线性回归算法

## 一、实验内容：

利用线性回归完成链接https://archive.ics.uci.edu/ml/datasets/Appliances+energy+prediction的任务

## 二、实验任务：

根据住宅各个区域内记录的温度，湿度对住宅整体（各耗电应用）的一个能耗情况进行预测。数据中的第B列就是我们所要预测的值，将C列到AA列作为特征值。暂时忽略最后两列的随机变量。

对数据进行分析：第C列数据为房屋内照明装置的能源使用。第D列到第U列分别为厨房，客厅，洗衣间，办公室，浴室，房屋外面（北侧），熨烫区，青少年活动中心，父母房间的温度（）和湿度（%）。第V,W,X,Y,Z,AA列分别为气象站测得的外部温度（），外部压力（mmHg），外部湿度（%），风速（m/s），能见度（km），结露点（）等天气特征。一共为25个属性特征。

此时，我们的任务可以用下列式子表示：

## 三、实验算法：

**线性回归模型是一个通过属性进行线性组合来进行预测的函数**

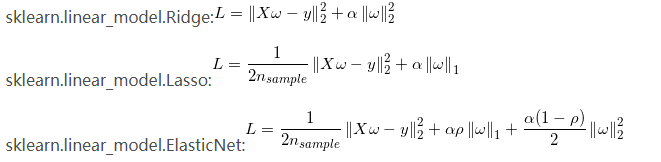
## 四、实验数据划分：

**一共有19735条数据，对数据进行随机划分，取2/3（13156条）的数据集作为训练集，1/3（6579条）的数据集作为测试集来评估模型性能。在划分过程中没有关注数据的分布是否一致的问题，因为数据量满足模型训练学习需求，同时考虑算法的计算开销，故未采用交叉验证法及自助法对数据进行划分。2/3只是一个经验划分方法。**

## 五、实验结果：

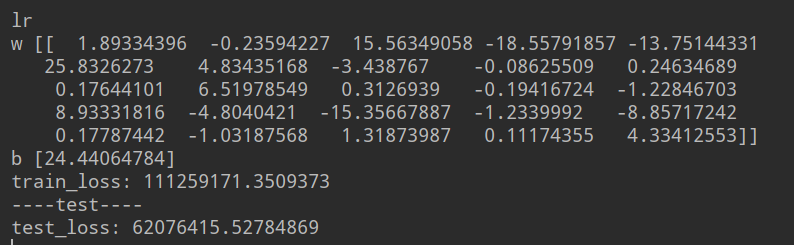
根据损失函数设置的不同，选择了如下四种线性回归模型

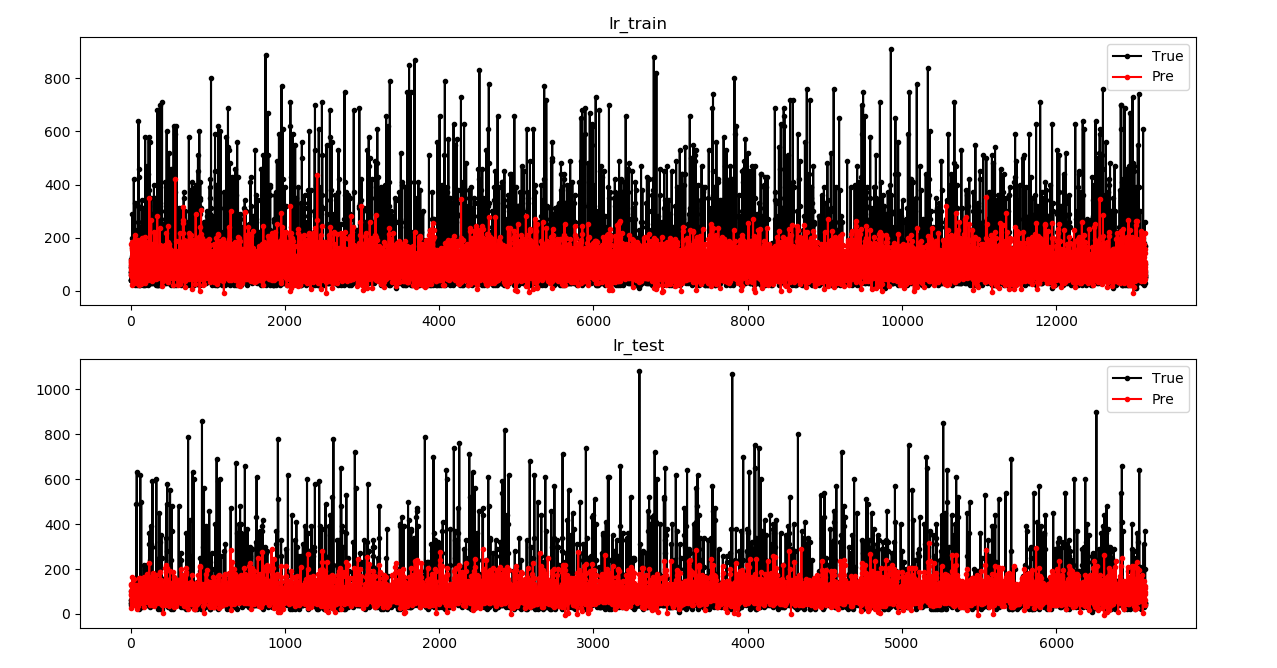
Sklearn.linear\_nodel.LinearRegression:



