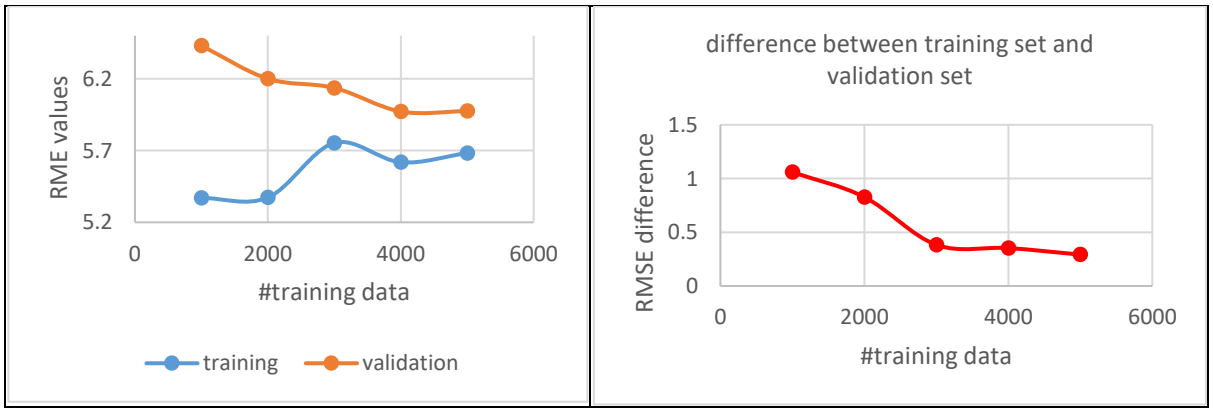
學號：B03901096　系級：　電機三　姓名：周晁德

1. 請簡明扼要地闡述你如何抽取模型的輸入特徵（feature）

答：因為是要使用九小時資料預測第十小時的 PM2.5，所以我把每個月 20 天，每天 24 小時的資料，共 480 小時的資料，每個連續 9 小時的資料當作 input，第十小時的 pm2.5 當作 label，所以每個月有 471 筆 training data，一年 12 個月就有 5652 筆 training data，我嘗試過選取 18 種不同的因素任意組合，最後發現 $pm2.5, O_3, CO, wind\_speed$, wind_Hr，搭配組合結果最好!

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：每筆 training data 是由每九小時的 18 種因素共 162 個資料，及第十天 pm2.5 的濃度所組成，使用一次方程式 $y = wx + b$，w, b initial 為 0，並且使用 ada_grad，initial learning rate=0.1，共跑 5000 個 iterations，validation set 是從 training set 中切出來的 652 筆 data，比較 training data 數量分別為 1000、2000、3000、4000、5000 筆資料的結果



隨著 training dat 的增加:

(1)training set 的 RMSE 值有增加的趨勢(上方左圖)。

(2)validation set 的 RMSE 值漸減(上方左圖)。

(3)training set 和 validation set 的 RMSE 值差距變小(上方右圖)。

(4)只用 1000 筆 training data overfitting 最嚴重(上方左圖)

由此可知隨著 training set 的增加，結果比較不容易 overfitting，預測結果也有越準確的趨勢。

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：1 st:$y = wx + b$;2 nd:$y = w_2x^2 + w_1x + b$ ； 1type:pm2.5 ； 18types:所有因素

traing set	1st	2nd
1 type	6.11956	6.103137
18 types	5.647987	5.454796

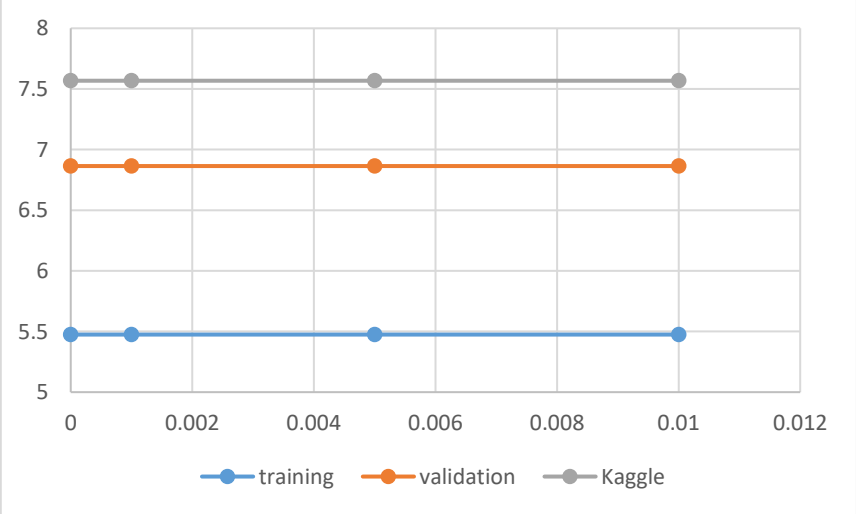
validation set	1st	2nd
1 type	6.158427	6.244267
18 types	6.022059	6.188182

由上表可知，在使用相同 data type(例如都使用 pm2.5 或 18 種因素都考慮)的情況下，在 training set 上一次式的 RMSE 比二次式的 RMSE 高，但在 validation set 上一次式的 RMSE 較二次式的 RMSE 低，由此可知提高 function 的複雜度，更容易發生

overfitting 的情況，預測的準確度可能因此下降。若在相同 function(同為一次式或二次式)之下，增加 data type，不論在 training set 或 validation set 的 RMSE 都下降，由此可知增加 data type 將使預測 pm2.5 時考慮到更多影響因素，預測較準確。

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

答：因為 regularization 是用來解決 overfitting 的問題，由 2.可知當只使用 1000 筆 training data 時，overfitting 最嚴重，所以選用 1000 筆 training，其他條件同 2.，比較 λ=0、0.01、0.005、0.001。



由左圖可知隨著 λ 值的不同對於預測準確度的影響並無太大差別，在我自己程式實作時也有試著加入 regularization，但只讓我在 Kaggle 上的 RMSE 值減少 0.00003，因此我認為加入 regularization 對於提升預測準確度並無太大幫助。

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 $X = [x^1\ x^2\ \cdots\ x^N]$ 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：X 矩陣 shape 是 N＊d 維，其中 N 維資料筆數，d 為 feature 數量，y 矩陣的 shape 為N＊1維

$$L = \sum_{n=1}^{N}(y^n - w \cdot x^n)^2 = (Xw - y)^T(Xw - y)$$

$$\nabla_w L = \nabla_w (Xw - y)^T(Xw - y) = \nabla_w (w^T X^T Xw - 2w^T X^T y + y^T y) = 2X^T Xw - 2X^T y = 0$$

$$\therefore X^T Xw = X^T y \rightarrow w = (X^T X)^{-1} X^T y$$

6. 我的 hw1.py 和 hw1_best.py

這兩個程式是相同的，model 都是選擇$pm2.5, O_3, CO, wind_speed, wind_Hr$，使用二次方程式$y = w_2 x^2 + w_1 x + b$。