

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

(1) 抽全部 9 小時內的污染源 feature 的一次項(加 bias)

用 Close form 為初始值的情況下

Public Score: 7.46325

Private Score: 5.50302

(2) 抽全部 9 小時內的 pm2.5 的一次項當 feature (加 bias)

用 Close form 為初始值的情況下

Public Score: 7.47217

Private Score: 5.59538

Ans: 由上述紀錄發現使用全部 feature 會使結果更好, 推論原因是因為未來 pm2.5 的值不只和過去 pm2.5 的值有關, 和其他 feature 也有相關, 而當使用到不相關的 feature 只要 train 後的 weight 為 0 也不會有太大的影響, 故會有上面的結果。

2. 將 feature 從抽前 9 小時改成抽前 5 小時, 討論其變化

(1) 抽全部 9 小時內的污染源 feature 的一次項(加 bias)

用 Close form 為初始值的情況下

Public Score: 7.68223

Private Score: 5.47722

(2) 抽全部 9 小時內的 pm2.5 的一次項當 feature (加 bias)

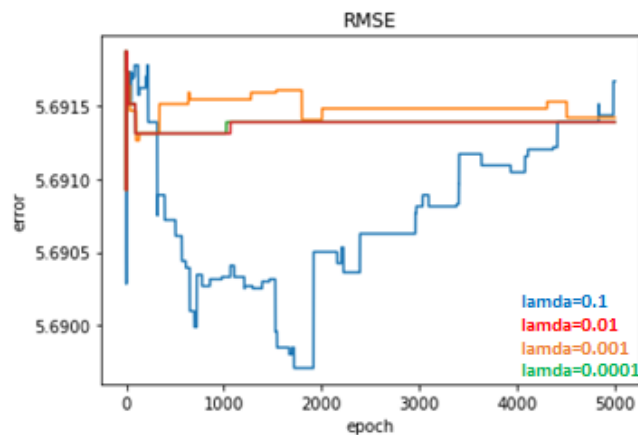
用 Close form 為初始值的情況下

Public Score: 7.56692

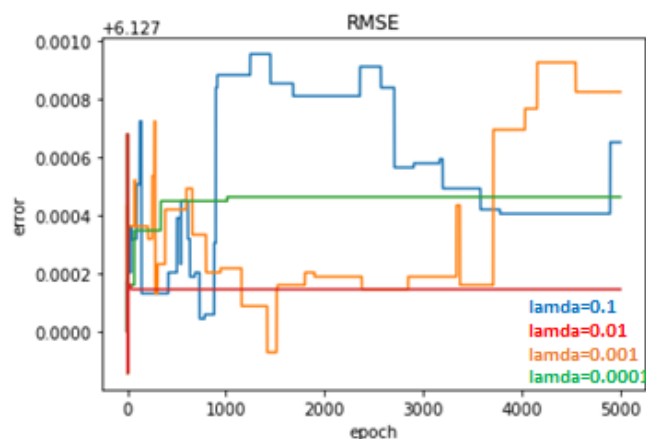
Private Score: 5.82022

Ans: 與(1)結果比較發現public score都變大了, 但private的結果卻不一定, 這個結果可能是因為當9小時改成5小時後, 代表feature變少了, 但是training data數量卻上升, 所以出來的結果就不一定變好或變差。

3. Regularization on all the weight with $\lambda=0.1$ 、 0.01 、 0.001 、 0.0001 ，並作圖
(1) 抽全部 9 小時內的污染源 feature 的一次項(加 bias)



- (2) 抽全部 9 小時內的 pm2.5 的一次項當 feature (加 bias)



4. 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 $\sum_{n=1}^N (y^n - x^n w)^2$ 。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 ？請寫下算式並選出正確答案。(其中 $X^T X$ 為 invertible)

Ans:

$$XW = y$$

$$(X^T X) W = X^T y$$

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

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

故選 (C)