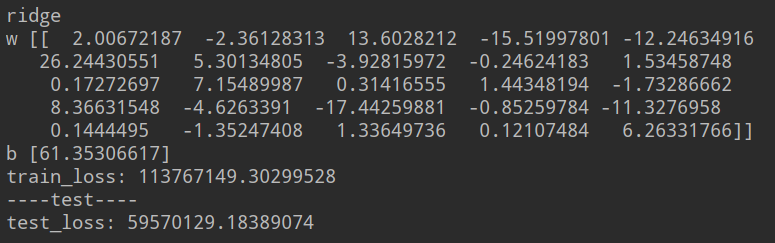
### 未对数据进行预处理

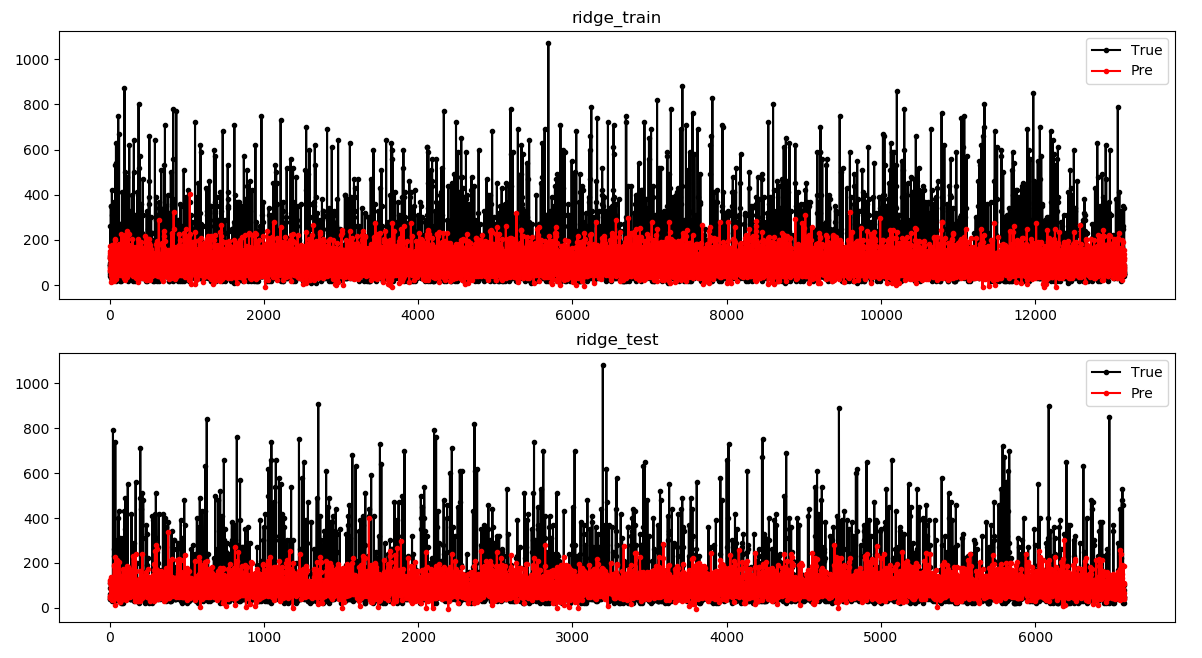
1、简单线性回归，损失函数为均方误差（平方损失），其实就等同于平方和误差函数。



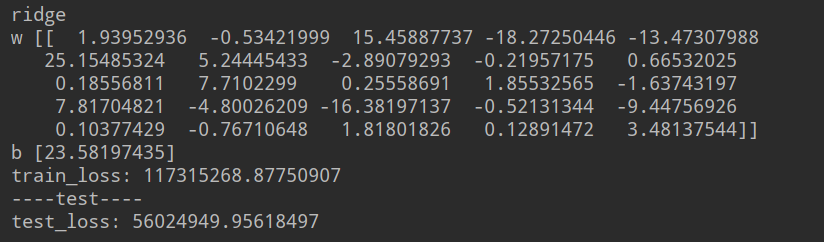


