學號：R06922130 系級： 資工碩一 姓名：葉韋辰

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

以下表格為**scaling各測項、一次項、iteration = 10000、採adagrad**所得到的結果：

feature	public分數	private分數	RMSE
所有汙染源	7.46424	5.42173	6.523420689
僅PM2.5	7.44013	5.62719	6.596241419

- RMSE表現：「所有汙染源」之表現優於「僅PM2.5」，較多參數（162 features vs. 9 features）或許是影響二者表現的原因

2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

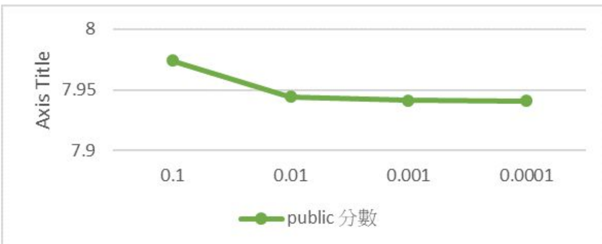
以下表格為**scaling各測項、一次項、iteration = 10000、採adagrad**所得到的結果：

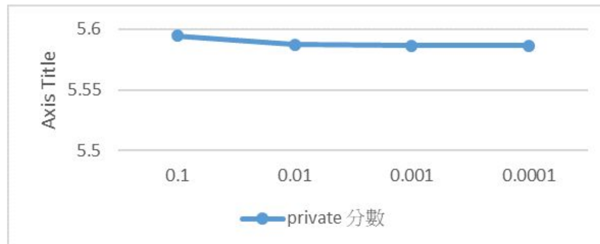
	public分數	private分數	RMSE
所有汙染源	7.77581	5.41925	6.701900842
僅PM2.5	7.57904	5.79188	6.744913686

- RMSE表現：如同9小時版本，在5小時情況下「所有汙染源」之表現仍優於「僅PM2.5」，較多參數（162 features vs. 9 features）或許是影響二者表現的原因

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

λ	0.1	0.01	0.001	0.0001
public分數	7.97418	7.94430	7.94131	7.94101
private分數	5.59466	5.58726	5.58654	5.58647
RMSE	6.887952061	6.867655233	6.865633026	6.865431045





從上述表格與圖表中，可判斷隨著lambda值從0.1一路趨近0.0001，regularization的結果愈來愈好，調整lambda值確實是重要議題之一

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 $X = [x^1\ x^2\ \ldots\ x^N]^T$ 表示，所有訓練資料的標註以向量 $y = [y^1\ y^2\ \ldots\ y^N]^T$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^TX為invertible)

(a) $(X^TX)X^Ty$
(b) $(X^TX)^{-0}X^Ty$
(c) $(X^TX)^{-1}X^Ty$
(d) $(X^TX)^{-2}X^Ty$

Ans: **(c)**，說明如下：

$$minimize\ \ ||y - Xw|| \Leftrightarrow Xw = proj_x y$$

$$\Leftrightarrow\ < y - Xw, b > \ = 0, \forall b \epsilon R(x)$$

$$\because b\epsilon R(x), let\ \ b = Xw'$$

$$\Leftrightarrow\ < y - Xw, Xw' > \ = 0$$

$$\Leftrightarrow\ < y, Xw' > \ - < Xw, Xw' > \ = 0 \ \ \Leftrightarrow\ < y, Xw' > \ = \ < Xw, Xw' >$$

$$\Leftrightarrow (Xw')^T y \ = \ (Xw')^T Xw \ \ \Leftrightarrow \ \ (w')^T X^T y = (w')^T X^T Xw$$

$$\Leftrightarrow < X^T y, w' > \ = \ < X^T Xw, w' > \ \Leftrightarrow \ \ X^T y = X^T Xw$$

$$\because X^TX\ \ is\ \ invertible\ \ \ \therefore w = (X^TX)^{-1}X^Ty$$