

# ML 2019 SPRING HW1

## B06902074 資工二 柯宏穎

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 score	private score
(1)All feature	5.66699	7.21391
(2)PM2.5	5.93022	7.24763

明顯地，只抓取pm2.5的來預測，因影響pm2.5的原因必定不只一個，我們不能純粹依照前9小時的pm2.5來training，會忽略掉過多可能的變因。

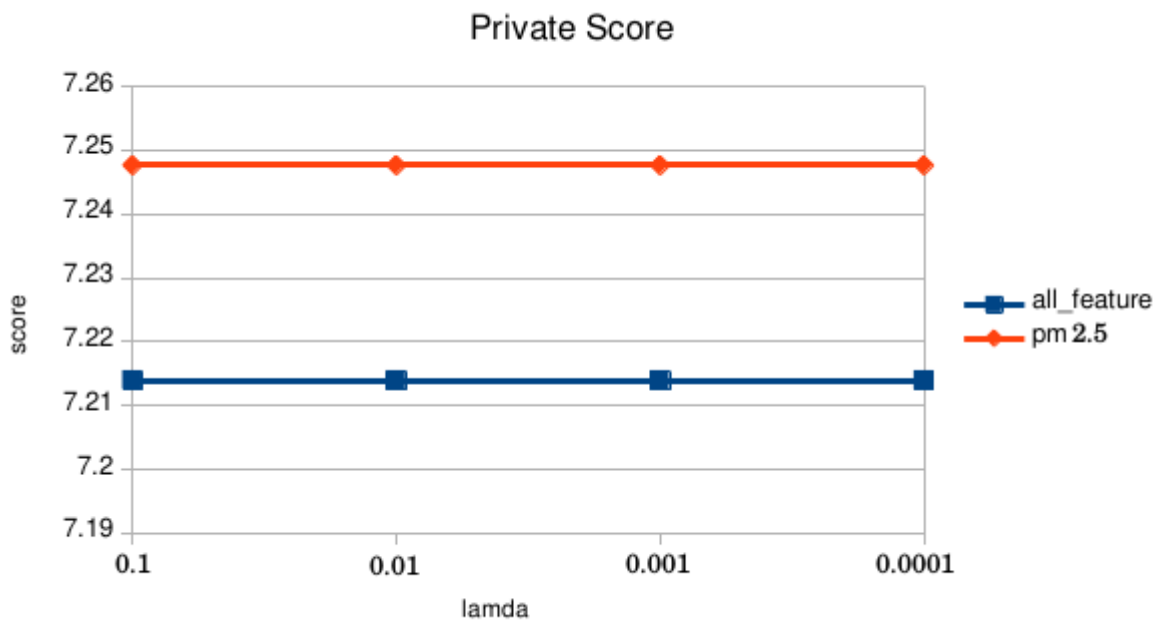
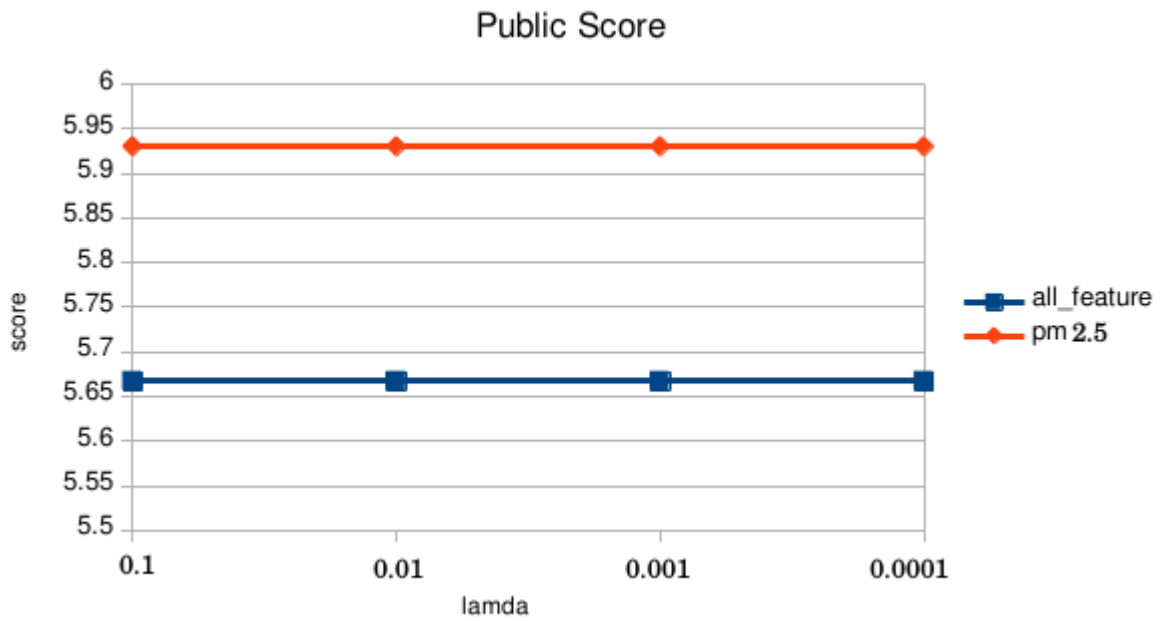
**2. (1%)**將feature從抽前9小時改成抽前5小時，討論其變化