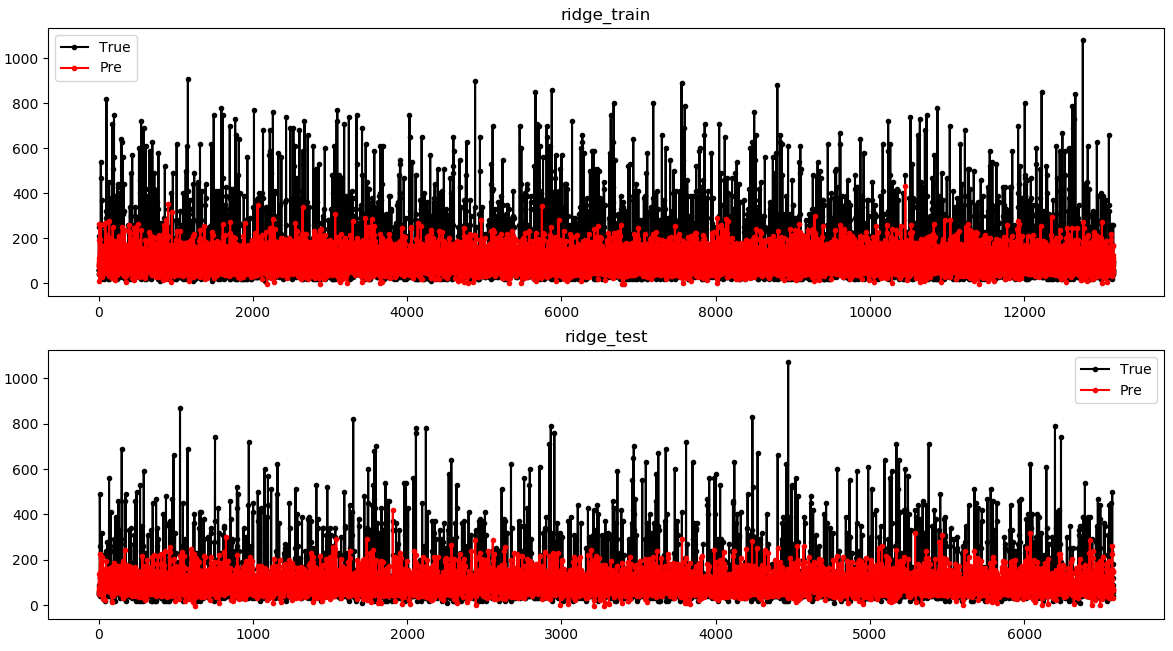
2、加了L2正则项的损失函数，参数alpha=1.0，控制参数w大小





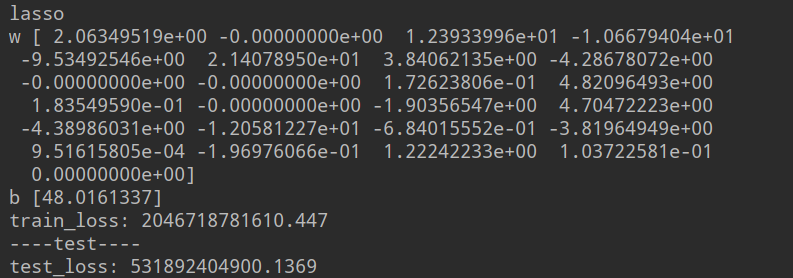
增大参数alpha，alpha设置为10，w参数减少，但是训练的模型效果依然很一般

