

學號：B05902018 系級：資工三 姓名：張凱程

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 $p = 9 * 1 + 1$

1. (2%)記錄誤差值 (RMSE)(根據 kaggle public+private 分數)，討論兩種 feature 的影響

(1) 全部 feature: $5.51664 + 7.07990$

(2) PM2.5 feature: $5.79583 + 7.17286$

因為(1)的 feature 比(2)多，可以做參考的項目會比較多，因此前者會去找非 PM2.5 的 feature 跟 PM2.5 的相關性，會增加其準確率。

附註：(1)是選沒有被選到但是比較好的結果。

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

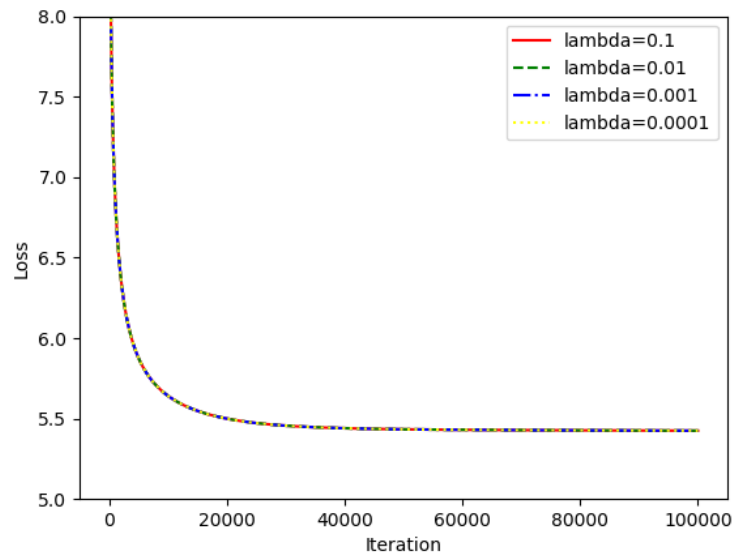
(1) 全部 feature: $5.51664 + 7.07990 \rightarrow 5.88784 + 7.11336$

(2) PM2.5 feature: $5.79583 + 7.17286 \rightarrow 6.16914 + 7.15599$

由於剩下的 4 小時還是有其參考性，他們的 weight 也需要進行學習，所以正確率可能會比較差。

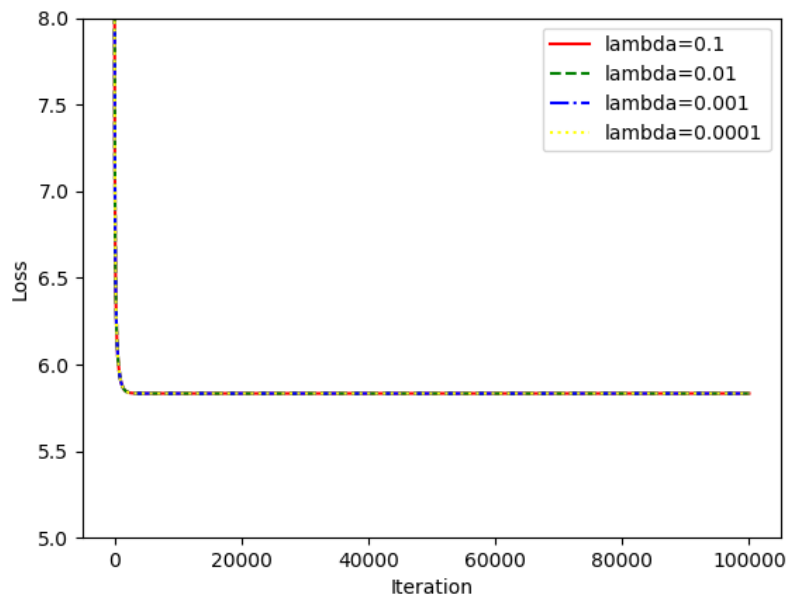
3. (1%)Regularization on all the weight with $\lambda=0.1$ 、 0.01 、 0.001 、 0.0001 ，並作圖

(1)



基本上四項是重疊的，因為誤差極小，差距大概 10 的-8 至-9 次方，但是 λ 大的，loss 還是會稍微大一點，因為 weight 的長度影響到 training 的程度比較高。

(2)



一樣四項都很接近，且收斂速度很快。

兩者的差別在，前者做 100,000 次 iteration 時 loss 還是會稍作更新，後者之後 loss 就完全不變。

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 $\sum_{n=1}^N (y^n - x^n w)^2$ 。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. $(X^T X)yX^T$
- c. $(X^T X)^{-1}X^T y$
- d. $(X^T X)^{-1}yX^T$

Answer: c.