	public score	private score
(1)All feature	5.98296	7.17309
(2)PM2.5	6.23692	7.24512

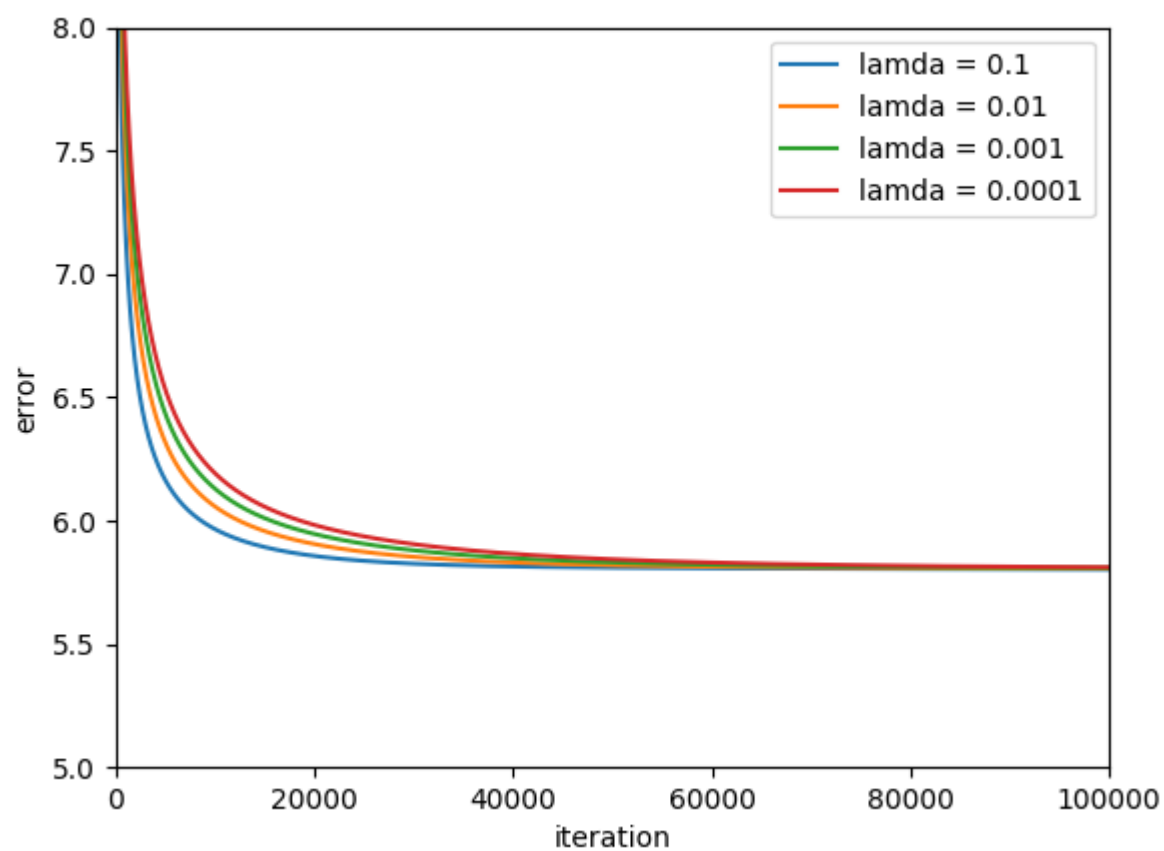
這次我取預測前5小時來做training，跟用前9小時的比較，差距不大。在public中，用比較長的時間去training，相對好比較多，這也比較符合想像：有更多的先備資料，能更準確地預測未來。private中則幾乎相同，較無討論空間。