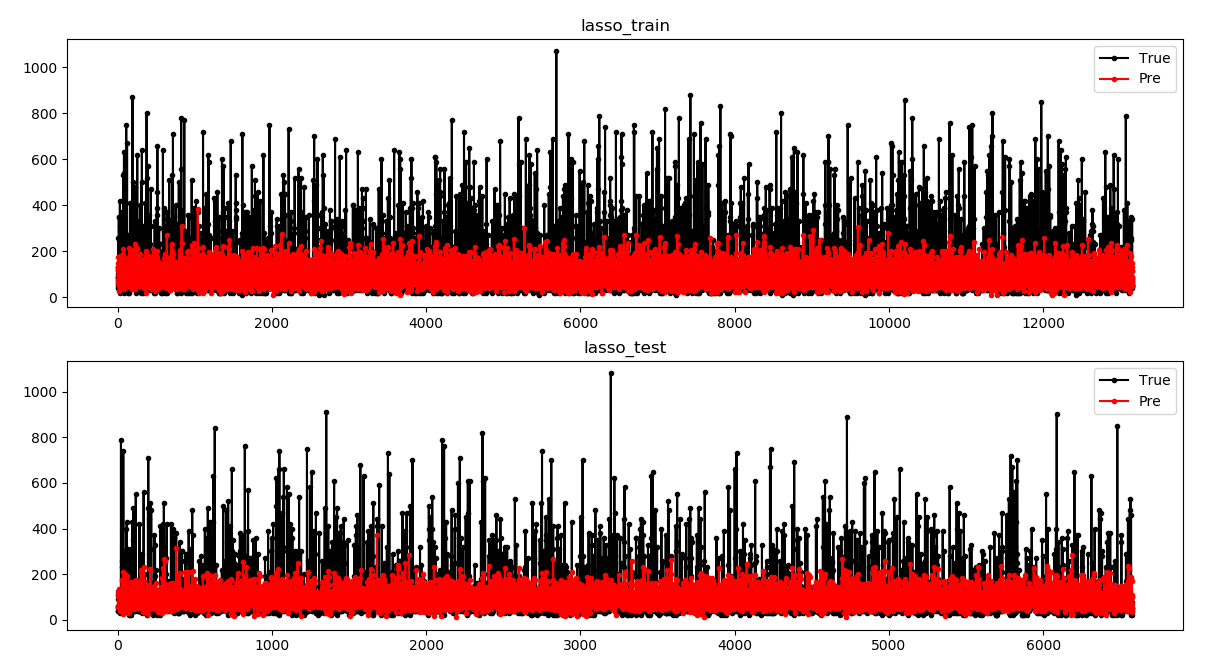




1. lasso

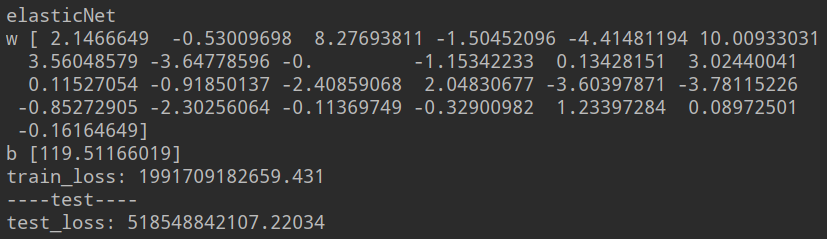
对损失函数加了L1正则项，alpha设置为1

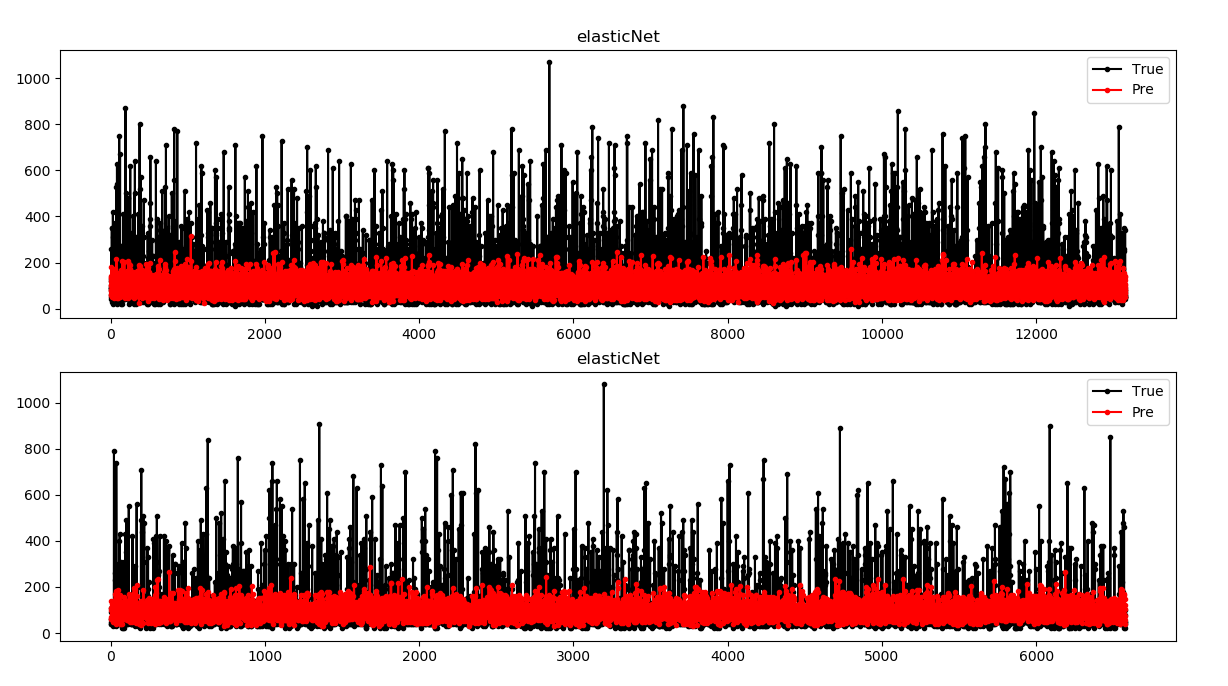




