

ML HW1 REPORT

學號：B04611015 系級：資工二 姓名：陳佳佑

1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)

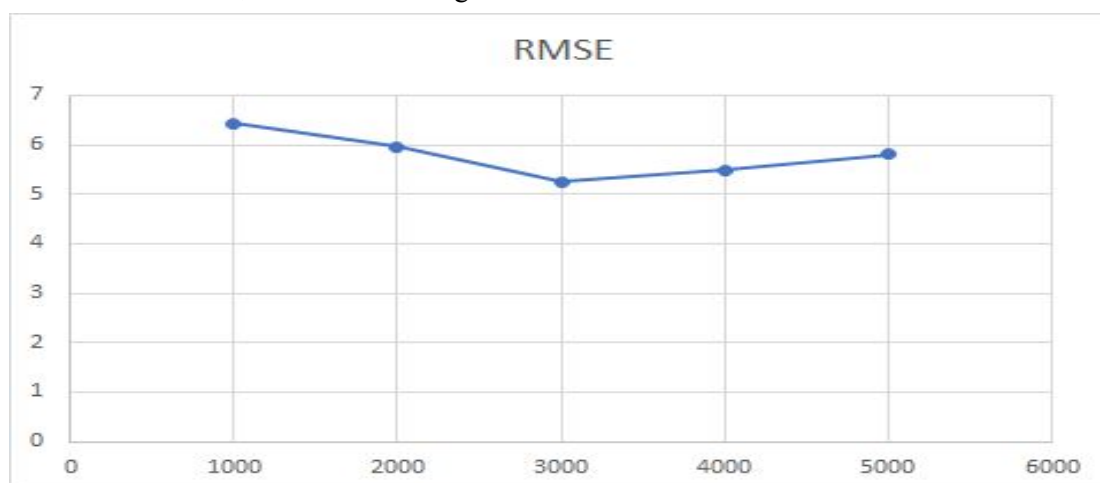
答：

一開始我直接使用前九天的所有資料一共是 18×9 個維度，在用fixing learning rate的情況下，收斂到的結果很慘。然後我就開始單跑各個資料從前1個小時到前9個小時之結果，看RMSE下降之程度，最後篩出了PM2.5取前9個小時，PM10取前5個小時，O3取前3個小時，RAINFALL取前2個小時，PM2.5的平方取前4個小時，PM10的平方取前兩個小時。

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：我以adagrad調整learning rate，並將eta用0.011，full batch 1000000 epo，validation set 使用除了training set的所有其他data。

結果出乎意料，當使用一半的資料時獲得之結果居然比5000筆好，估計是後面資料有雜訊造成此現象，亦或是overfitting。



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：

全部data前9個小時	5.684863377
pm2.5 前9個小時	5.871061182
model in 1.	5.810728487

雖然在validation時，full的rmse明顯較低但送上kaggle時，卻是另兩個model有比較好的效果。此結果應該是因為複雜度過大造成variance過大，而資料量又沒大到足以讓variance下降造成之效果。

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

答：

理論上正規化會改善縮小variance。但對於這個Model (第一題中取的feature)，正規化並沒有幫助，反而產生更差的效果，這應該是因為model的bias過大，而variance大一點，反而讓Model有可能跑到較好的結果，而縮小variance就減少了其機率。

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：

可轉換為 $\min |X^T w^T - y|^2$

可以看成是min向量的norm，因此兩向量的方向應相同

$X^T w^T = y$ ，因為 X^T 為長方矩陣，沒有反矩陣，因此同乘 X

$XX^T w^T = Xy$, $w = ((XX^T)^{-1} Xy)^T$ if XX^T 可逆 else find Pseudo inverse