

學號：R05631027 系級：生機碩二 姓名：楊皓文

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

Learning rate = 0.5, Repeat = 20000

由下表我們可以發現當有較多的污染源時，public 的成績會較高，然而，private 的部分卻差不多甚至更差一些，因此可以推論，其他污染因子不是全部都有正面的貢獻，也有雜訊的部分，反而會使得準確率較低

	Kaggle public	Kaggle private
9 小時內全部污染源	7.45884	5.44789
9 小時內 pm2.5	7.44015	5.62719

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

Learning rate = 0.5, Repeat = 20000

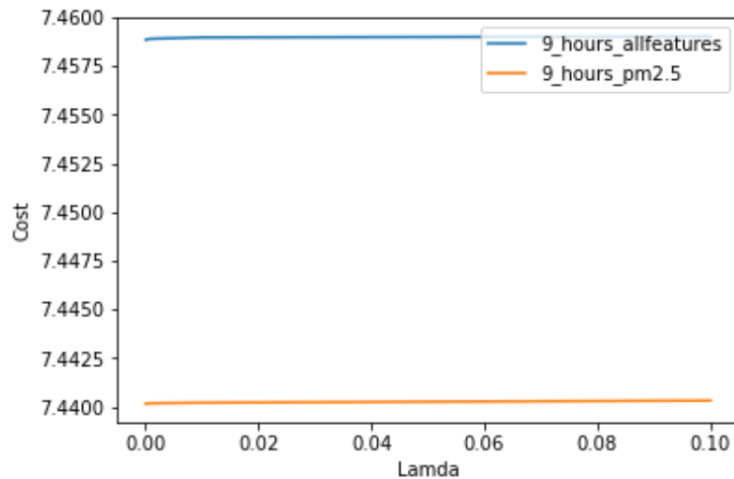
由下表我們可以看到 5 小時和 9 小時相比，平均來說 9 小時的準確率還是較高一些，5 小時又只有 pm2.5 的情況可能由於 feature 太少，使得 error 最高

	9 小時 public 分數	5 小時 public 分數
全部污染源	7.45884	7.64207
只有 pm2.5	7.44015	8.27832

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

由圖表可以看出在本次研究中，regularization 的影響並不大，甚至會有更大的 cost 出現

	Lambda=0.1	Lambda=0.01	Lambda=0.001	Lambda=0.0001
全部污染源	7.45902	7.45897	7.45890	7.45885
只有 pm2.5	7.44032	7.44021	7.44018	7.44016



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

$y = Xw$  帶入  $X^T y^T = X^T X X w$

所以  $w = (X^T X)^{-1} X^T y$

所以答案選 c