增大参数alpha，alpha设置为10，w参数值减少，但是训练的模型效果依然很一般

4、elasticNet



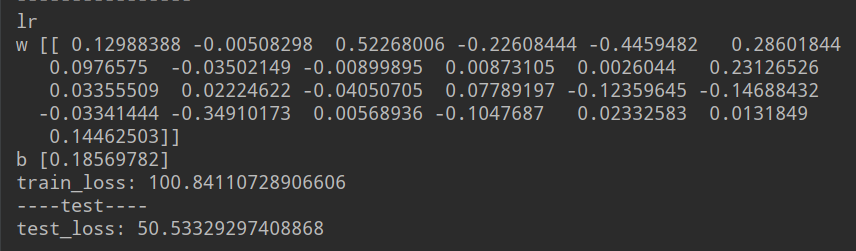


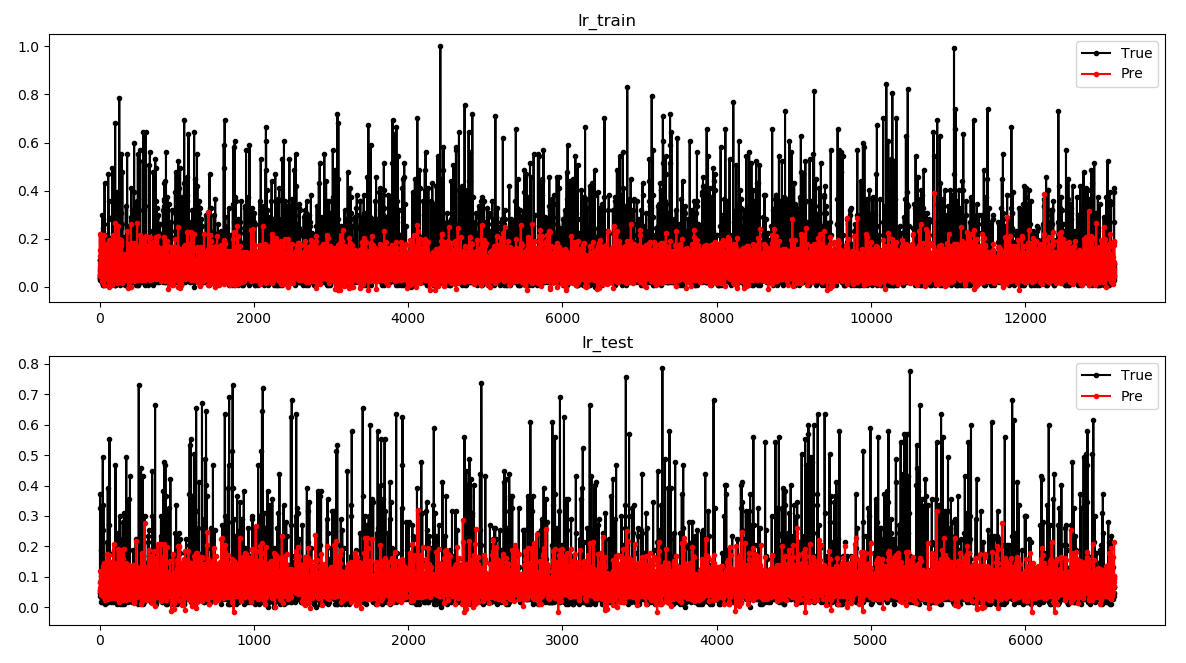
### 2、对数据进行归一化

对X和Y都采用了min-max归一化，其余参数保持不变

