

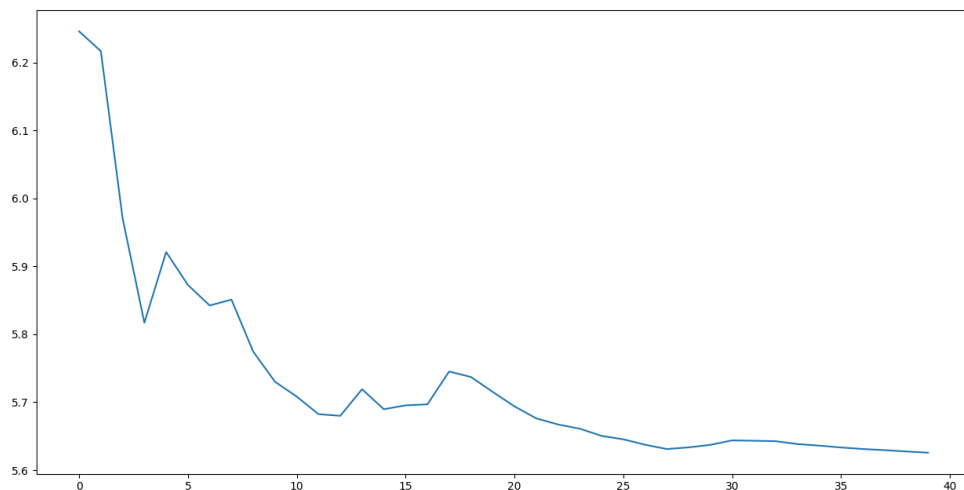
1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)

答：

先把資料按月份分為 12 組，再從每一組連續的 480 小時中取出 471 筆 data，最後總共有 $471 \times 12 = 5652$ 筆 data。接著再計算第 1 小時到第 9 小時的每一個觀測項以及其的平方對第 10 小時的 pm2.5 的相關係數，最後取較高的作為 feature。最後我使用的 feature 為前 9 個小時的 pm2.5 以及前 9 個小時的 pm2.5 的平方，總共 18 個 feature。

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

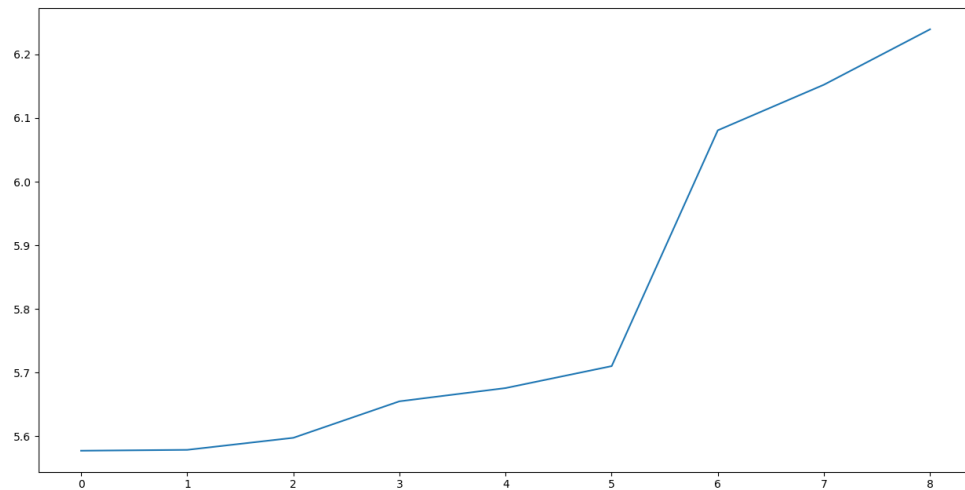
答：



我從所有 data 中分了 1652 個 element 作為 validation set，第 3 題也是相同作法。橫軸代表最後 training set 的大小，從 100 個到 4000 個，縱軸代表相應的 validating error。觀察此圖，可以發現大致上訓練資料量預測準確率呈正相關。

3. 請比較不同複雜度的模型對於 PM2.5 預測準確率的影響

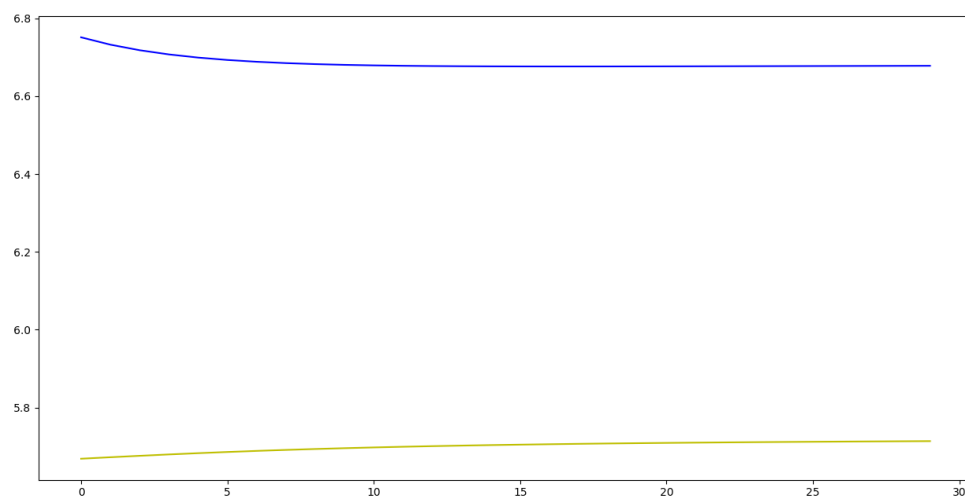
答：



此題我只有使用 pm2.5 作為 feature。橫軸代表從第 (x+1) 個小時開始取，例如 0 代表從第 1 個小時開始取，取到第 9 個小時，然後預測第 10 個小時的 pm2.5，以此類推。由圖可見，模型複雜度與預測準確度呈正相關。

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

答：



由於經過實驗之後，發現小樣本似乎比較容易出現 over fitting 的問題，因此我取前 200 個 element 作為 training set，剩下的 5452 個 element 作為 validation set。圖中藍色的線為 validating error，黃色的線為 training error，而橫軸為 regularization 中

λ 的值。可以發現隨著 λ 的變大，training error 逐漸變大，而 validating error 逐漸變小，但 λ 超過一個臨界點之後，兩者的 error 都緩慢地上升。

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：

$$\text{Loss function} = L = (y - Xw)^T(y - Xw)$$

We want to minimize the value of loss function:

$$\Rightarrow \frac{\partial L}{\partial w} = 0$$

$$\Rightarrow \frac{\partial}{\partial w} (y - Xw)^2 = 0$$

$$\Rightarrow -2X^T(y - Xw) = 0$$

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

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