

1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)

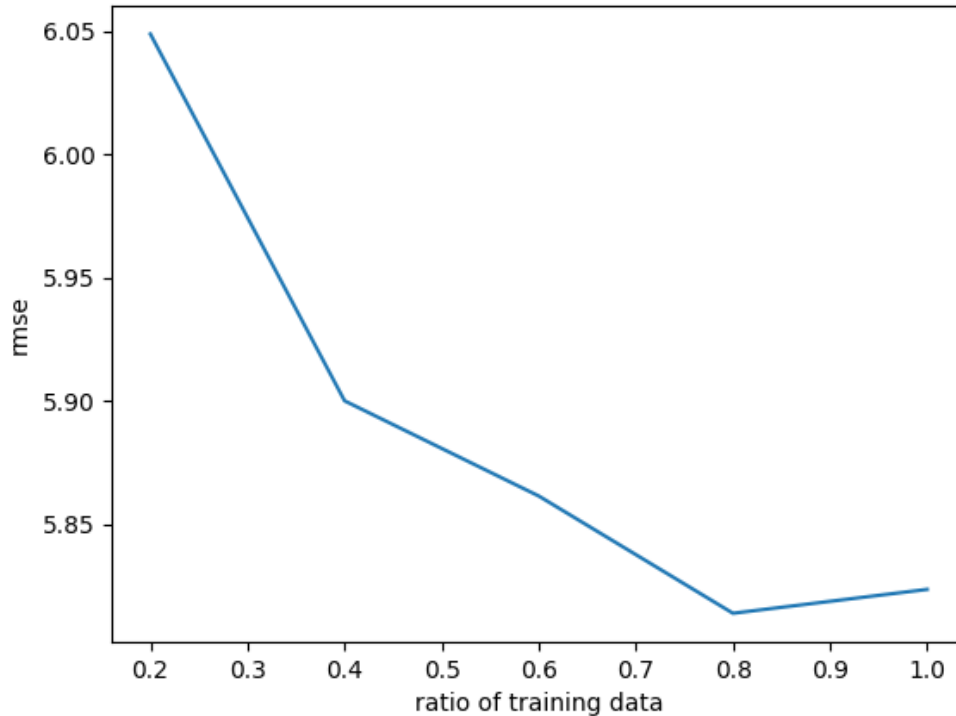
答：

首先刪除 `WD_HR` 這個特徵，然後把 `WIND_DIREC` 從角度換成二維平面上的兩個向量，因此每個小時的空汙指標依然是 18 維。取前 9 個小時的所有空氣污染指標（18 維），共 162 維，先做一次線性迴歸，再從線性迴歸的權重中挑出那些絕對值大於 0.1 的特徵。

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：

共有 5652 筆訓練資料，以下為分別使用 0.2, 0.4, 0.6, 0.8, 1.0 的訓練資料所跑出的結果。



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：

後來嘗試在做第一次線性迴歸之前，把每一維的資料平方之後接在每組訓練資料之後，形成 324 維的訓練資料。然而在相同的參數下，方均根誤差從原來的 5.81400 升高到了 5.94744。可見增加模型複雜度不見得會增加預測準確率。

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

答：

理論上而言，因為線性迴歸所要最小化的損失函數是凸函數，所以有全域最佳解。而正規化也是線性的操作，因此不論是否有做正規化，皆可以得到一組等價的最佳解。所以正規化理論上對預測準確度不會有影響。

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 $X = [x^1 \ x^2 \ \cdots \ x^N]$ 表示，所有訓練資料的標註以向量 $y = [y^1 \ y^2 \ \cdots \ y^N]^T$ 表示，請以 X 和 y 表示可以最小化損失函數的向量 w 。

答：

$$\begin{aligned} & \sum_{n=1}^N (y^n - w \cdot x^n)^2 \\ &= \|y - Xw\|^2 \\ & \frac{\partial}{\partial w} \|y - Xw\|^2 \\ &= 2X^T(y - Xw) \\ & 2X^T(y - Xw) = 0 \\ & w = (X^T X)^{-1} X^T y \end{aligned}$$