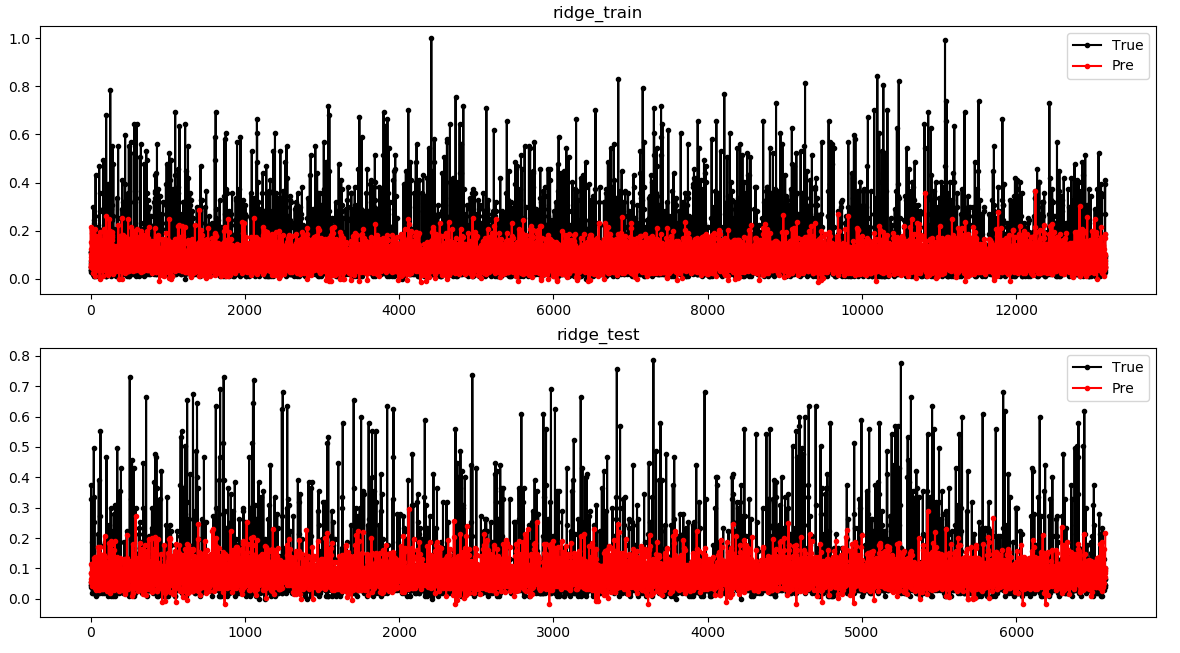


#### 1、简单线性回归，损失函数为均方误差（平方损失）

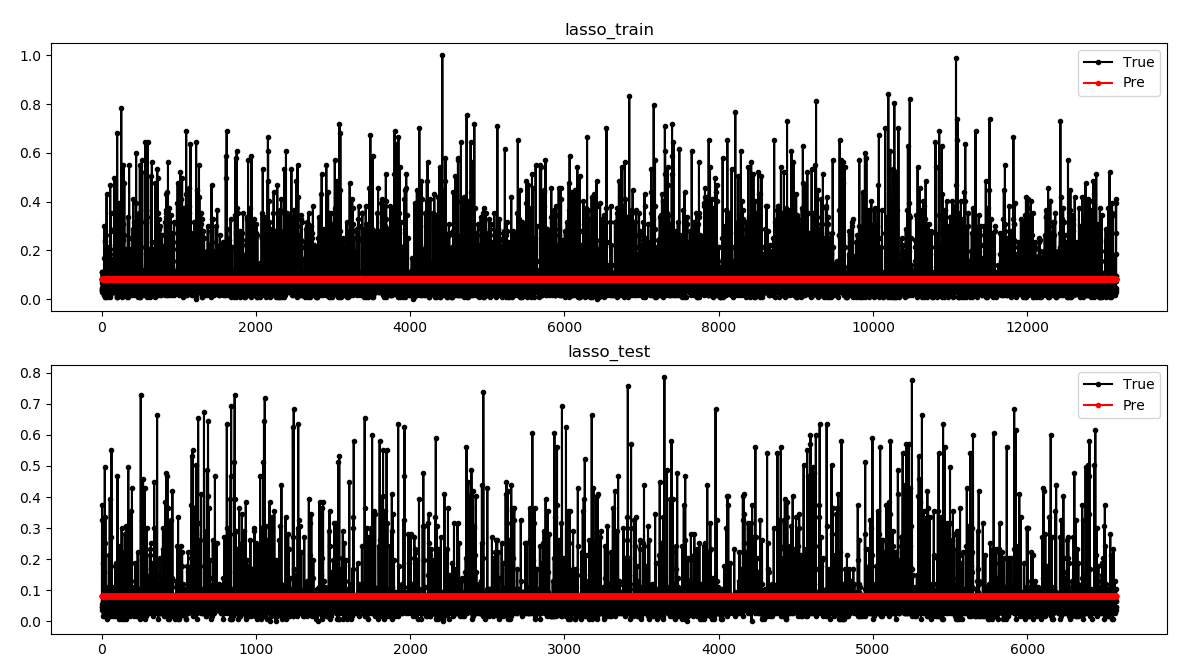




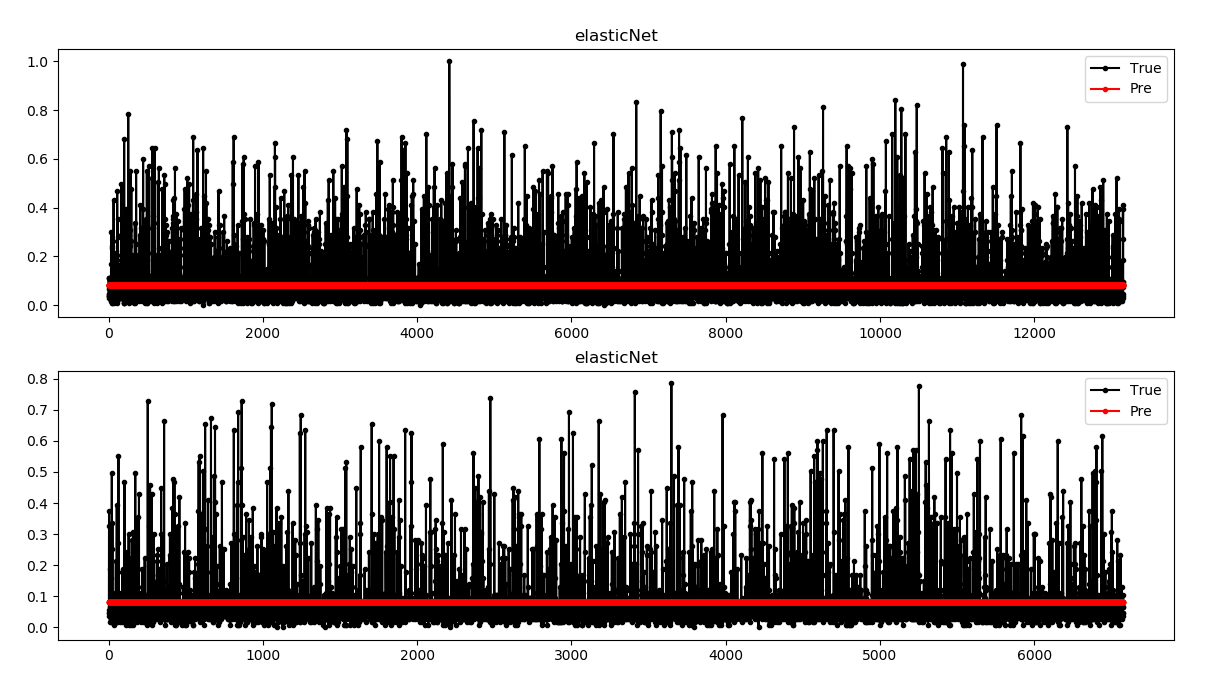
#### 2、ridge



#### 3、lasso

