

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：

- a. NR 請皆設為 0，其他的數值不要做任何更動
- b. 所有 advanced 的 gradient descent 技術(如: adam, adagrad 等) 都是可以用的
- c. 第 1-3 題請都以題目給訂的兩種 model 來回答
- d. 同學可以先把 model 訓練好，kaggle 死線之後便可以無限上傳。
- e. 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

	kaggle Public	kaggle Private
9 小時內 PM2.5	5.90263	7.22356
9 小時內所有的 feature	5.65650	7.27081

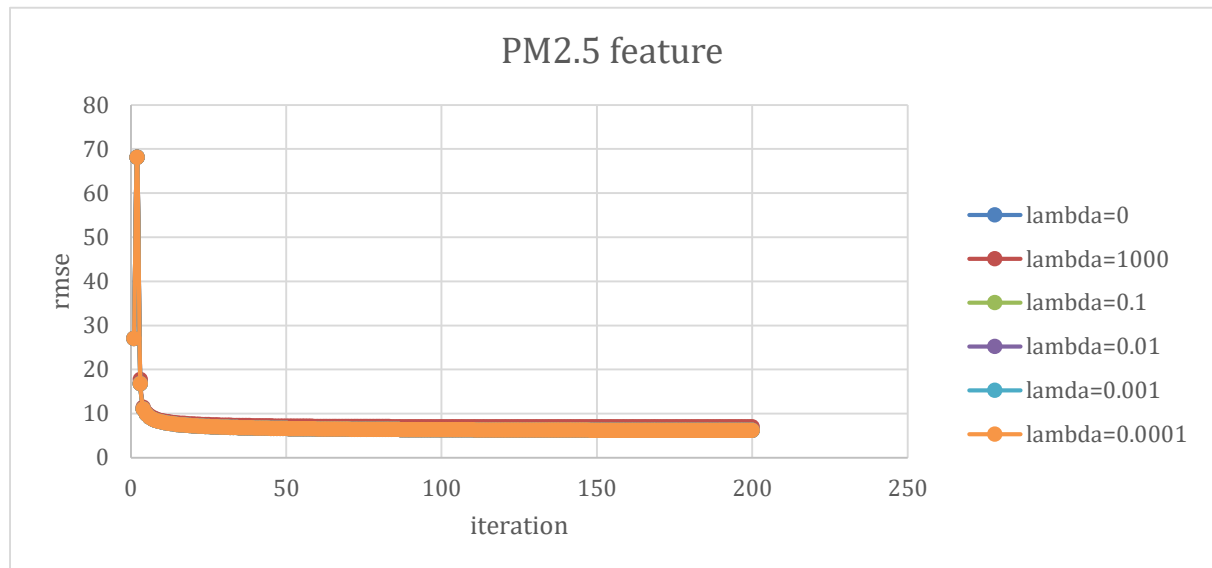
只取 pm2.5 在 public 分數差於取所有 feature，但在 private 的分數卻較好，應該是取所有 feature 會有 overfit 的問題，只在 public 效果較好。然而，在實作時若對資料做過處理，並取所有 feature 下去 train，其實效果仍較佳，因此兩種 feature 說明 pm2.5 並不是只受 pm2.5 自己的前時段影響，而是有其他因素會影響濃度。

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

	kaggle Public	kaggle Private
9 小時內 PM2.5	5.90263	7.22356
5 小時內 PM2.5	19.53866	20.21173
9 小時內所有 feature	5.65650	7.27081
5 小時內所有 feature	18.99177	19.39797

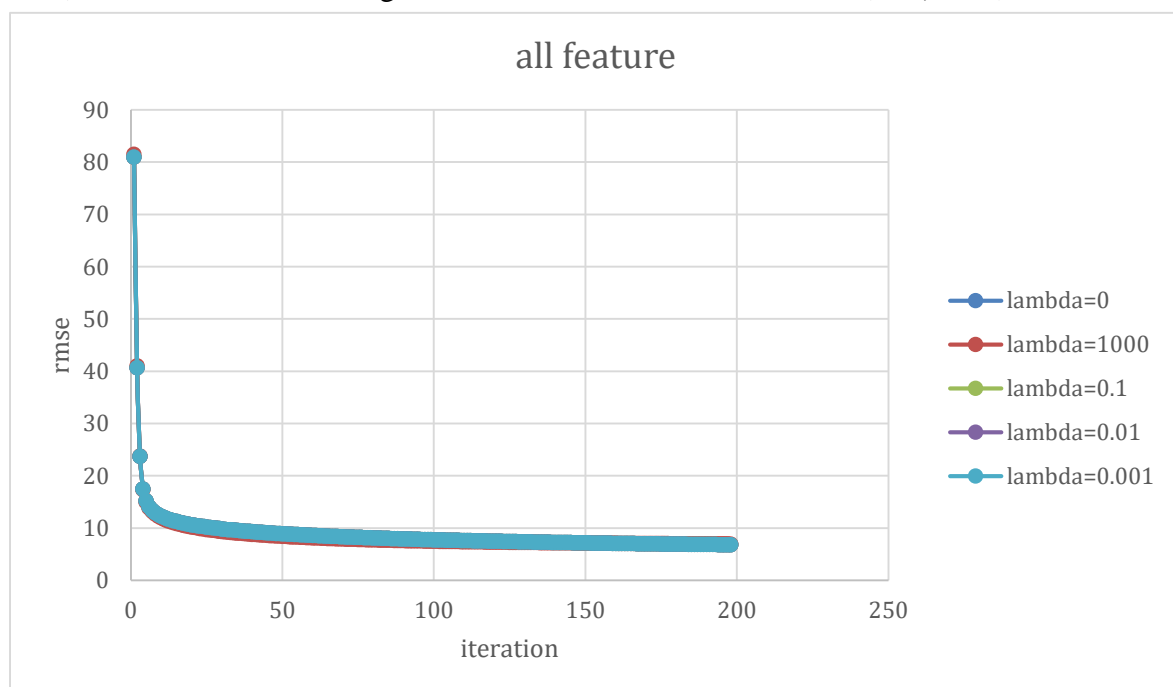
若只取前 5 個小時的資料，會發現不論何種方式誤差皆大幅度增加，可能 data 取的量不足有關，必須更多時間的資料去預測。此外，也有 overfit 的問題，private 實際結果都沒有比較好。之後應該嘗試若取 10 個小時或是更多時間的資料去看是否會對結果有所影響。

3. (1%)Regularization on all the weight with  $\lambda=0.1$ 、 $0.01$ 、 $0.001$ 、 $0.0001$ ，並作圖  
只取 pm2.5 作為 feature，且 learning rate=10, iteration=10000 取前 200 筆方便觀察



由圖形可以發現， $\lambda=1000$ 時下降速度慢，是因為曲線變得更平滑，但進步時既沒有很多。而其他  $\lambda$  的差別難以用圖形去呈現，因此推測若只取 pm2.5 會有 underfit

取所有 feature，且 learning rate=10, iteration=10000 取前 200 筆方便觀察



本資料去除前兩筆 iteration=1 和 2，因為在 iteration=2 時 RMSE 高達 600 對圖形分析較為困難，推測 learning rate 設得太大。由圖形可以得知，在  $\lambda=0,0.1,0.01,0.001,0.00001$  時，由於設定值太小，對是否有 overfit 或是有解決 overfit 較難判斷。因此將  $\lambda$  設為 1000 時，可以發現下降速度會變快，效果有略為增加，加入 regularization 能使結果略有進步，判斷有些許 overfit。

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

**b**

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