學號：B04901025 系級： 電機三 姓名：陳鴻智

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

```
cost_sqrt:  6.50423323304
```

```
cost_sqrt:  6.5962445333
```

一開始我是用(1)的model, 可是似乎在public set的成績不太好，後來改用第二種才改善。 但用正確答案算出rmse，發現其實第一種model好像error比較小，可能是pm2.5相對來說data比較少的關係。

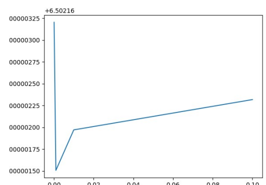
2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

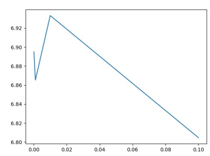
```
cost_sqrt:  26.4400626784
```

```
cost_sqrt:  26.7351554608
```

改成抽前五小時，雖然在training的error可以降到3，但是出來的結果並不好，應該是發生overfitting

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





C

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) $(X^TX)^{-2}X^Ty$

let $J(b) = \sum_{i=1}^{N} (y_i - x_i^T b)^2$

$J(b) = (y - Xb)^T (y - Xb)$

$= (y^T - (Xb)^T)(y - Xb)$

$= (Xb)^T(Xb) - y^T(Xb) - (Xb)^T y + y^T y$

$= b^T X^T X b - 2(Xb)^T y + y^T y$

let $\frac{\partial J}{\partial b} = 0 \Rightarrow 2X^T X b - 2X^T y = 0$

$\Rightarrow X^T X b - X^T y = 0 \Rightarrow b = (X^TX)^{-1}X^Ty$ #