

學號：B03202006 系級：物理四 姓名：陳昱蓁

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
Ans:

	Public score	Private score	RMSE
全部 9 小時內的污染源(Model 1)	7.6218	5.37166	12.99346
全部 9 小時內 pm2.5 (Model 2)	7.44056	5.62746	13.06802

由表格可以觀察出，**Model 1** 的表現大致上較 **Model 2** 好，這是因為 **Model 1** 的 **feature** 維度比較高，從 **training set** 裡，**model** 可以學到較多東西。

但是兩者的差別是相近的。經過自己設不同的 **Model** 實驗，發現 **PM2.5** 本身就是一個很有用的 **feature**，所以單單用前九小時的 **PM2.5**，也可以跟用全部污染源做 **feature** 的模型媲美。

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

Ans:

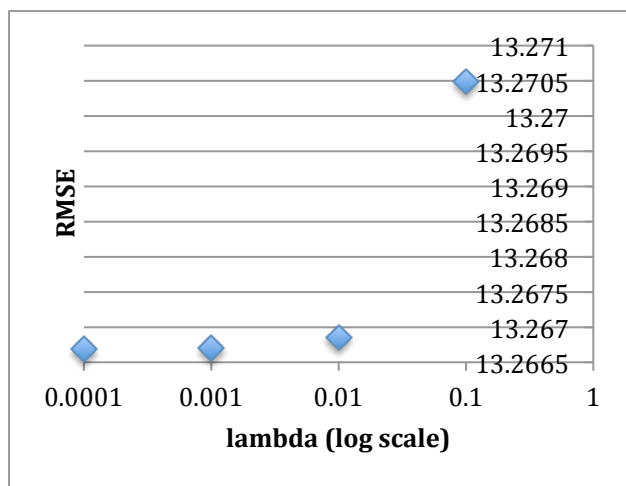
	Public score	Private score	RMSE
全部 5 小時內的污染源(Model 3)	22.31702	15.52603	37.84305
全部 5 小時內 pm2.5 (Model 4)	22.5667	16.73367	39.30037

跟第一小題比較，發現這兩個模型明顯都變差了，雖然 **training data** 變多，但是並不能彌補 **feature** 維度變少、模型過於簡化這個損失。

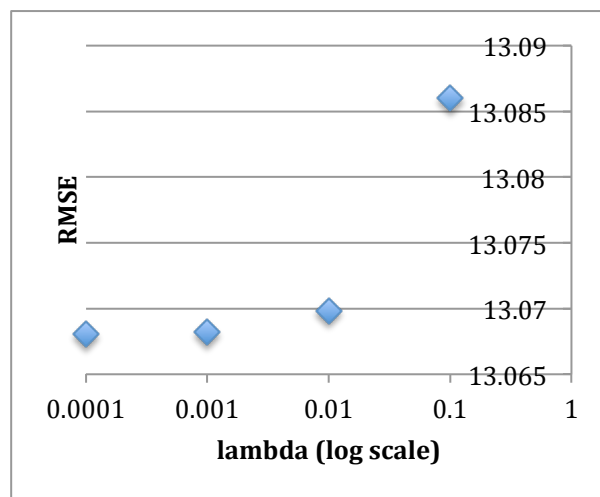
3. (1%)Regularization on all the weight with $\lambda=0.1$ 、 0.01 、 0.001 、 0.0001 ，並作圖

Ans:

	全部 9 小時內的污染源	全部 9 小時內 pm2.5
lambda = 0.1	13.27049	13.08602
lambda = 0.01	13.26685	13.06982
lambda = 0.001	13.26671	13.06819
lambda = 0.0001	13.26669	13.06804



All features with regularization



PM2.5 with regularization

(因為時間關係，所有污染源的模型並沒有 **train** 到最佳的狀態；第一小題的所有污染源模型 **iteration step = 18000**，本題所有污染源模型 **iteration step = 11000**)

加了 **reularization** 反而 **performance** 變差，可能是因為本來的 **Model** 就沒有 **overfit** 的情形，所以不需要加 **reularization**。

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

Ans: (c)

Loss function : $L = \sum_{n=1}^N (y^n - x^n \cdot w)^2 = \sum_{n=1}^N (y^n - \sum_i (x_i^n \cdot w_i))^2$

(將第 n 筆資料的第 i 個 **feature** 記為 x_i^n ，將第 i 個 **feature** 的 **weight** 記為 w_i)

欲最小化損失函數，亦即找到一組模型參數 w 使得

$$\frac{\partial L}{\partial w_j} = -2 \sum_{n=1}^N x_j^n \times (y^n - x^n \cdot w) = 0 \quad (\forall j)$$

此為一線性方程組，可整理成矩陣形式：

$$X^T (y - Xw) = 0$$

移項得

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