

學號：B03705049 系級：資管四 姓名：廖寬璿

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

all	7.65941 + 5.52192
pm2.5	8.06277 + 5.42832

Train了同樣的次數後 pm2.5在 public比較差, 在private比較好但都差不多 由此可得知pm2.5以外的feature沒那麼重要

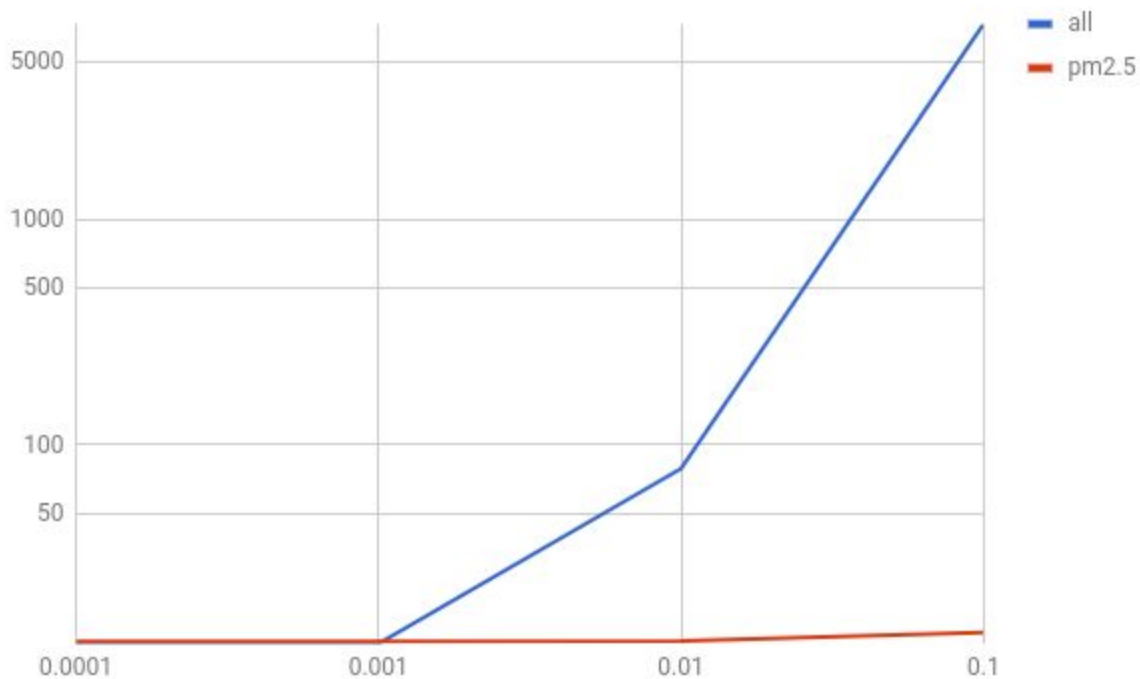
2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

all	7.89145 + 5.57374
pm2.5	8.00052 + 5.73149

Train了同樣的次數後 pm2.5在 public比較差, 在private比較好但都比9小時的差由此可得知前4小時的feature還是重要的

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

	0.1	0.01	0.001	0.0001
all	3612.03473 + 3677.38759	35.63916 + 43.00978	7.67517 + 5.52795	7.65941 +5.51513
pm2.5	8.93308 + 5.86017	8.08290 + 5.45588	8.06742 + 5.42985	8.06277 + 5.42832



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)

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

$$loss = \sum (Y - XW)^2$$

$$\frac{\partial loss}{\partial W} = 2(Y - XW)(-W) = 0$$

$$Y = XW$$

$$X^T Y = X^T XW$$

$$(X^T X)^{-1} X^T Y = W$$