

學號：R07922104 系級：資工碩一 姓名：林傳祐

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

	public	private
all feaure	5.78115	7.24601
pm2.5 only	5.93024	7.24786

在 public 上看，只用 pm2.5 參數數量似乎不足以來預測

但在 private 上卻幾乎得到相同的 loss

或許如果能有更好的方式做出更好的 model

只用 pm2.5 也能預測得很好也說不定

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

	public	private
all feaure	5.98585	7.13413
pm2.5 only	19.51523	20.19665

在這裡我們可以看到，對於使用全部 feature 來說

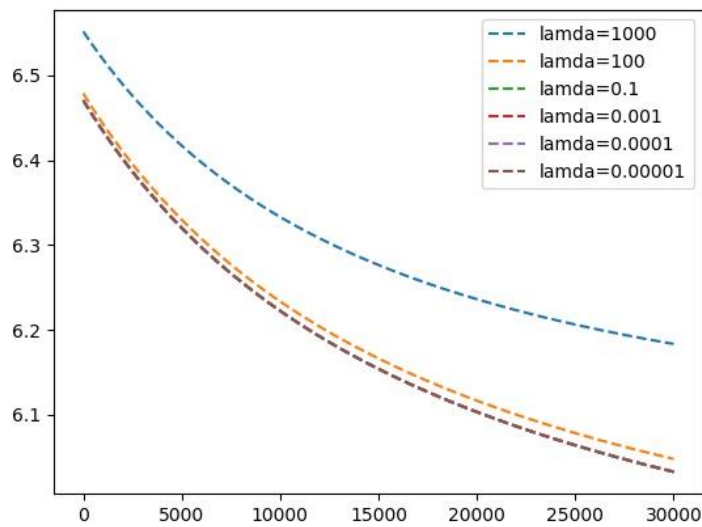
使用 5 或 9 小時，在 Public & private 各有勝負

應該都是不錯的方式

但只用 pm2.5 又只有 5 小時，加上 bias 只有 6 個參數

可能是 loss 如此高的原因

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



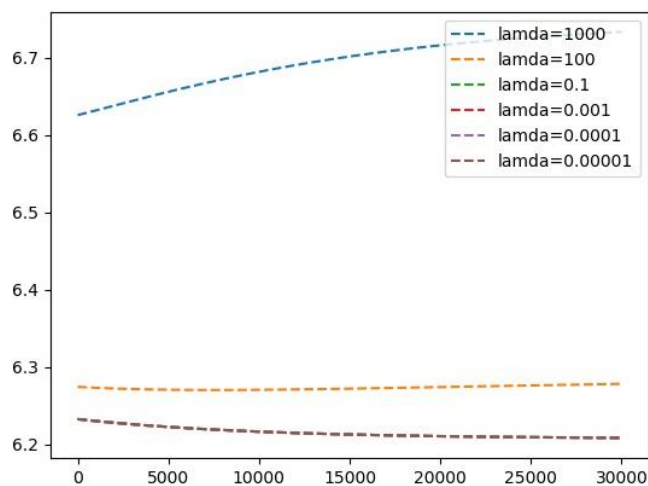
(iter40000 次，取後面 30000 次做圖)

因為只使用一次式，regularization 效果無法出來

加上 lambda 都差不多，所以 4 個都重疊

所以我再實作差很多倍的，可以稍微看出影響

大 lambda 會有較高的 loss



(iter40000 次，取後面 30000 次做圖)

這裡也可以看出 lambda 大有較大的 loss

值得一提的是，lambda 過大時，會使 loss 反而變高

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

C