學號：R05546003 系級： 工工碩二 姓名：謝凱靜

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

		public	private	RMSE
9 hr	all features	7.46631	5.30105	6.474833
	only pm2.5	7.44013	5.62719	6.596241

只考慮 PM2.5 這個特徵值的時候 Public Score 高於使用所有的特徵值，但在 Private Score 使用所有的特徵值高於只考慮 PM2.5 這個特徵值，兩者一起看算出 RMSE ，只考慮 PM2.5 這個特徵值高於使用所有的特徵值，推測仍有些變數還是有用的，所以只考慮 PM2.5 這個特徵值的 RMSE 還是輸給了使用所有特徵值。

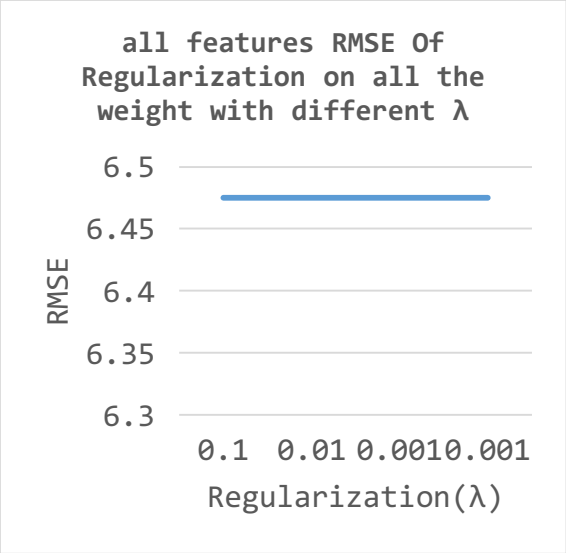
2. (1%)將 feature 從抽前 9 小時改成抽前 5 小時，討論其變化

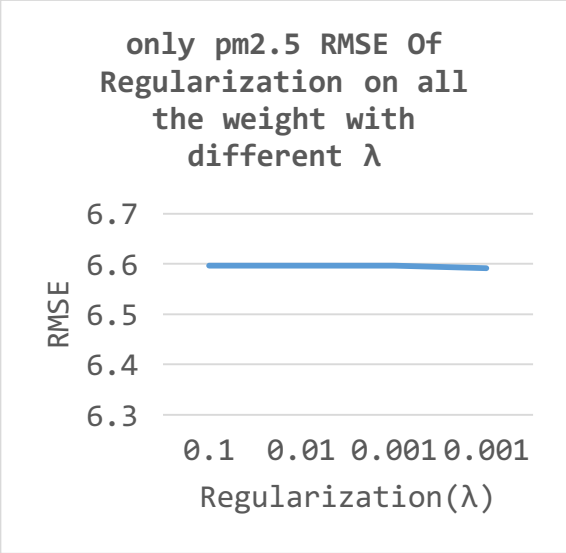
		public	private	RMSE
5 hr	all features	7.66477	5.3299	6.601384
	only pm2.5	7.57904	5.79187	6.744909

可以從表格中的數字觀察，不管是 Public Score, Private Score, RMSE,皆沒有比抽取前九個小時的好，表示考量比較多的時間，對於預測 PM2.5 來說比較有解釋力。

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

		Regularization(λ)	public	private	RMSE
9 hr	all features	0.1	7.46632	5.30104	6.474834
		0.01	7.46631	5.30104	6.474829
		0.001	7.46631	5.30104	6.474829
		0.001	7.46631	5.30105	6.474833
	only pm2.5	0.1	7.44011	5.62721	6.596239
		0.01	7.44013	5.62719	6.596241
		0.001	7.44013	5.62719	6.596241
		0.001	7.39719	5.6734	6.591885





從上面表格中的數字，以及所作的圖，可以看到幾乎是一條線，可能是因為沒有做 Feature Scaling 導致 Regularization 並沒有很明顯，並沒有隨著 λ 的改變有所改變

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 $\mathbf{x}^n$，其標註(label)為一存量 $\mathbf{y}^n$，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 $\mathbf{X} = [\mathbf{x}^1\ x^2 \dots x^N]^T$ 表示，所有訓練資料的標註以向量 $\mathbf{y} = [\mathbf{y}^1\ y^2 \dots y^N]^T$ 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 $\mathbf{X}^T\mathbf{X}$ 為 invertible)

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

$$\sum_{n=1}^{N}(y^n - x^n \cdot w)^2$$

$$= y^{n\prime}y^n - y^{n\prime}x^nw - w'x^{n\prime}y^n + w'x^{n\prime}x^nw$$

$$= y^{n\prime}y^n - 2w'x^{n\prime}y^n + w'x^{n\prime}x^nw$$

$$let\ \frac{\partial(L)}{\partial w} = -2x^{n\prime}y^n + 2x^{n\prime}x^nw = 0$$

$$=> (x^{n\prime}x^n)w = x^{n\prime}y^n => w = (x^{n\prime}x^n)^{-1}x^{n\prime}y^n$$