

請實做以下兩種不同 feature 的模型，回答第 (1) ~ (3) 題：

(1) 抽全部 9 小時內的污染源 feature 當作一次項(加 bias)

(2) 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

(1)RMSE=(5.89279+7.71956) / 2 = 6.806175(2) RMSE=(6.07042+6.63952) / 2 = 6.35497

Iteration 皆為 100000，在處理之前記得先把跨月份的 data 拿掉，使用 adagrad，可以發現到說在 public 上所有污染源 feature 表現來的比只拿 pm2.5 來的好，但是很明顯地在 private 上 pm2.5 比所有污染源 feature 來的好，平均起來 pm2.5 表現較佳，可以猜測所有污染源有可能發生 overfit

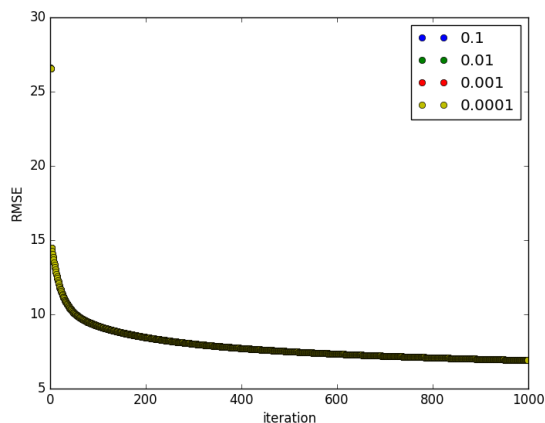
2. (1%)將 feature 從抽前 9 小時改成抽前 5 小時，討論其變化

(1)RMSE=( 18.96290+ 19.42313) / 2 = 19.193015 (2) RMSE = (19.57443+ 21.05805) / 2 = 20.31624

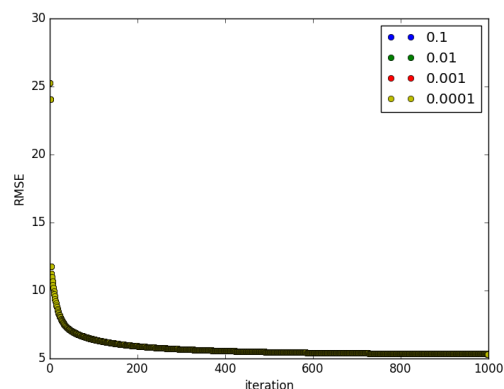
Iteration 同樣為 100000，同時也把跨月份 data 處理掉，可以發現到說兩者的表現並不大好，勉強可以說所有污染源優一點，但還是與取 9 小時差異甚大

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

(1)



(2)



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

- (a)  $(X^T X) X^T y$
- (b)  $(X^T X) y X^T$
- (c)  $(X^T X)^{-1} X^T y$
- (d)  $(X^T X)^{-1} y X^T$

ANS:(c)， $Y = X \cdot W$ ， $X^T Y = W$ ，同時最小化所以須加上 $(X^T X)^{-1}$ 純量。