學號：B03901030 系級： 電機三 姓名：蕭晨豪

1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)

答：

在hw1\_best中我取前九個小時的PM2.5一次項以及前兩小時的PM2.5的平方項作為feature

2.請作圖比較不同訓練資料量對於PM2.5預測準確率的影響

答：

以下三筆資料均使用stochastic training, iteration 60萬次，在最後1000次時不使用stochastic而使用所有的值做一般的gradient descent

model 1:十二個月中每個月隨機取五天

model 2:十二個月中每個月隨機取十天

model 3:十二個月中每個月二十天全取

在此方法下越多的訓練資料量達到越好的訓練效果

3. 請比較不同複雜度的模型對於PM2.5預測準確率的影響

答：

我主要比較了三種:

model 1:取前九個小時的PM2.5一次項當作feature

model 2: 取前九個小時的PM2.5一次項以及前一小時的PM2.5的平方項當作feature

model 3: 取前九個小時的PM2.5一次項以及前兩小時的PM2.5的平方項當作feature

此三種model的training方式同第2題，可看出越複雜的model有越好的效果

4. 請討論正規化(regularization)對於PM2.5預測準確率的影響

答：

當我的feature取前九個小時的PM2.5一次項時，不論我的lambda(regularization參數)取多少performance都沒有進步(大部分都比沒有regularization差一點點)

但當我的feature取前九個小時除了rainfall以外的所有參數的一次項時，regularization將原本的RMSE從6.11286降至6.1009(取lambda = 0.05)

我認為這是因為regularization主要是加一個penalty在取過多的參數以及過高的weight上面，當我取很多參數時regularization就會顯現其效果，可是當我只取前九個小時的PM2.5一次項時，可能因為參數已經夠少了，regularization並不會讓效果變好

5. 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 ，其標註(label)為一純量，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 。若將所有訓練資料的特徵值以矩陣 X =表示，所有訓練資料的標註以向量 y = 表示，請以 X 和 y 表示可以最小化損失函數的向量 w 。

答：

在取出特徵後我們的預測結果為，因此我們想找到w使最小(也就是讓我們定義的loss function最小)，將loss function對w做偏微分後得到，使其為零得到