

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

抽取全部(163 個 feature)的誤差值為 private：7.19763、public：5.63644

抽取 pm2.5(9 個 feature)的誤差值為 private：7.22578、public：5.91715

可以發現兩種 feature 都能夠通過 simple_baseline，因此先前的 pm2.5 對於之後 pm2.5 的影響是相當大的，然而也不是只有 pm2.5 會影響，因此在抽取較多 feature 的狀況下，並沒有出現 overfitting。

Collaborator：黃資翔 b06705057

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

我認為的抽取五小時為：同樣的九個小時，捨棄前四小時的 feature，然後只用後面的五個小時當作 feature，也就是 $[0,0,0,0,w_0,w_1,w_2,w_3,w_4,y]$ ，再將結果上傳至 kaggle。

(1)

抽取五小時的誤差值為 private：7.15633、public：5.97597

(2)

在只有九個 feature 的狀況下，抽取五小時的誤差值為

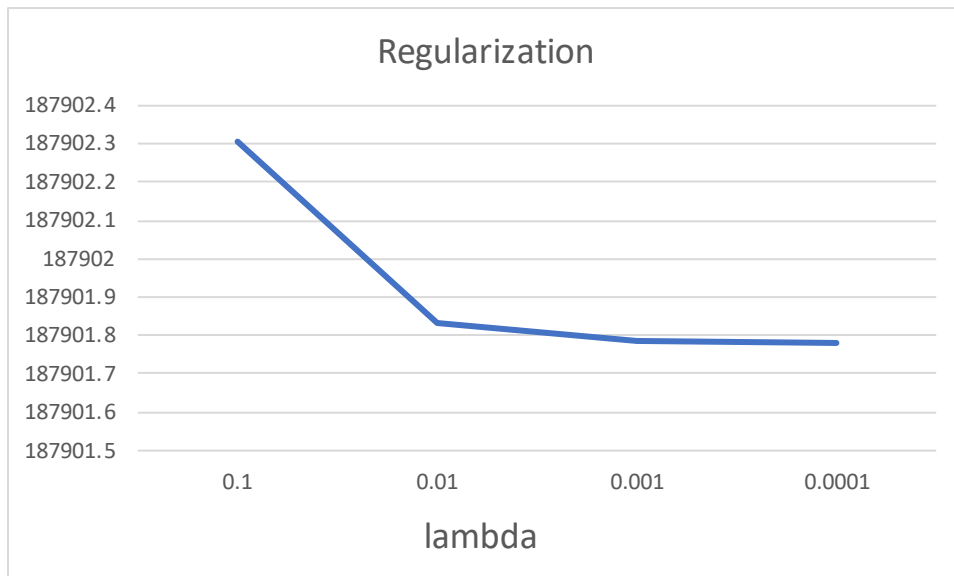
private：7.15811、public：6.29355

兩者的誤差值在 public 上升，但在 private 下降。推測應該是運氣好，命中了 private 的測資，而因為刪掉了一些參數，所以在 public 的表現就比原本九小時的差。

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

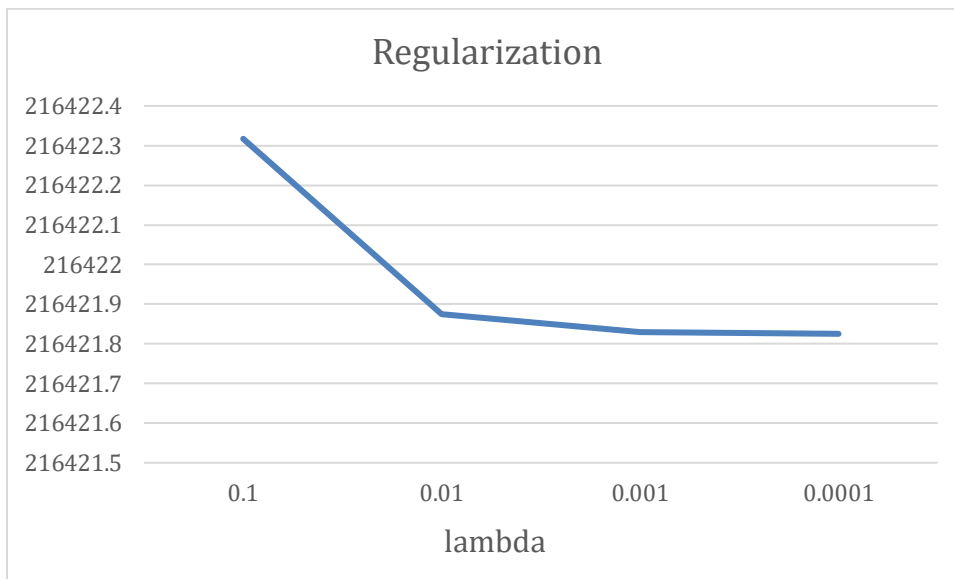
(1)

163 個 feature。



(2)

9 個 feature 。



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) $(X^T X)^{-1} y X^T$

答案是 C。