學號：R06942128 系級： 電信碩一 姓名：許祐銘

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

RMSE	Private Score	Public Score	Public + Private
All feature	5.28983	7.48251	6.47959
Only PM2.5	5.62719	7.44013	6.59624

(with learning rate = 100, iteration = 50000)

討論：雖然只抽PM2.5在Public的RMSE較低，但是對於Private或是整體來說，抽取全部 feature 的RMSE 相對比較低，代表除了PM2.5之外的 feature 還是會影響到 PM2.5

2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

RMSE	Private Score	Public Score	Public + Private
All feature	5.32875	7.66521	6.60117
Only PM2.5	5.79187	7.57904	6.74491

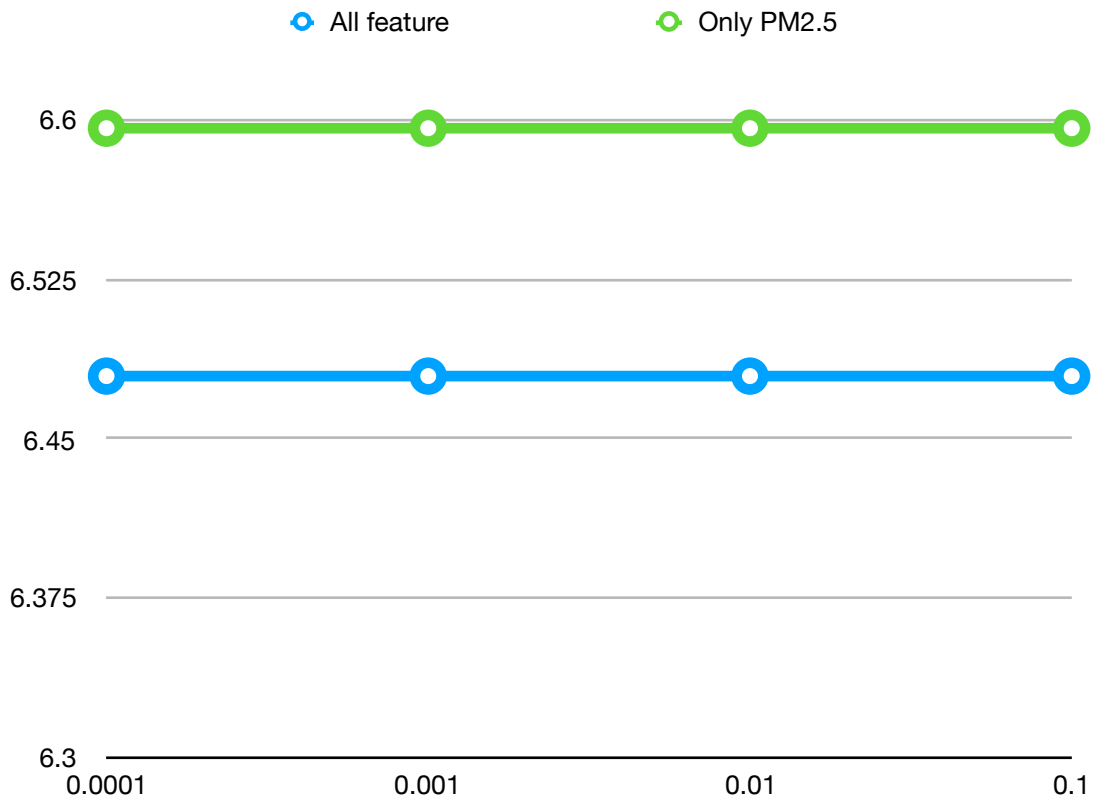
(with learning rate = 100, iteration = 50000)

討論：很明顯的，跟抽9小時相比的話，RMSE無論是Public或是Private都是變高，這也代表說抽取9小時的 feature 對 traning 是比較有幫助的。

3. (1%)Regularization on all the weight with λ=0.1、0.01、0.001、0.0001，並作圖顯示訓練後的差別(with RMSE)

RMSE (240 data)	*λ = 0.1*	*λ = 0.01*	*λ = 0.001*	*λ = 0.0001*
All feature	6.47958	6.47958	6.47958	6.47958
Only PM2.5	6.59624	6.59624	6.59624	6.59624

(with learning rate = 50, iteration = 50000)



對於題目的四個 λ ，我的model RMSE幾乎沒有變化，但是如果增加 λ 的話會如下表格，以All feature來看，可以看得出來 λ 越大，RMSE會偏向增加，應該是函數過於平滑了。

RMSE (240 data)	$\lambda = 1$	$\lambda = 10$	$\lambda = 100$	$\lambda = 1000$
All feature	6.47959	6.47956	6.47975	6.48716

(with learning rate = 50, iteration = 50000)

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)為一純量 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 $\sum_{i=1}^{n}(y^n - x^n \bullet w)^2$。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^TX為invertible)

$$L = \sum_{i=1}^{n} (y^n - x^n \bullet w)^2 = (Y - Xw)^T (Y - Xw)$$

$$= (Y^T - (Xw)^T)(Y - Xw)$$

$$= Y^T Y - Y^T Xw - (Xw)^T Y + (Xw)^T (Xw)$$

$$= w^T X^T Xw - 2(Xw)^T Y + Y^T Y$$

$$\frac{\partial L}{\partial w} = 2X^T Xw - 2X^T Y = 0$$

$$2X^T Xw = 2X^T Y$$

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