

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的一次項(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

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 feature * 9hr	7.46251	5.33797
pm2.5 * 9hr	7.44013	5.62719

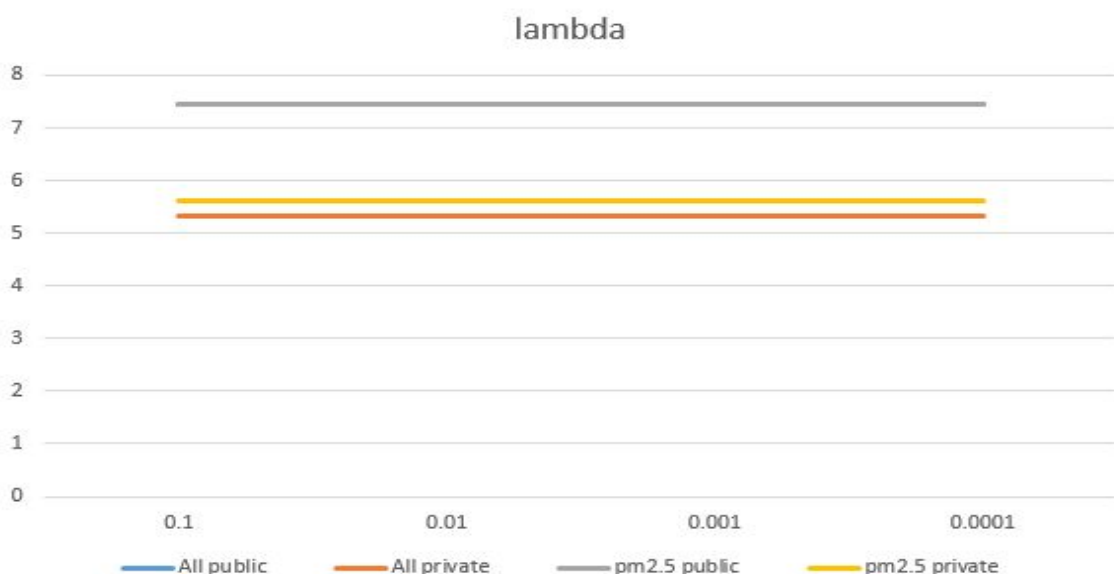
我們可以看到選取全部feature時在private中有比pm2.5更好的表現，但在public中僅小輸一點點，在自己train時，all feature的表現也比pm2.5來的好，整體看來，我會認為all feature 的表現比pm2.5好一些。

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 feature * 5hr	7.66522	5.35001
pm2.5 * 5hr	7.57904	5.79187

跟上面那題9小時的對照，整體的誤差值都變大了，在僅改變選取時間範圍的情況下，我們可以推論5個小時前到9個小時前的數值，對於pm2.5還是有一定影響力的。

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



折線圖中看不出來太大差別，實際數值也只在小數點後五位有正負一的誤差，或許是training次數不夠還沒到overfitting的程度，也可能是採用的loss function只有一次項沒有高次項，因此overfitting比較不容易發生。

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)為一存量  $y^n$ ，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  $w$ ？請寫下算式並選出正確答案。(其中  $X^T X$  為 invertible)

- (a)  $(X^T X) X^T y$
- (b)  $(X^T X)^0 X^T y$
- (c)  $(X^T X)^{-1} X^T y$
- (d)  $(X^T X)^{-2} X^T y$

Ans : (c)

令 feature 數為  $Z$ ， $X = N \times Z$ ， $W = Z \times 1$ ， $y = N \times 1$

$$\text{Loss} = \sum_{n=1}^N (y^n - X^n \cdot w)^2, \quad \hat{Z} X = CS(X)$$

$$\Leftrightarrow \|y - X \cdot W\|_{\min}$$

$$\Leftrightarrow \text{proj}_X(y) = X \cdot W$$

$$\Leftrightarrow \langle y - XW, \vec{x} \rangle = 0, \quad \forall \vec{x} \in X$$

$$\Leftrightarrow \langle y - XW, X\vec{b} \rangle = 0, \quad \forall \vec{b} \in Z \times 1$$

$$\Leftrightarrow (X\vec{b})^T (y - XW) = 0, \quad \forall \vec{b} \in Z \times 1$$

$$\Leftrightarrow \vec{b}^T X^T y - \vec{b}^T X^T XW = 0, \quad \forall \vec{b} \in Z \times 1$$

$$\Leftrightarrow \vec{b}^T (X^T y - X^T XW) = 0, \quad \forall \vec{b} \in Z \times 1$$

$$\Leftrightarrow X^T y - X^T XW = 0$$

$$\Leftrightarrow W = (X^T X)^{-1} X^T y$$