学号：11849325  专业：计算机技术 姓名：丁刘辉

请考虑以下2种不同feature的模型，回答第 (1) ~ (3) 題：

1. 抽全部9小時內的污染源feature当作一次項(加bias)  
   9x18+1=163種feature（9小時內所有18種測量值+bias項）
2. 抽全部9小時內pm2.5的一次項當作feature(加bias)  
   9x1+1=10種feature（9小時內所有PM2.5值+bias項）

备注 :   
      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

     c. 第1-3題請都以題目給訂的兩種model來回答

1. (2%)記錄誤差值 (RMSE)(根據 public+private分數)，討論兩種feature的影響

本数据集的数据属性和带预测数据具有比较大的线性相关性。

在做本一问时，暂时未使用偏置项，使用的是最基本的梯度下降方法，学习率固定，参数设置例子如下：



Alpha即为学习率，eps作为迭代退出条件，iteration是最大迭代次数。

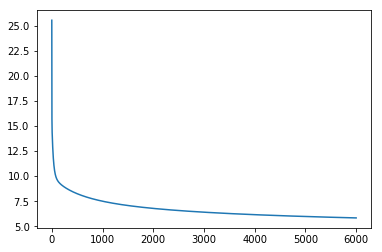
一次项系数向量w从0点出发，w0=[0,0,0...,0]

(1)在使用全部污染源属性作为一次项的情况下，结果如下：

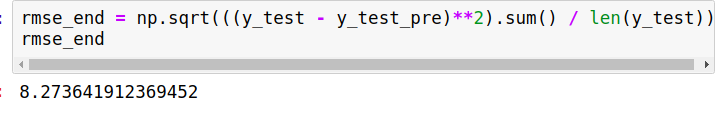
参数：



RMSE变化曲线



使用测试集进行预测，最终的RMSE值：

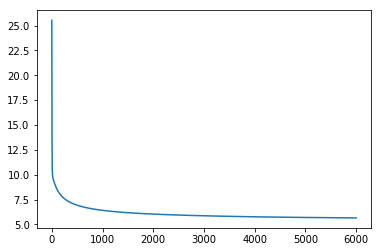


(2)在使用单PM2.5属性作为一次项的情况下，结果如下：

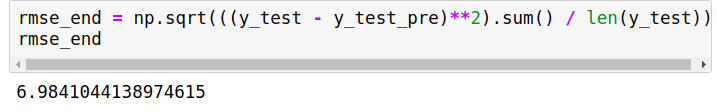
参数：



RMSE变化曲线：



使用测试集进行预测，最终的RMSE值：



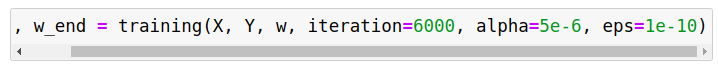
分析：

由于线性相关性较强，两种feature选择得到的结果差异不大。由训练情况来看，只用10种属性作为feature效果更好，不过这个主要是因为使用更少的feature可以接受更大的学习率，收敛可能更快。更多的feature意味着（初始）学习率必须设置更小，不然很容易越过optimal value。

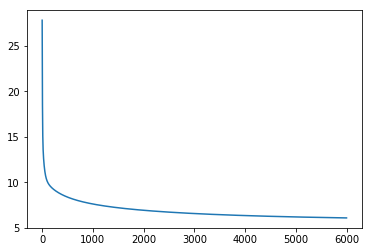
2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

(1)在使用全部污染源属性作为一次项的情况下，结果如下：

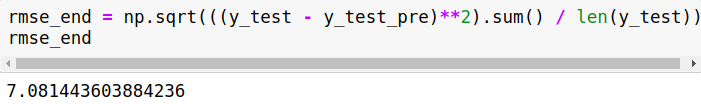
参数：



RMSE变化曲线：

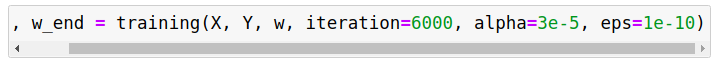


使用测试集进行预测，最终的RMSE值：

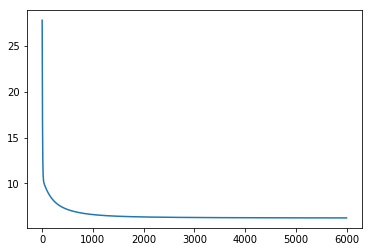


(2)在使用单PM2.5属性作为一次项的情况下，结果如下：

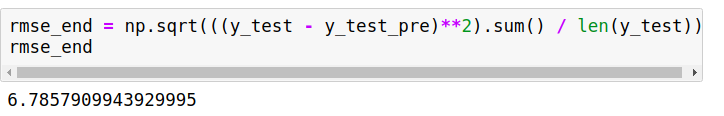
参数：



RMSE变化曲线：



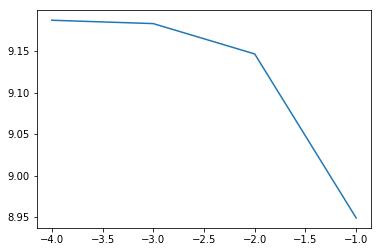
使用测试集进行预测，最终的RMSE值：



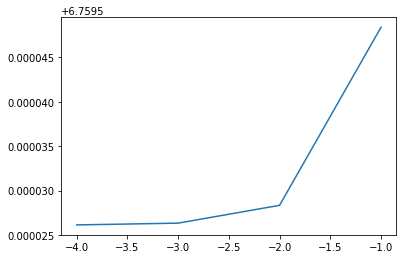
分析：同上一问一样，使用更少的参数可以接受更大的学习率，收敛更快。与第一题对比，只用5个小时的数据进行预测效果稍微好于用10个小时，单总体而言差别不大。

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

(1)在使用全部污染源属性作为一次项的情况下，结果如下：



(2)在使用单PM2.5属性作为一次项的情况下，结果如下：



横坐标使用的是log10(lambda)，从上图可以看出，使用更多的属性，引入正则项，预测效果变好，而更少的属性预测效果变差，但是从图二的纵坐标尺度可以看出，影响比较小。

总体而言引入正则项后预测效果有所增加，可能在一些数据分布更为明显的数据集上效果会更加明显。