

1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)

答：先只以 pm2.5 跑程式算出對全部 data 的 loss 值，在以 pm2.5 跟其他 feature 做兩兩配對算出 loss 值，找出能使 loss 下降的那些 features，再將所得的結果合起來作些微調整。得出的結果也與預想的差不多，除了跟相關係數高的某幾個 feature 有關之外，降雨、風向、風速等也會影響 pm2.5 的變化。

有趣的是，相關係數較高的 feature 不一定能建構出較準確的 model。

最後是取前九個小時的 [PM25, 2], [O3, 2], [PM10, 2], [RAINFALL, 1], [WD_HR, 1], [WIND_DIREC, 1], [WIND_SPEED, 1], [WS_HR, 1], [CO, 2] (後面數次代表次方)

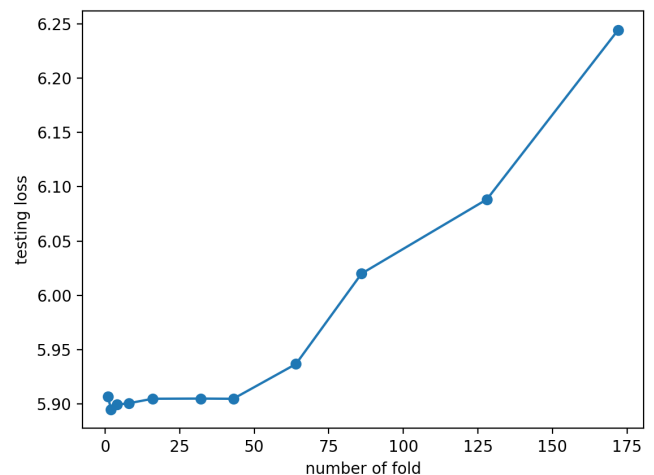
2. 請作圖比較不同訓練資料量對於PM2.5預測準確率的影響

答：我是以 k-fold cross validation 的方式，對每個 fold 進行gradient descent 算出 k 組參數，求平均後再計算對全部 data 的 loss 值。

k 值有[1,2,4,8,16,32,43,64,86,128,172]

對應的資料量為

[5504,2752,1376,688,344,172,128,86,64,43,32]



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：以下是以 16-fold cross validation 平均後對全部 data 的 loss 值

model (數字代表次方)	loss
[PM25, 1]	6.658376864
[PM25, 2]	6.302697106
[PM25, 3]	7.237896774
[PM25, 4]	7.832158745
[PM25, 5]	28.81379346
[PM25, 2], [PM10, 2]	6.144403964
[PM25, 2], [WD_HR, 2]	6.219301722
[PM25, 2], [WIND_DIREC, 2]	6.25203333
[PM25, 1], [O3, 1], [PM10, 1], [RAINFALL, 1], [WD_HR, 1], [WIND_DIREC, 1]	6.425252943

[PM25, 2], [O3, 2], [PM10, 2], [RAINFALL, 2], [WD_HR, 2], [WIND_DIREC, 2]	5.984947168
[PM25, 1], [AMB_TEMP, 1], [CH4, 1], [CO, 1], [NO2, 1], [NOx, 1], [O3, 1], [PM10, 1], [RAINFALL, 1], [SO2, 1], [WD_HR, 1], [WIND_DIREC, 1], [WIND_SPEED, 1], [WS_HR, 1]	6.525823718
[PM25, 2], [AMB_TEMP, 2], [CH4, 2], [CO, 2], [NO2, 2], [NOx, 2], [O3, 2], [PM10, 2], [RAINFALL, 2], [SO2, 2], [WD_HR, 2], [WIND_DIREC, 2], [WIND_SPEED, 2], [WS_HR, 2]	6.259758871

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

答：lambda 小時對整體影響不大，而隨著 lambda 增加，loss 值亦增加

lambda	loss
0	5.906580019
1	5.906638664
5	5.907147048
10	5.908337136
100	5.993581452
1000	7.055613877
10000	10.31704481

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 $\sum_{n=1}^N (y_n - w^T x_n)^2$ 。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：

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