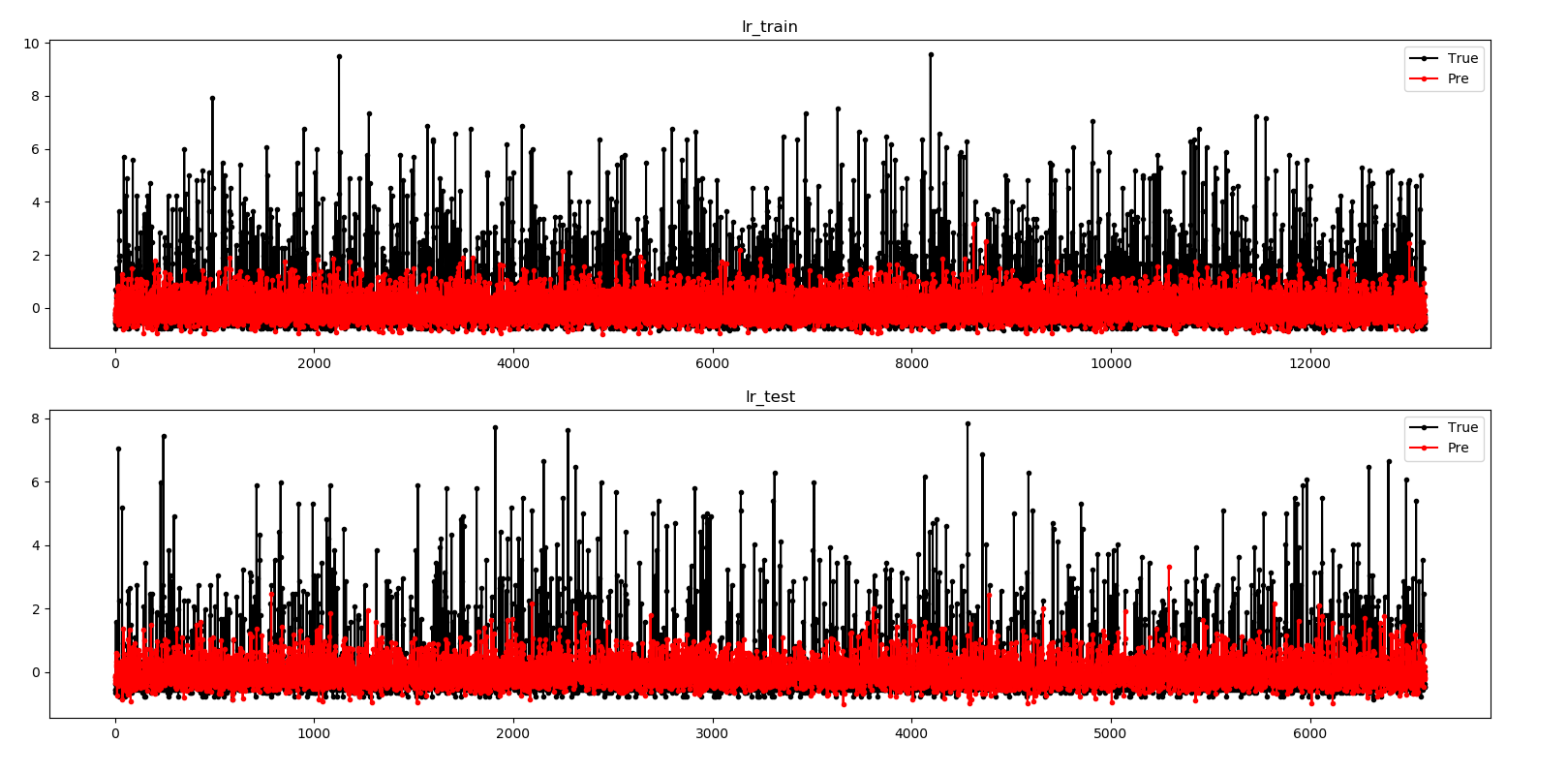


#### lasso



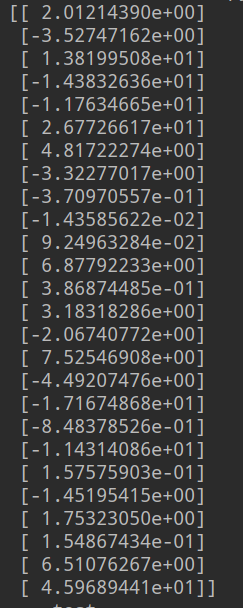
对数据进行高斯标准化

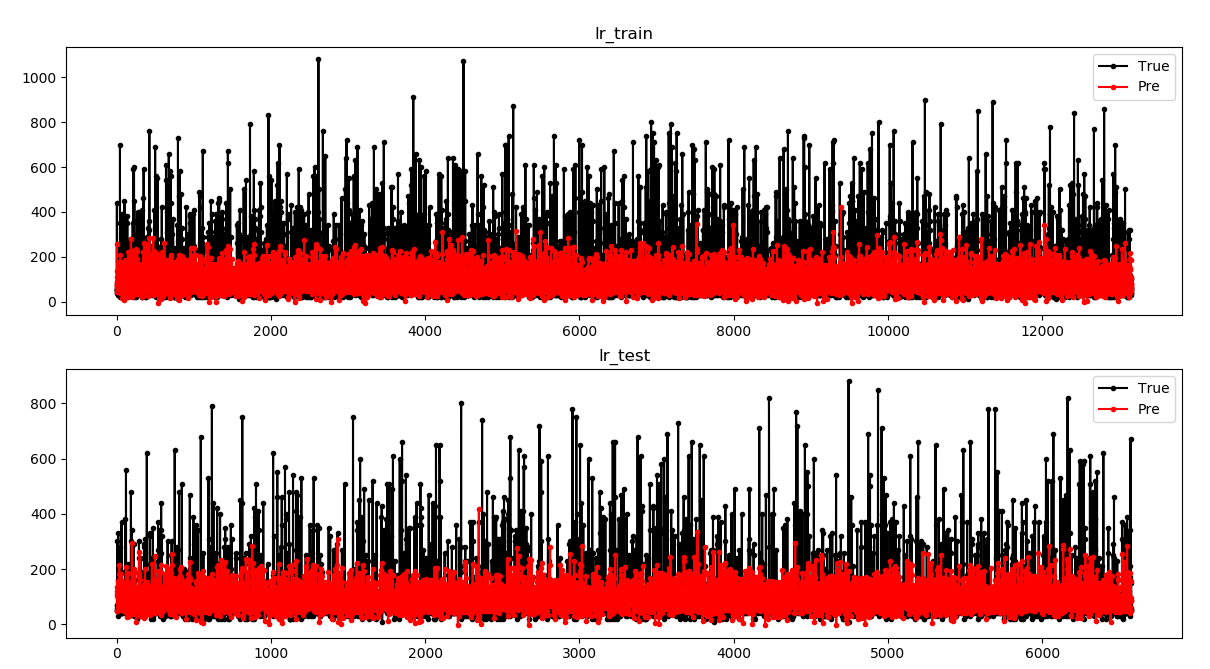




### 3、手动求最优解

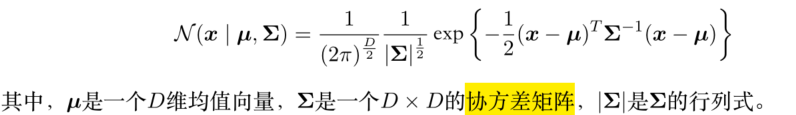
加了正则项的结果，最后求得的参数如下。





### 4、对数据进行基变换

对原始数据变量进行某种固定形式的预处理或者特征抽取，如果原始变量为向量，那么原始变量的特征可以用基函数{},用多个多元高斯分布，构建一个新的，具有J维特征。



### 5、结论

尝试了多种方法，均没有训练出较理想的模型，训练阶段就存在较大误差。