

1. (2%)記錄誤差值 (RMSE)(根據 kaggle public+private 分數), 討論兩種 feature 的影響

(1)9 小時所有 feature : public=7.4623, private=5.53033, RMSE=6.567742

(2)9 小時 PM2.5 : public=7.44013, private=5.62719, RMSE=6.596241

取所有 feature 作為模型 RMSE 值小於只取 PM2.5 的模型，顯示 PM2.5 預測受到過去其他項污染源的影響。

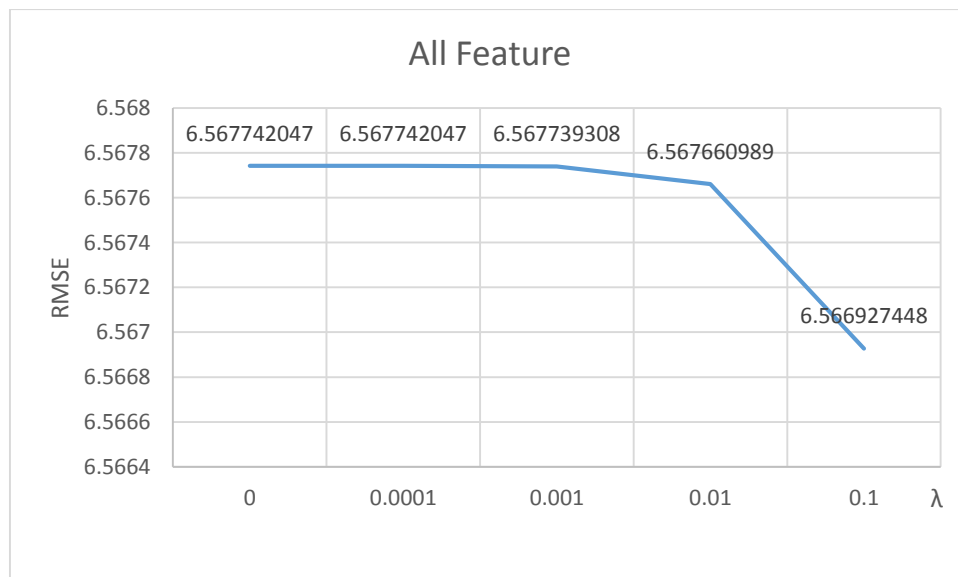
2. (1%)將 feature 從抽前 9 小時改成抽前 5 小時，討論其變化

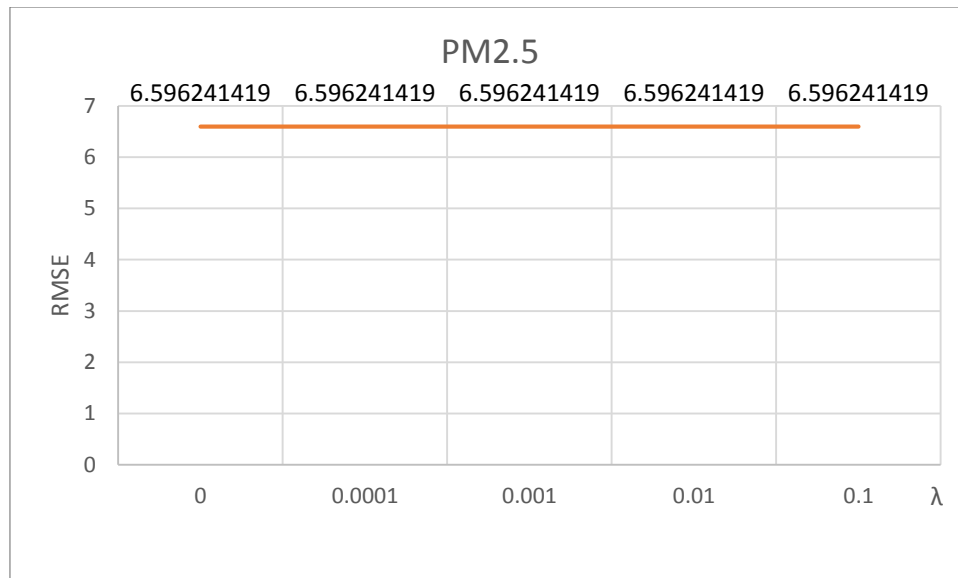
(1)5 小時所有 feature : public=7.65923, private=5.44089, RMSE=6.643308

(2)5 小時 PM2.5 : public=7.57904, private=5.79187, RMSE=6.744909

將 feature 改為抽取 5 小時後，兩者 RMSE 皆上升，其中取所有 Feature 作為模型的 RMSE 上升了 1.15%，而只取 PM2.5 的模型 RMSE 上升了 2.25%，再次顯示 PM2.5 數值受到其他項污染源的影響，並且受到過去 6~9 小時的影響。

3. (1%)Regularization on all the weight with  $\lambda=0.1$ 、0.01、0.001、0.0001, 並作圖





對取所有 feature 的模型進行 regularization 後，其 RMSE 有些微的下降，然而對於只取 PM2.5 的模型沒有影響。

4. (1%) 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  $\sum_{n=1}^N (\mathbf{x}^n - \mathbf{x}^n \cdot \mathbf{w})^2$ 。若將所有訓練資料的特徵值以矩陣  $\mathbf{X} = [\mathbf{x}^1 \mathbf{x}^2 \dots \mathbf{x}^N]^T$  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，請問如何以  $\mathbf{X}$  和  $\mathbf{y}$  表示可以最小化損失函數的向量  $\mathbf{w}$ ？請寫下算式並選出正確答案。(其中  $\mathbf{X}^T \mathbf{X}$  為 invertible)

- (a)  $(\mathbf{X}^T \mathbf{X}) \mathbf{X}^T \mathbf{y}$
- (b)  $(\mathbf{X}^T \mathbf{X})^{-0} \mathbf{X}^T \mathbf{y}$
- (c)  $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$
- (d)  $(\mathbf{X}^T \mathbf{X})^{-2} \mathbf{X}^T \mathbf{y}$

要使 Error  $E = \|\mathbf{y} - \mathbf{X}\mathbf{w}\|^2$  最小，須讓  $\mathbf{X}\mathbf{w}$  為  $\mathbf{y}$  在  $\mathbf{X}$  的 column space 的 Orthogonal Projection，也就是  $\mathbf{X}\mathbf{w} = \mathbf{P}_w \mathbf{y}$ ，其中  $\mathbf{P}_x = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ ，因此  $\mathbf{w} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ 。答案為 (c)。