

學號：B05705053 系級：資管三 姓名：蔡涵如

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  
使用adagrad, learning rate開1000的情況下，將所有pm2.5小於零的training都刪掉，並且  
10feature運算速度快很多，也比較快收斂

	Private	Public
所有的feature	7.27654	5.75537
只用pm2.5	7.21994	5.90006

結論:抽取全部的feature未必會比只抽取pm2.5好。

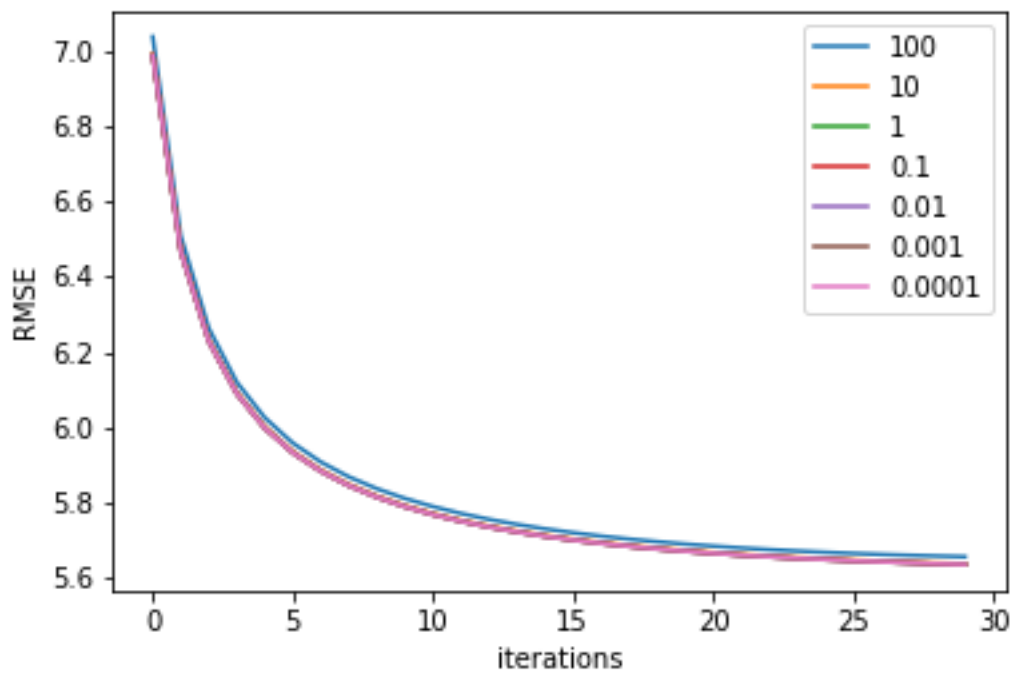
2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

