSE	實驗設計： • 取 PM2.5 為特徵
0	6.1050	6.2191	
0.001	6.1064	6.2186	結果討論： • 當 labmda 等於 0.01，驗證 RMSE 出現最小值，因為正規化能讓模型變得較平滑一點，也較能在驗證資料上有好表現。 • 但變好的幅度並沒有很大，我的想法是因為訓練資料量太大 (5000 筆)，導致特徵本來就沒有辦法完全 fit 所有的點，因此正規化平時讓 overfit 的曲線平滑的優點就不明顯了。
0.005	6.1061	6.2177	
0.01	6.1060	6.2176	
0.05	6.1057	6.2185	
0.1	6.1059	6.2193	
1	6.1226	6.2474	
10	6.4809	6.6457	
100	8.0594	8.2279	

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：

$$\begin{aligned}
 Loss &= \sum_{n=1}^N (y^n - w \cdot x^n)^2 \\
 &= (X^T w - y)^T (X^T w - y) \\
 &= (w^T X - y^T)(X^T w - y) \\
 &= w^T X X^T w - 2w^T X y + y^T y \\
 \nabla_w Loss &= \nabla_w (w^T X X^T w - 2w^T X y + y^T y) \\
 &= 2X X^T w - 2X y \\
 \nabla_w Loss = 0 &\Rightarrow X X^T w = X y \Rightarrow w = (X X^T)^{-1} X y
 \end{aligned}$$