

學號：R05922096 系級：資工碩一 姓名：李哲安

1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)

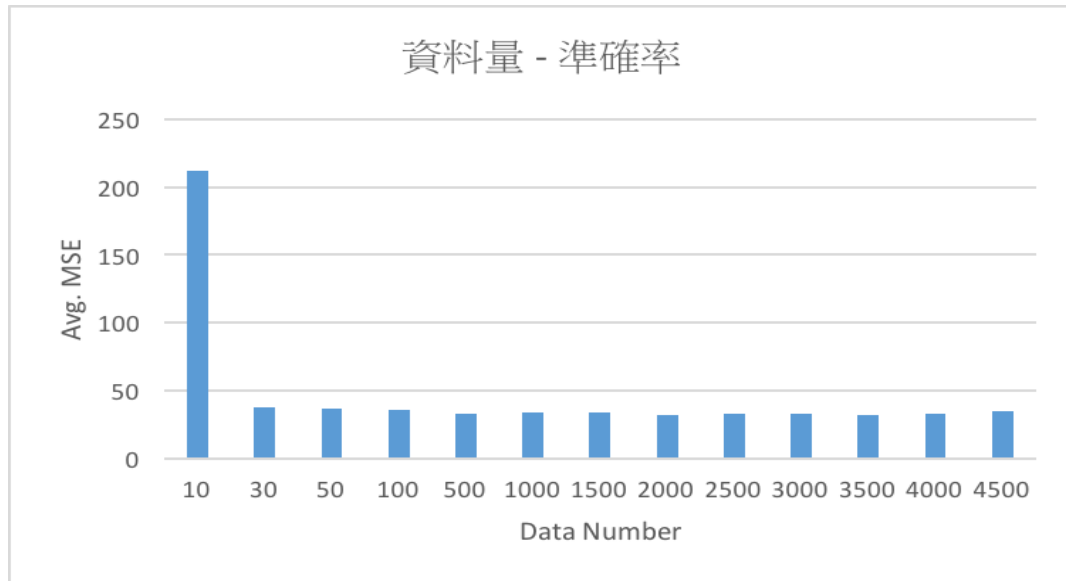
答：

只取連續 9 小時的 pm2.5 值當作 data，第 10 小時的 pm2.5 當作答案。

$\text{train\_x} = [x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9]$

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

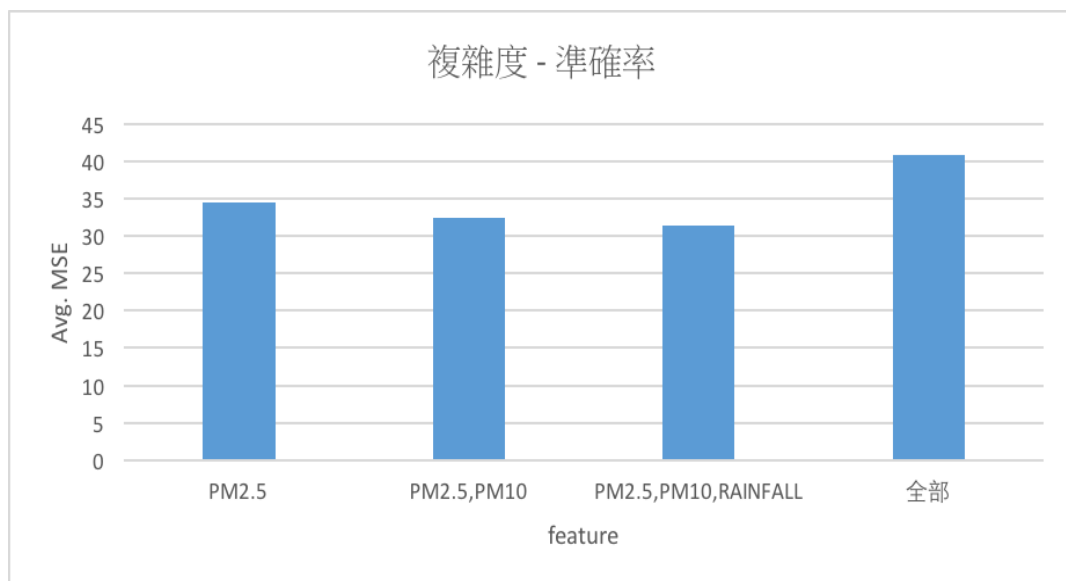
答：



資料量越多，準確率越高。但當資料量足夠多時，對準確率的影響很小。

3. 請比較不同複雜度的模型對於 PM2.5 預測準確率的影響

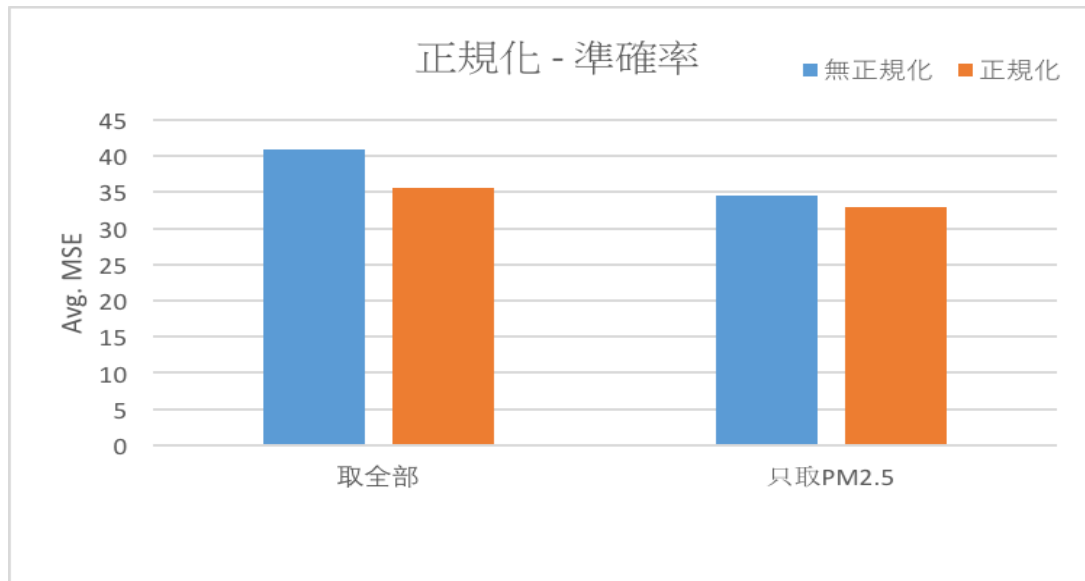
答：



提高複雜度(增加 feature 數)能提高準確率，但若複雜度太高則反而會有反效果。

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

答：



正規化能避免 model overfitting，在 PM2.5 中對於提高準確率幫助。

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：

k: data 維度

e:  $N$  by 1 (loss)

y:  $N$  by 1

w: 1 by k

X:  $N$  by k

$$\min \|e^T\| = \min \|(y - Xw^T)^T\|$$

當  $Xw^T = y$  對  $X$  的正交投影時，會有最小的 loss

$$\Rightarrow Xw^T = P_X y = X(X^T X)^{-1} X^T y \quad (P_X \text{ 為正交投影矩陣})$$

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

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