

學號：R07922004 系級：資工碩一 姓名：吳星耀

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

	(1) $p = 9 \times 18 + 1$	(2) $p = 9 \times 1 + 1$
RMSE (Private / Public)	7.27081 / 5.65650	7.22356 / 5.90263
RMSE (Average)	6.463655	6.563095

由表可見(1)在Public的表現較好，而(2)在Private的表現較好，由於我們無法得知Private跟Public的分別內容，因此我們將兩個測資平均起來討論其差別。平均結果可以發現(1)的預測結果比(2)還好，由於(1)比(2)多取了其他許多空氣相關feature，因此相較於(2)只用PM2.5去預測之後的PM2.5，(1)有更多相關數據可以參考，預測比較準是合理的。

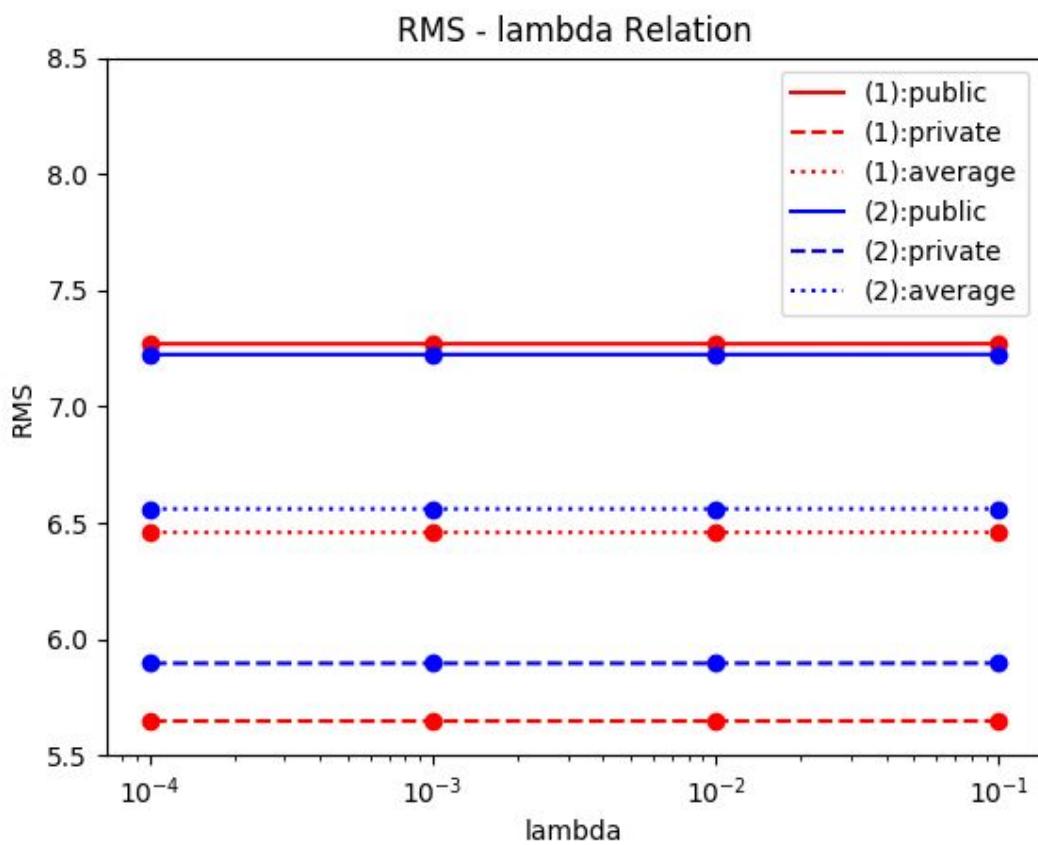
2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

	(1) $p = 9 \times 18 + 1$	(2) $p = 9 \times 1 + 1$
RMSE (Private / Public)	7.21542 / 5.96400	7.22464 / 6.22749
RMSE (Average)	6.58971	6.726065

與1.的原因相同，2.中(1)比(2)有更小誤差的原因也是因為取了更多空氣相關feature。而與1.比較，2.因為取的時間數較少，因此得到更少的feature，造成更多的誤差。同時因為1.的(2)比2.的(1)有更多的誤差，因此可以猜測少抽4小時比只取PM2.5的影響更大。

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

	(1) p = 9*18+1	(2) p = 9*1+1
$\lambda=0.1$ RMSE (Private / Public / Avg)	7.26989 / 5.64784 / 6.458865	7.22405 / 5.89631 / 6.56018
$\lambda=0.01$ RMSE (Private / Public / Avg)	7.26968 / 5.64749 / 6.458585	7.22358 / 5.89592 / 6.55975
$\lambda=0.001$ RMSE (Private / Public / Avg)	7.26966 / 5.64746 / 6.45856	7.22354 / 5.89589 / 6.559715
$\lambda=0.0001$ RMSE (Private / Public / Avg)	7.26966 / 5.64746 / 6.45856	7.22353 / 5.89588 / 6.559705



由圖與表可以看見 $\lambda$ 的大小對private, public與average的RMS幾乎沒有影響，推測因為所取的feature皆為一次feature，由於regularization是為了避免高次項所造成的overfitting，而(1)、(2)中的feature皆為一次項，因此並沒有帶來特別的好處。

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)