

學號：B06705057 系級：資工二 姓名：黃資翔

請實做以下兩種不同feature的模型，回答第 (1) ~ (3) 題：

抽全部9小時內的污染源feature當作一次項(加bias)

抽全部9小時內pm2.5的一次項當作feature(加bias)

備註：

- NR請皆設為0，其他的數值不要做任何更動
- 所有 advanced 的 gradient descent 技術(如: adam, adagrad 等) 都是可以用的
- 第1-3題請都以題目給訂的兩種model來回答
- 同學可以先把model訓練好，kaggle死線之後便可以無限上傳。
- 根據助教時間的公式表示，(1) 代表 $p = 9 \times 18 + 1$ 而(2) 代表 $p = 9 \times 1 + 1$

1. (2%)記錄誤差值 (RMSE)(根據kaggle public+private分數)，討論兩種feature的影響

Model \ RMSE	Public	Private
Case 1 (163 個參數)	5.63779	7.21546
Case 2 (10 個參數)	5.90263	7.22356

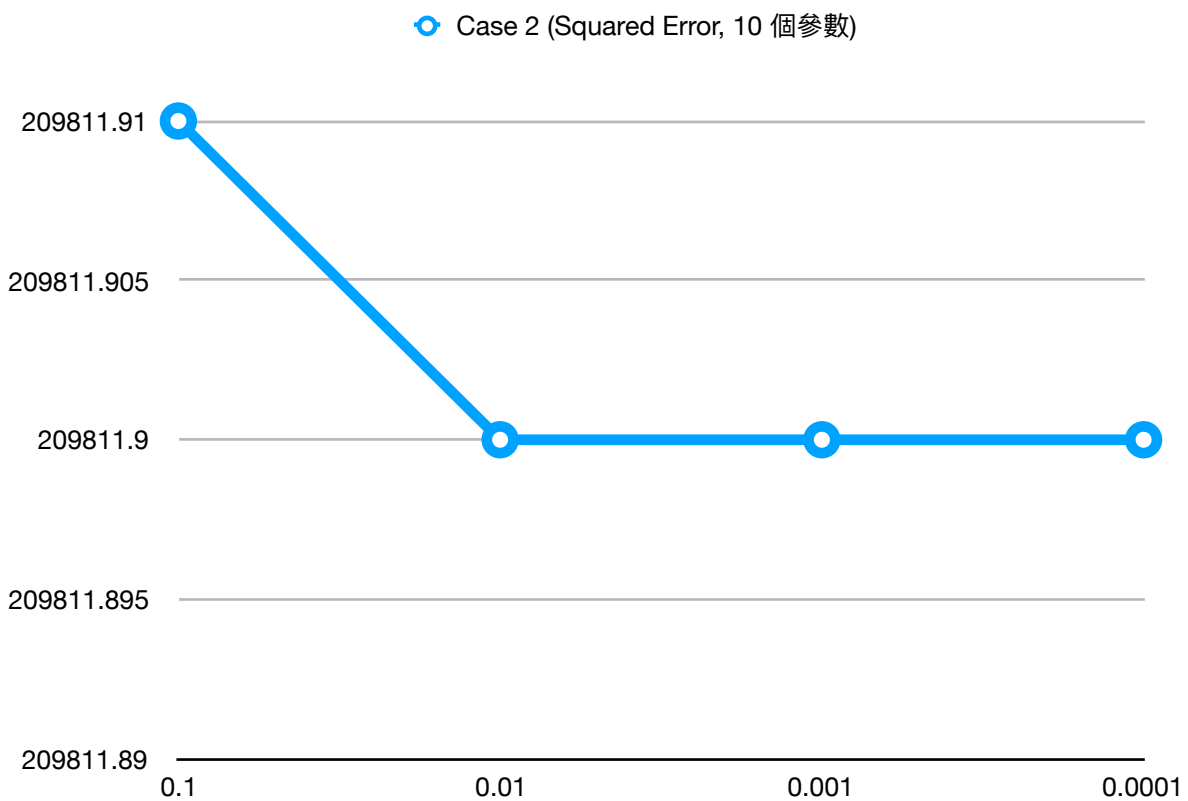
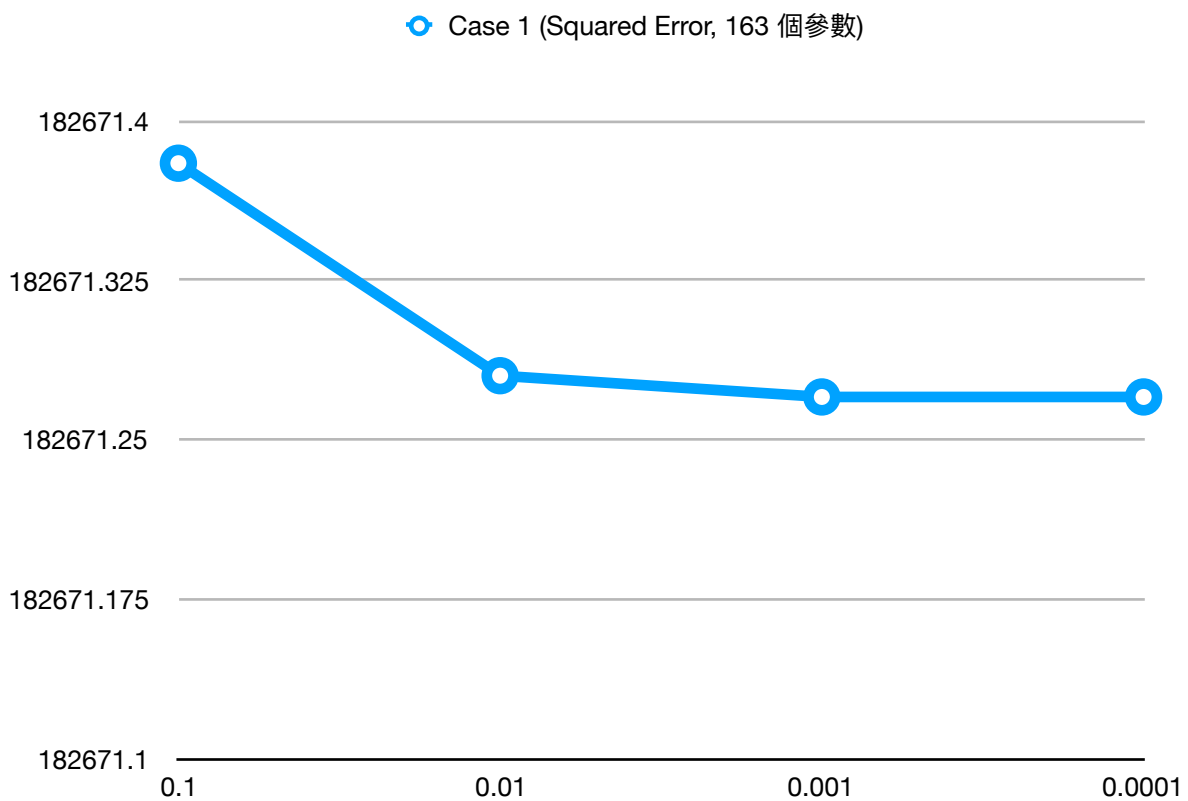
平均而言 Case 1 的表現比較好，因為參數較多，其所包含的 function set 較大，data 有五千多筆足夠多，因此不太可能 overfitting。Case 2 雖然只考慮 pm2.5，但結果也不錯，由此可知前幾小時的 pm2.5 對下一小時的 pm2.5 佔很重要的因素

2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

Model \ RMSE	Public	Private
Case 1 (91 個參數)	5.98287	7.16743
Case 2 (6 個參數)	6.22749	7.22464

與第一題很像，參數較多的模型平均而言表現比較好。若 data 沒有那麼多的話，則參數不可以太多，只考慮前 5 小時也是不錯的模型。

3. (1%)Regularization on all the weight with $\lambda=0.1$ 、 0.01 、 0.001 、 0.0001 ，並作圖



4. (1%)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)為一純量 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 $\sum_{n=1}^N (y_n - x_n w)^2$ 。若將所有訓練資料的特徵值以矩陣 $X = [x_1 \ x_2 \ \dots \ x_N]^T$ 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，請問如何以 X 和 y 表示可以最小化損失函數的向量 w ？請選出正確答案。(其中 $X^T X$ 為invertible)

Ans：(C)

與 b06705058 劉品桉 討論