

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

這邊使用手刻並有分大小的 linear regression model

(1)**Public:5.56462:Private:6.63584** 若給電腦全部的污染源則它可能可以由這些污染源看出一些關係，能預測的較為精準。我認為其中影響最大的是風向，特別是原檔案的風向是角度，需轉為 **cos** 及 **sin** 兩項。修正後 **error** 可以減少個 0.2~0.3，非常可觀。

(2)**Public:5.83309:Private:7.25859** 若只有 **PM2.5** 的資料，則只能觀察它前九個小時的起落(兩項的差)來預測最後第十小時的值，若有較為異常(前面都增加，第十小時卻減少)的狀況發生，可能就無法預測。

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

這邊使用手刻並有分大小的 linear regression model

(1)**Public:5.94978:Private:6.99282** 根據一些我的觀察，若是把一些不重要的 **feature** 去除 **error** 可能會降低。但是這邊的結果顯示前 4 個小時的 **data** 也是有其重要性存在。

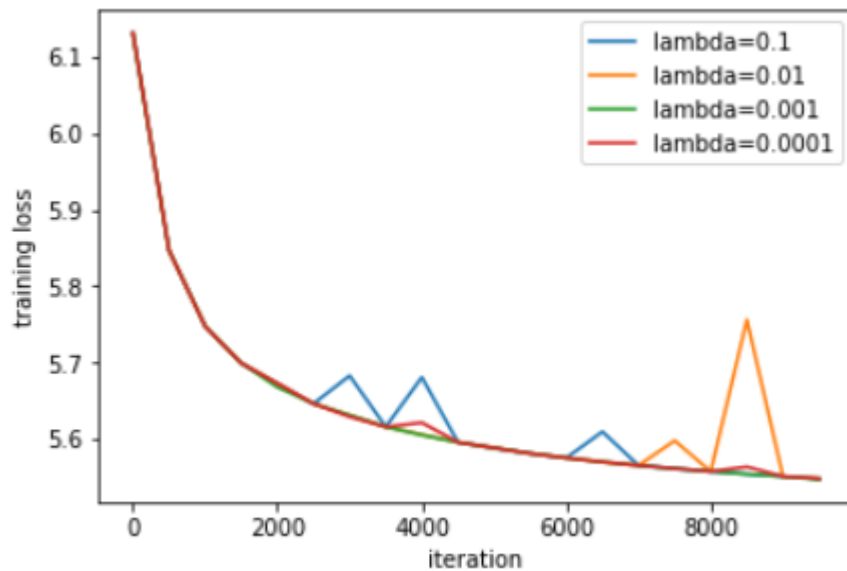
(2)**Public:6.22829:Private:7.28860** 這邊 **feature** 非常少，基本跟 1-2 一樣，只是能觀察的趨勢又更少，顯然 **error** 會更大。

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

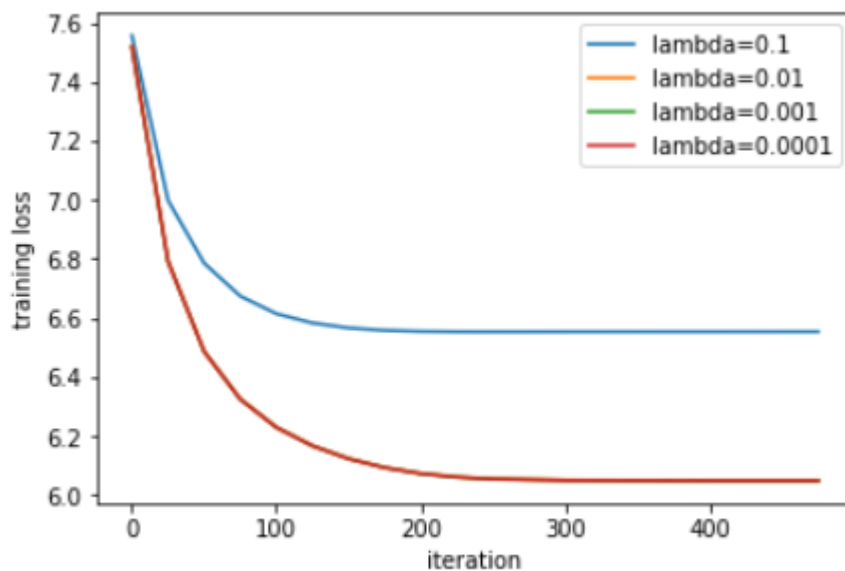
這邊使用手刻並有分大小的 linear regression model

(1)從原始數據可以看出 **training error** 在小數點後三位有不同，但是整體影響基本上從這個例子看不出來。我的猜想是由於 **Regularization** 是為

了避免 **overfit**，然而這個例子沒有高次項，看不出有嚴重 **overfit** 的問題，所以 **Regularization** 在此例子是看不出什麼作用



(2)同 3-1，但是這裡 feature 更少了。然而這裡可以看到  $\lambda=0.1$  對 weight 限制太大，導致 error 壓不下來。



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  $\mathbf{x}^n$ ，其標註(label)為一純量  $y^n$ ，模型參數為一向量  $\mathbf{w}$  (此處忽略偏權值  $b$ )，則線性回歸的損失函數(loss function)為  $\sum_{n=1}^N \frac{1}{2} (y^n - \mathbf{x}^n \cdot \mathbf{w})^2$ 。若將所有訓練資料的特徵值以矩陣  $\mathbf{X} = [\mathbf{x}^1 \mathbf{x}^2 \dots \mathbf{x}^N]^T$  表示，所有訓練資料的標註以向量  $\mathbf{y} = [y^1 y^2 \dots y^N]^T$  表示，請問如何以  $\mathbf{X}$  和  $\mathbf{y}$  表示可以最小化損失函數的向量  $\mathbf{w}$ ？請選出正確答案。(其中  $\mathbf{X}^T \mathbf{X}$  為 invertible)

(c)  $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$