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

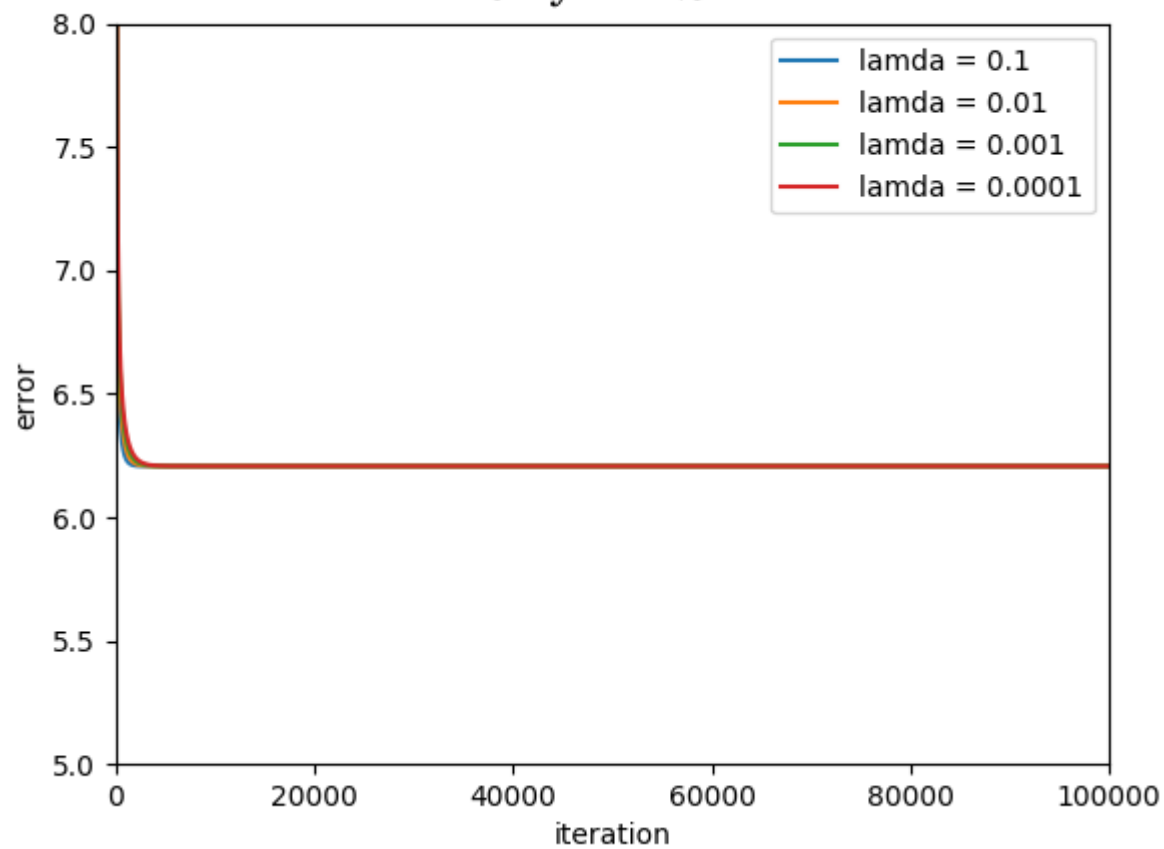
	$\lambda=0.1$	$\lambda=0.01$	$\lambda=0.001$	$\lambda=0.0001$
All feature(public)	5.66700	5.66699	5.66699	5.66699
PM2.5(public)	5.93022	5.93022	5.93022	5.93022
All feature(private)	7.21391	7.21391	7.21391	7.21391
PM2.5(private)	7.24762	7.24763	7.24763	7.24763



All feature



Only PM2.5



由多次training可發現，在本次的預測中，可能因為有加入 $adagrad$ 的關係， $\lambda$  極小時，並無太大的差別，因為越接近optimal點時，下降速度越慢，本來就不太容易過於接近極值點造成overfitting。有嘗試過將 $\lambda$ 調大，會造成underfitting，版上的score會變高。

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  $\sum_{n=1}^N (y^n - x^n * w)^2$ 。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)

Ans:(c)