

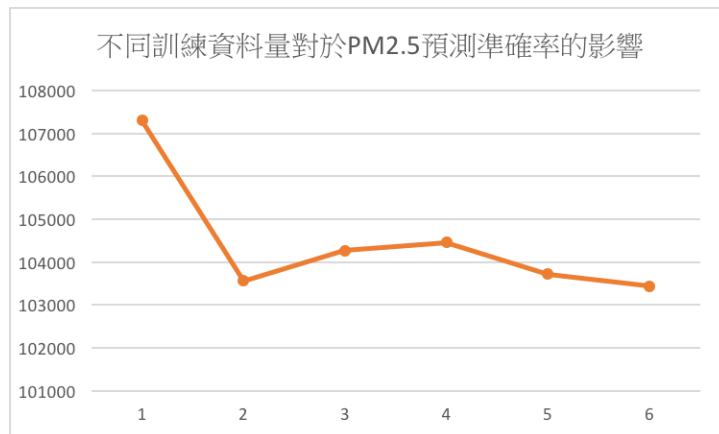
1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)

答：因為要預測 pm2.5 的值，故只抽取每小時 pm2.5 的值來做訓練，每組 data 含有連續九小時的 pm2.5 的值。而模型為對 pm2.5 為二次函數，其他項為一次函數。之後發現加入 o3 跟 rainfall 效果會變好因此有加入，但以下題目為了方便都只抽取 pm2.5

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

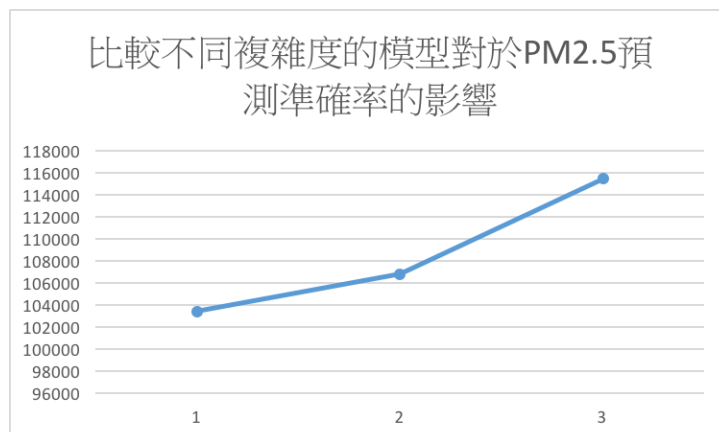
答：做法是取用 n 筆 data 作 training，再用 2826 筆（共六個月）data 作 test，比較不同訓練資料量對於 PM2.5 預測準確率的影響。

橫軸是 data 取的月數，分別從一到六個月，而縱軸是把每筆 loss 相加後的結果，可發現基本上資料量越多，loss 是會降低，但在某些情況下可能會上升一點，但不至於上升太多，可能跟剛好取到的 data 有關（可能造成像 2 的低點）。



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：做法是用不同次方的模型（分別從一次方到三次方函數），方法也是只取 pm2.5 再用 2826 筆（共六個月）data 作 test，比較不同次方函數對 PM2.5 預測準確率的影響。

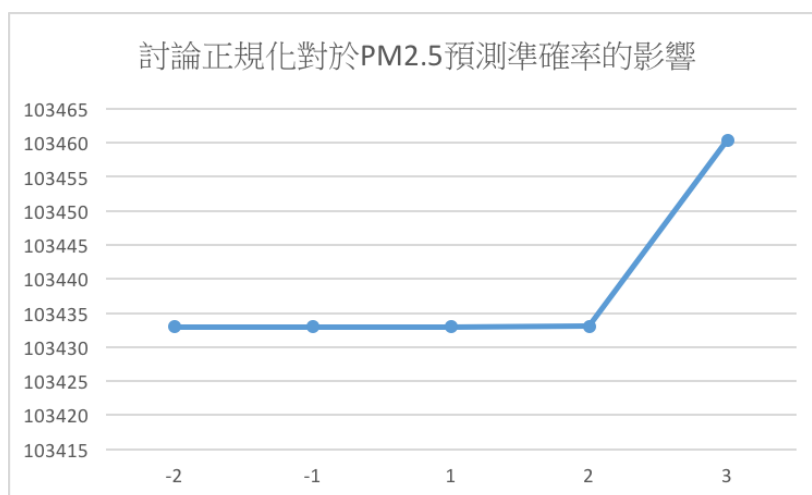


可發現一次方的偏差值最低，可能是次方變高會造成 overfitting 的現象，或是因為用 gradient descent 只能保證走到區域最小值，不能保證是 global minimum

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

答：做法是用不同係數來正規化作 training，再用 2826 筆（共六個月）data 作 test，比較不同係數對 PM2.5 預測準確率的影響。橫軸代表的是 lambda 的十的次方（從負二到三），縱軸是把每筆 loss 相加後的結果，可發現數值在 3 次方的時候突然變高，而最低點是發生在一次方的時候。

而跟沒做正規化的結果比起來正規化出來的結果似乎不會比較好，可能是因為是一次方的模型，所以平滑程度可能都差不多，故影響不大。



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 \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 \mathbf{b})，則線性回歸的損失函數(loss function)為 $\sum_{n=1}^N (y^n - \mathbf{w} \cdot \mathbf{x}^n)^2$ 。若將所有訓練資料的特徵值以矩陣 $\mathbf{X} = [\mathbf{x}^1 \mathbf{x}^2 \dots \mathbf{x}^N]$ 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，請以 \mathbf{X} 和 \mathbf{y} 表示可以最小化損失函數的向量 \mathbf{w} 。

答：

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