	Private	Public
前九個小時	7.27654	5.75537
前五個小時	7.23824	5.99182

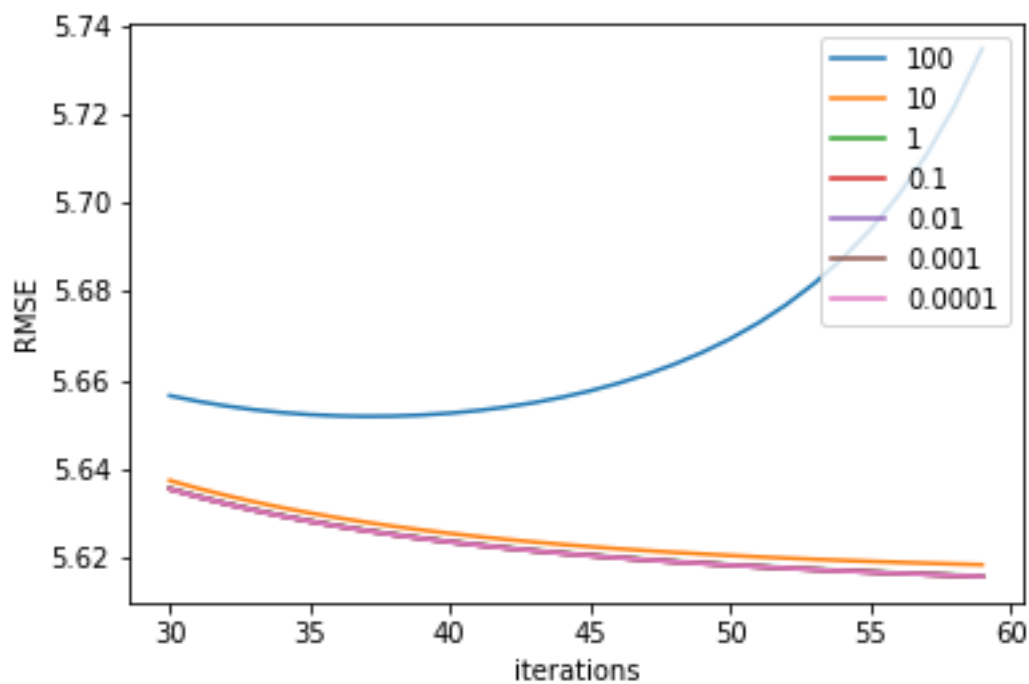
結論:抽取全部的feature未必會比只抽取pm2.5好。

因為如果有noise的feature在linear regression裡面是會影響RMSE的，所以還是必須做  
feature extraction來避免noise feature的干擾，並且可以發現private public set上面沒有一定的  
相對高低。

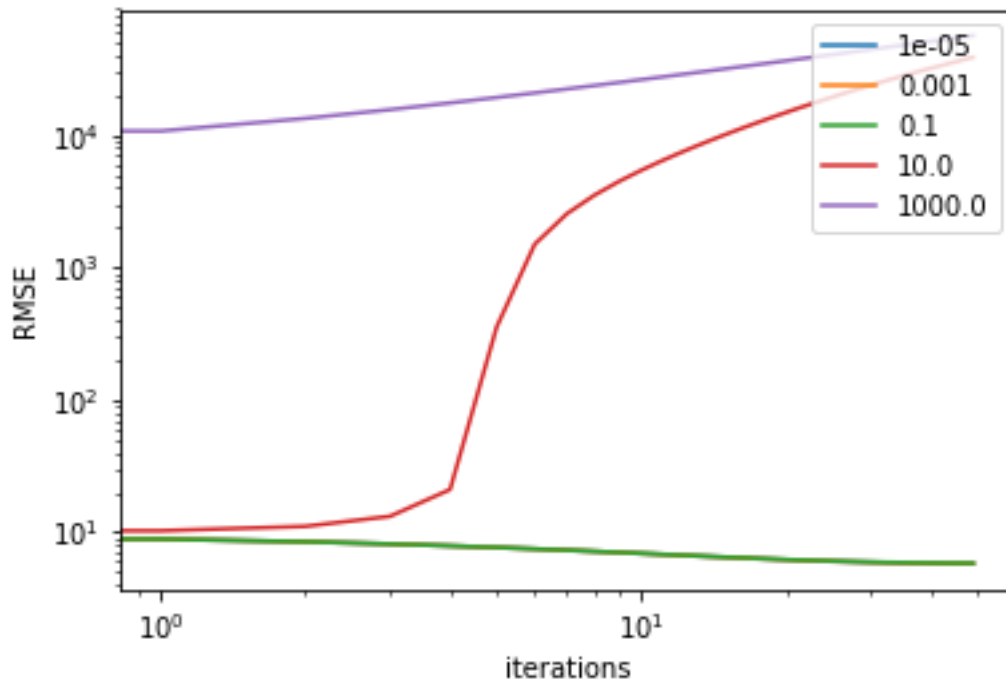
3. (1%)Regularization on all the weight with  $\lambda=0.1$ 、 $0.01$ 、 $0.001$ 、 $0.0001$ ，並作圖  
因為我不斷的調整參數所有的alpha值的線還是重疊在一起，所以我決定多算了幾個alpha分  
別是100, 10, 1, 0.1, 0.01, 0.001, 0.0001，對其做圖的結果如下，我將之分成train的前三十個  
iteration,以及後30個iteration,如以下圖所示，RMSE alpha為一百的一開始就大於其他的數  
值，而後三十個iteration, 則表示出 $\alpha > 100$ ，會讓越過最佳解而gradient無限上升，而前面



iteration點可以觀察的出來,rmse其實都保持著定值，原因是因為線性模型調整alpha會讓 regression line 平移。



然而因為觀察這個結果之後，只有會不會一次跳太大而不而越過終點線的特徵，其實有點類似 learning rate 的效果，我另外又實作了二次函數的 regression, 將所有的 feature 的平方的都加進 feature 當成新的 feature, 以下圖是我所發現的結果, alpha 太大或太小都會讓 model testing set 的 RMSE 越來越大，唯有設 0.1 時會讓 model 的 testing set 得 RMSE 會變最小。



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)