學號：B03705027 系級： 資管三 姓名：鄭從德

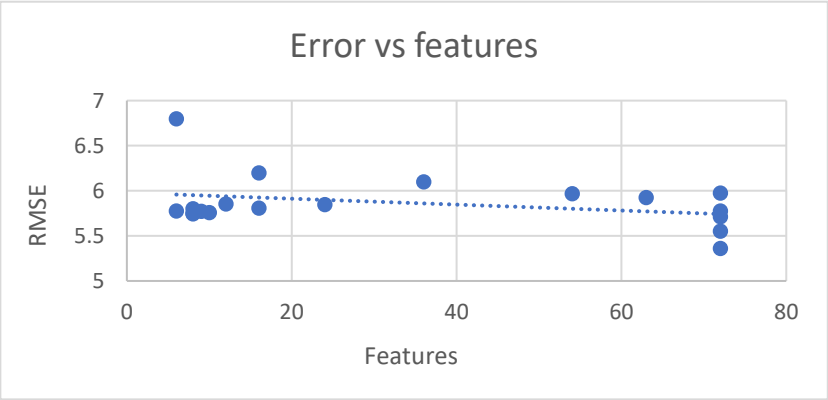
1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)

我一共抽取了八個觀測值：分別是 pm2.5、pm2.5 的平方、SO2、O3、風速、風向、一小時平均風速(WS_HR)、一小時平均風向(WD_HR)。這些都是取前九期的資料，所以一共有 72 個 features。我的選擇方法是每個側項都各別單獨與 pm2.5 跑一個 model 出來，error 下降的我就會在我最後的 model 中選擇它。

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

這是我用 Kaggle 上的 public set 得分與 dataset 大小關係做的圖。我僅挑 model 接近的幾筆資料進行做圖，因此資料點較少。由圖中可以看出，在我挑選 training data 多的時候，每次 Error 較近，變異數較小；後來利用 random 的方式切小 training set，error 的變異數因此增加，分布的更分散。總題而言還是小的 training set 誤差比較小。我覺得可能是因為取較多資料的話，同個資料點會被取到很多次，可能因此造成 overfit。

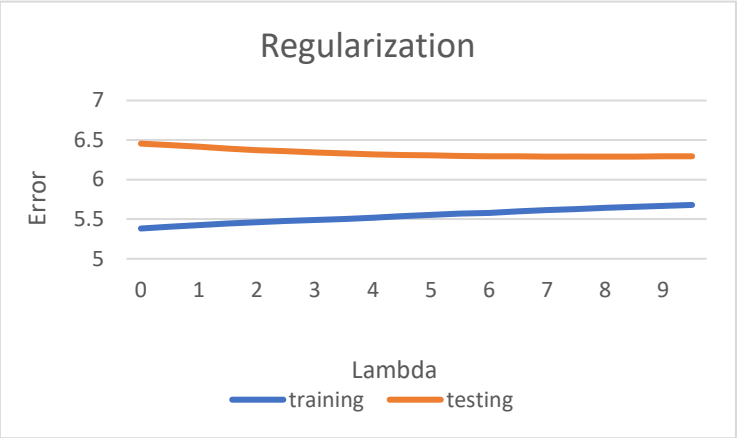
3. 請比較不同複雜度的模型對於 PM2.5 預測準確率的影響



若我們以 feature 數目判斷複雜度，則基本上 feature 越多 error 越低。

若我們以方程式複雜度討論，則由實作結果可以歸納：PM2.5 取二次方有助於 error 降低；其他的 features 的二次式則沒有顯著較果。三次方時則會使 test error 飆升(overfit)，因此我在最後 model 中僅使用二次式的模型。

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



上圖為重建一組 model 並隨機的進行 train test split 後的做圖結果。 testing set 最低誤差出現在 lambda = 6 的時候。再加入 lambda 的過程中，testing error 有變小的趨勢(幅度可達 0.3)，但過了一個臨界點後，testing error 還是會上升。
結論：適當的挑選 lambda 值，真的可以降低 testing error!!

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 ... x^N] 表示，所有訓練資料的標註以向量 y = [y^1 y^2 ... 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 。

$\varepsilon = y - w \cdot X$

$$\sum \varepsilon^2 = \varepsilon^T \varepsilon = (y - w \cdot X)^T (y - w \cdot X)$$

$$\frac{\partial}{\partial w} \varepsilon^T \varepsilon = 0$$

$$\frac{\partial}{\partial w} (y - w \cdot X)^T (y - w \cdot X) = 0$$

$$\frac{\partial}{\partial w} (y - w \cdot X)^2 = 0 \qquad -2X^T (y - w \cdot X)^2 = 0$$

$$X^T y = (X^T X) w$$

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