

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

RMSE	public	private
9hr all feature	7.46237	5.53562
9hr pm2.5	7.44013	5.62719

2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

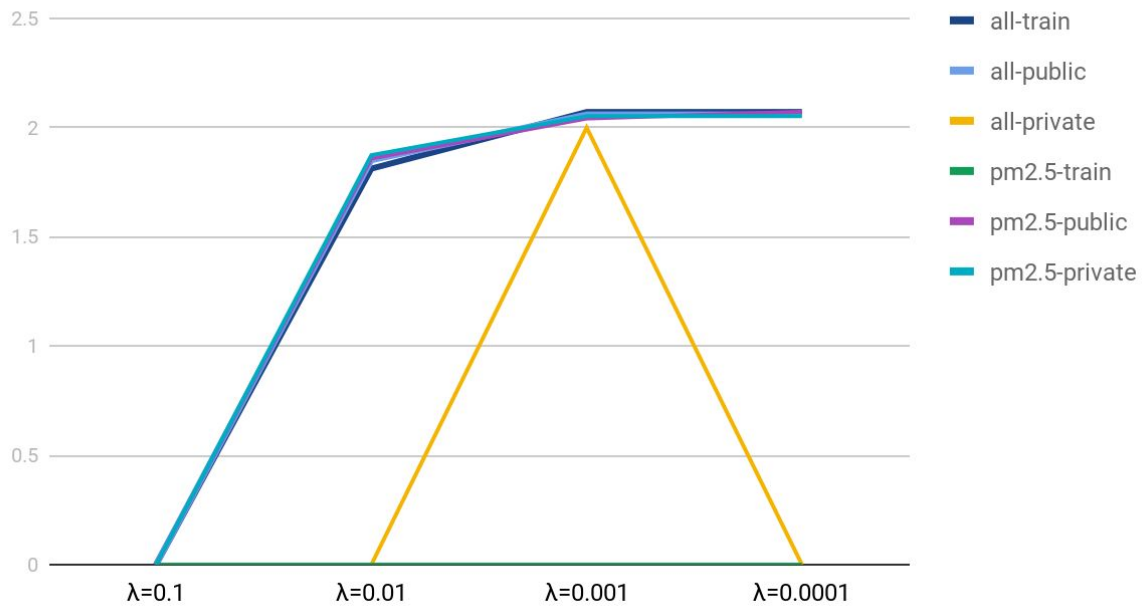
RMSE	public	private
5hr all feature	7.65925	5.44092
5hr pm2.5	7.57904	5.79187

從第一題和第二題中，public 和 private 的變化並不一樣，而且方均根差也有一定的差距，難以發現使用 all feature 還是單純 pm2.5 能得到比較好的模型，只能推測 public 和private的資料可能來自不同規則所選出的，不過大致上取較長的時間可以降低方均根差的值。

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

RMSE	$\lambda$	train	public	private
9hr all feature	0.1	5.682134	7.45556	5.40722
	0.01	5.682141	7.45618	5.40717
	0.001	5.682142	7.45625	7.45625
	0.0001	5.682142	7.45625	5.40716
9hr pm2.5	0.1	6.123022	7.43945	5.62696
	0.01	6.123022	7.44006	5.62717
	0.001	6.123022	7.44012	5.62719
	0.0001	6.123022	7.44013	5.62719

RMSE(normalized)



為了方便做在同一張圖，我把方均根差算完後再做處理，上圖中

$$y = (RMSE - \min(RMSE_i)) \div \text{標準差}$$

$\lambda$ 較大代表訓練出來的模型越平滑，在模型中發現  $\lambda$  太小時誤差就逐漸變大。

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

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

(d)  $(X^T X)^{-2} X^T y$

(c)

$$loss = \sum_{i=1}^N (y^n - x^n w)^2$$

minimize loss  $\rightarrow$  derivative = 0

$$\frac{\partial loss}{\partial w} = 2X^T (Xw - y) = 0$$

$$X^T X w = X^T y$$

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