學號：R05922164 系級： 資工碩一 姓名：柯百嶽

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

以下為train 50000次後的結果，lamda皆是3，learningrate固定

	private	public
全部9小時內的污染源feature的一次項(加bias)	5.30345	8.08463
抽全部9小時內pm2.5的一次項當作feature(加bias)	6.22831	7.43017

討論：

首先取得traindata從每個月連續20小時取不同的十小時當作訓練資料，因此上面兩種都有(20*24-9)*12= 5652，但(1)的每一筆資料有9*18筆+bias共163個參數，而(2)是9*1+bias共10個參數。

看參數就決定了訓練的速度，此外選取更多feature，在kaggle上的表現上public，只有PM2.5比較好一點，但在private全選比較好。

個人認為PM2.5可能不只跟自己相關，所以全選的話在真正預測上會比較準確一點。

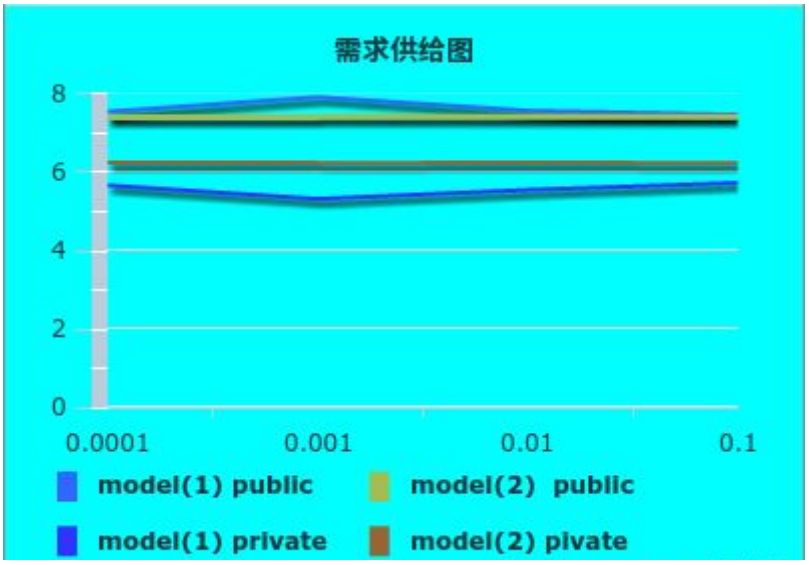
2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

	private	public
全部9小時內的污染源feature的一次項(加bias)	5.52268	7.66656
抽全部9小時內pm2.5的一次項當作feature(加bias)	6.17755	7.54457

討論：
不管是哪個模型，只選取5個小時都是比較差的，可以看出考慮更長的連續資料是有用的，但要選多少，相關程度可能就是要嘗試後才能知道。
因此我最後使用的還是選取9個小時，然後選了大概10個feature當作train來用。

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

名称	model(1) public	model(2) public	model(1) private	model(2) pivate
0.0001	7.54866	7.42851	5.66739	6.23295
0.001	7.92715	7.42962	5.31792	6.23126
0.01	7.57231	7.40663	5.55184	6.25205
0.1	7.47579	7.42483	5.73512	6.22931



感覺上Regularization對一次式沒什麼影響的感覺，沒有特別的走向。

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^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

令loss function偏微分等於0得到$-2X^T(y-Xw) = 0$

$===> -2X^Ty+2X^TXw=0 ===> X^Ty = X^TXw ===> w =(X^TX)^{-1}X^Ty$