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)

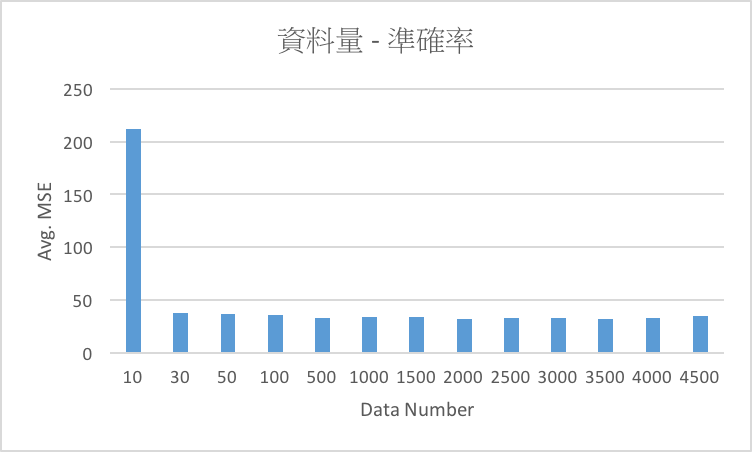
答：

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。

train\_x = [x1, x2, x3, x4, x5, x6, x7, x8, x9]

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

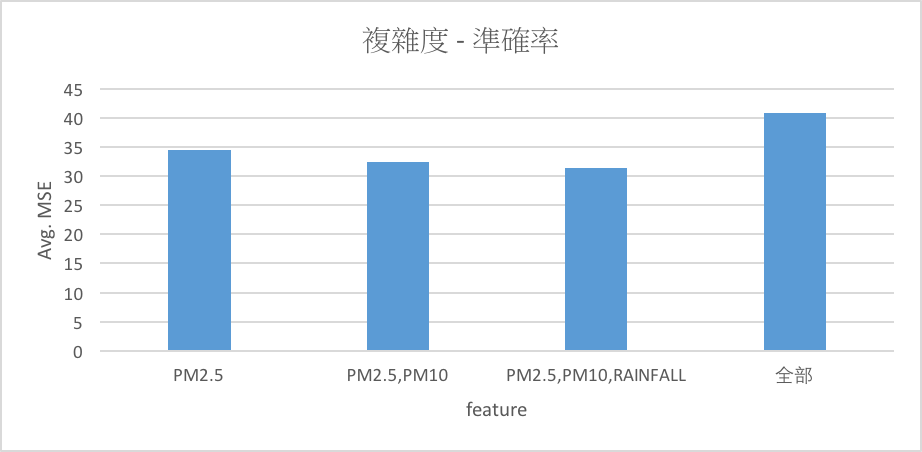
答：



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

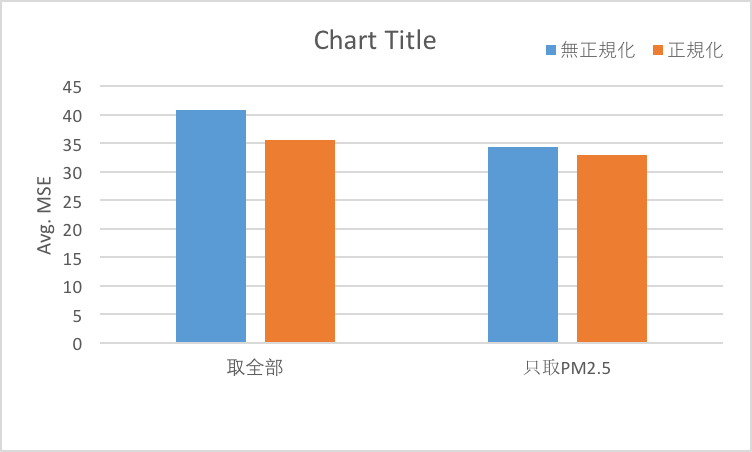
答：



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 xn，其標註(label)為一存量 yn，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 。若將所有訓練資料的特徵值以矩陣 X = [x1 x2 … xN] 表示，所有訓練資料的標註以向量 y = [y1 y2 … yN]T表示，請以 X 和 y 表示可以最小化損失函數的向量 w 。

答：

e: 1 by N y: 1 by N w: 1 by k X: k by N

min || e || = min || y – wX ||

=> wX = Pwy = w(wTw)-1wTy

=> X = (wTw)-1wTy

=> X = w-1(wT)-1wTy = w-1y

=> w = yX-1