學號：R06921005 系級： 電機碩一 姓名：陳昱文

請實做以下兩種不同 feature 的模型，回答第 (1) ~ (3) 題：
抽全部 9 小時內的污染源 feature 的一次項(加 bias)
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

A：

RMSE	public	private	private + public
9hr + 全	7.97087	5.60933	6.89200
9hr + PM2.5	7.39590	5.83648	6.66198

在 public 看起來會有 0.6 幅度的進步 然而實際上在 private 卻是 0.2 幅度的退步，一方面源自資料本身的偏差，另一方面或許表示刪掉過多的 feature，除了 PM2.5 上存在有幫助的 feature

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

A：

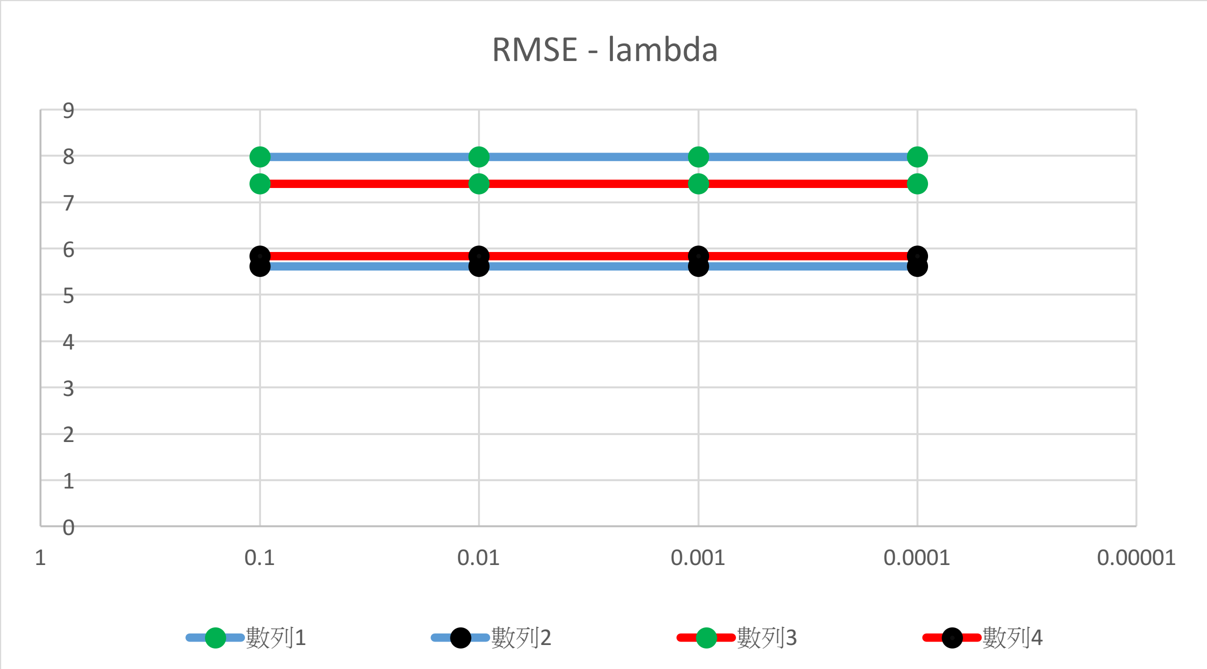
RMSE	public	private	private + public
5hr + 全	7.84647	5.38140	6.72780
5hr + PM2.5	7.53030	5.88723	6.75888

在 public 和 private 有著不一樣的改變情況，可能源自於資料本身的偏差，關於時間更加的改善方式或許是對各個 feature 做客製化的時間長度資料選取

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

A：

RMSE	λ	public	private	all
(1)	0.1	7.97087	5.60933	6.89200
(1)	0.01	7.97087	5.60933	6.89200
(1)	0.001	7.97087	5.60933	6.89200
(1)	0.0001	7.97087	5.60933	6.89200
(2)	0.1	7.39591	5.83649	6.66199
(2)	0.01	7.39590	5.83648	6.66198
(2)	0.001	7.39590	5.83648	6.66198
(2)	0.0001	7.39590	5.83648	6.66198



橫軸：λ (log scale)；縱軸：RMSE (linear scale)

藍線為(1)情況下各 λ值之 RMSE；紅線為(2)情況下各 λ值之 RMSE

綠點為 public data；黑點為 private data

可以發現幾乎全為水平線，也就是說 regularize 完全沒有起到效果，估計是因為值取太小，相較於 object function (loss) 小到足以忽略，若將值取大將能看到 regularize term 對 training 的影響

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 xn，其標註(label)為一存量 yn，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 = [x1 x2 … xN]T 表示，所有訓練資料的標註以向量 y = [y1 y2 … yN]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 XTX 為 invertible)

(XTX)XTy

(XTX)-0XTy

(XTX)-1XTy

(XTX)-2XTy

A：(c)

我們想要最小化的函數可等效為：min $\|y - Xw\|$ ，可以線性代數解之

存在w_0使得 $Y - Xw_0 \in \mathrm{Null}(X)$，也就是說 $< Xw , Y - Xw_0 > = 0$ 對於所有 w，

由內積性質推得 $< w , X^*(Y - Xw_0) > = 0$ 對於所有 w，

故推得 $X^*(Y - Xw_0) = 0$，所以 $w_0 = (X^*X)^{-1}X^*y$，因此得答案為(c)