

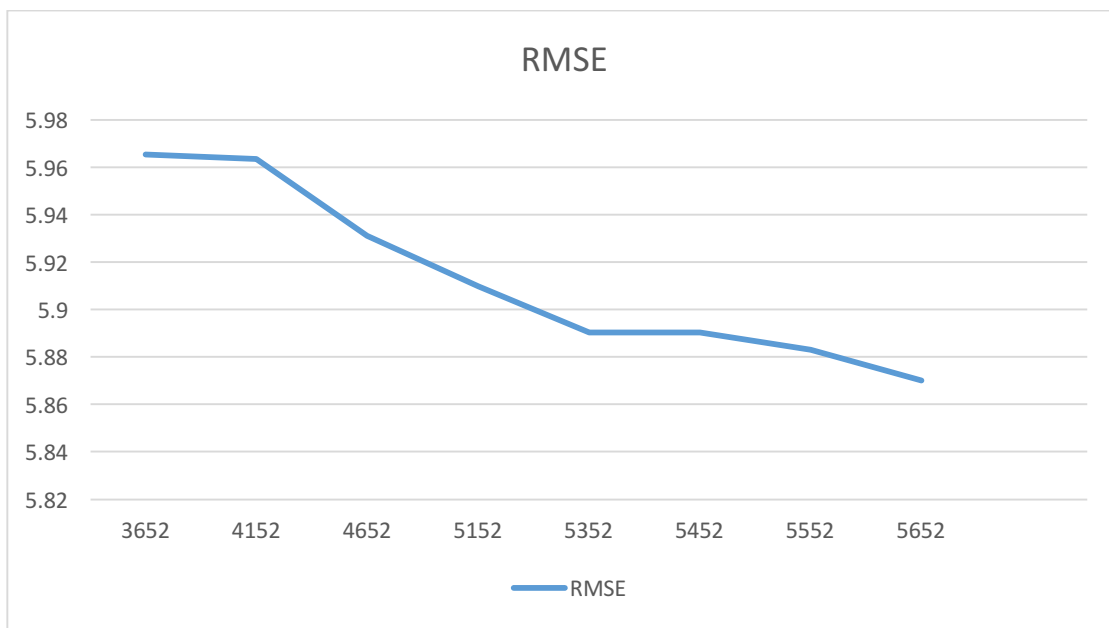
1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)

答：

每個月 20 天 * 24 小時，只取連續 9 小時所以總共能取 $20 * 24 - 9$ ，訓練資料共有 12 個月，故可以取得 $12 * (20 * 24 - 9) = 5652$ 筆訓練資料，另外每筆訓練資料的特徵一次方我只取 ['PM2.5', 'PM10', 'O3', 'AMB_TEMP', 'RH'], 二次方我只取 ['PM2.5', 'PM10'] 效果為最佳。

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：



另外在實驗中亦有發現，取較少的訓練資料量結果仍然有可能變好，原因可能在於資料本身 (例如: 極端值剛好沒被取到、取到的資料比較一致...等)。

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：

使用越複雜的模型在訓練資料下效果會比較好，但是在測試資料下效果則未必比較好，經過測試使用二次方效果為最佳（在訓練及測試下都比一次方都來得好）。

特徵為 ['PM2.5', 'PM10', 'O3', 'AMB_TEMP', 'RH']，VALIDATION SIZE 為 1000

	TRAIN RMSE	VALIDATION RMSE
1 次方	5.918155	5.898155
2 次方	5.797118	5.792701
3 次方	5.775844	5.869875
4 次方	5.582636	14.757032

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

答：

使用正規化可以避免 overfitting 的問題，但在本次作業的狀況比較不顯著，只會有些為影響。

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

答：

$$w = (X^T X)^{-1} X^T y$$