

學號: R07943095 系級: 電子碩一 姓名: 劉世棠

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) all feature : 5.63779(public)+7.21544(private)=12.85323

(2) pm 2.5 only : 5.90263(public)+7.22356(private)=13.12619

=> 由上述的資料可以發現在指定的 **learning rate** 和 **iteration times** 中看所有 **feature** 的效果較好

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

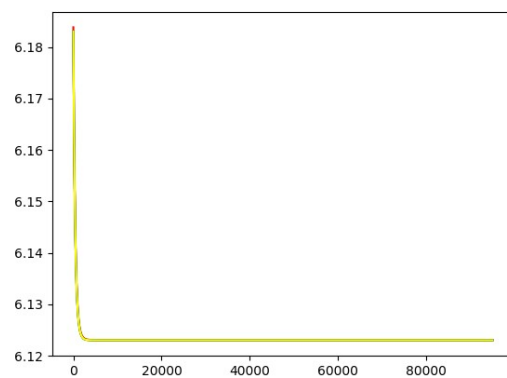
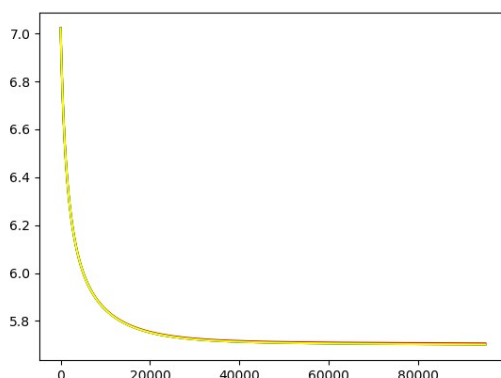
(1) all feature : 5.98231(public)+7.16746(private)=13.14977

(2) pm 2.5 only : 6.22732(public)+7.22552(private)=13.45284

=> 由上述的資料可以發現在指定的 **learning rate** 和 **iteration times** 中, 資料預測的結果比九小時的差, 如果不做其它修改則除了 **all feature** 的 **private** 有改進 (有可能是剛好才這樣) 之外並無其它優勢。

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

根據實驗結果, 因為我的模型只有使用 linear model, 所以曲線並不會明顯地分開, 以下分別為 all feature (左) 和 PM 2.5 only (右):



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