**形状

中度可信度描述已自动生成**

**智能制造质量预测**

组长： 张庆伟 3210103414

组员： 葛涛 3210103407

组员： 刘祥盛 321010XXXX

2023年11月08日

目录

[**一、** **问题介绍** 2](#_Toc150333697)

[**1.1.** **活动背景** 2](#_Toc150333698)

[**1.2.** **活动目标** 2](#_Toc150333699)

[**二、** **数据处理** 3](#_Toc150333700)

[**三、** **线性回归** 4](#_Toc150333701)

[**3.1** **立项阶段** 4](#_Toc150333702)

[**3.2** **外部对接1** 4](#_Toc150333703)

[**3.3** **外部对接2** 4](#_Toc150333704)

[**3.4** **筹备阶段** 4](#_Toc150333705)

[**3.5** **活动阶段1** 4](#_Toc150333706)

[**3.6** **活动阶段2** 4](#_Toc150333707)

[**3.7** **总结阶段** 4](#_Toc150333708)

[**四、** **Lasso回归** 5](#_Toc150333709)

[**五、** **Ridge回归** 5](#_Toc150333710)

[**六、** **小组分工** 5](#_Toc150333711)

1. **问题介绍**
   1. **活动背景**

从数字火炬

业做好准备。

* 1. **活动目标**

本次参观

1. **数据处理**

本次活动的主要流程如下表所示：

具体时间安排见下节“时间轴”。

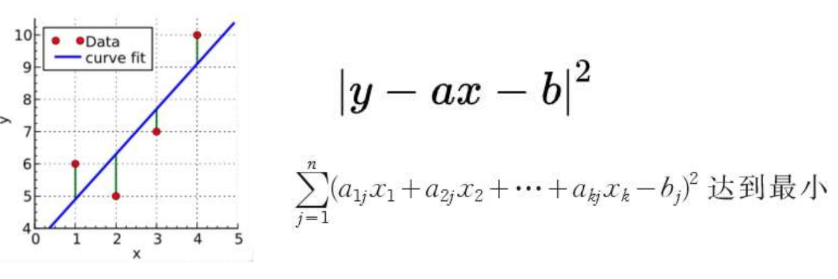
1. **线性回归**
   1. **基本含义**

在统计学中，线性回归（Linear Regression）是利用称为线性回归方程的最小平方函数对一个或多个自变量和因变量之间关系进行建模的一种回归分析。这种函数是一个或多个称为回归系数的模型参数的线性组合。只有一个自变量的情况称为简单回归，大于一个自变量情况的叫做多元回归。

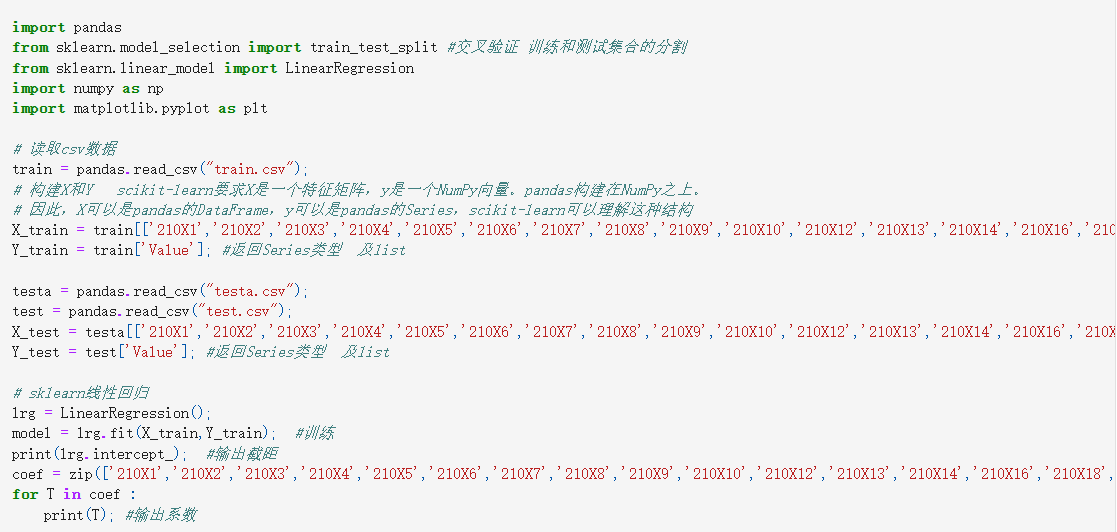
* 1. **拟合方程**

一般来说，线性回归都可以通过最小二乘法求出其方程，一般地，影响y的因素往往不止一个，假设有x1，x2，...，xk，k个因素，通常可考虑如下的线性关系式：

用“距离直线的竖直距离的平方”反应误差



* 1. **代码部分展示**

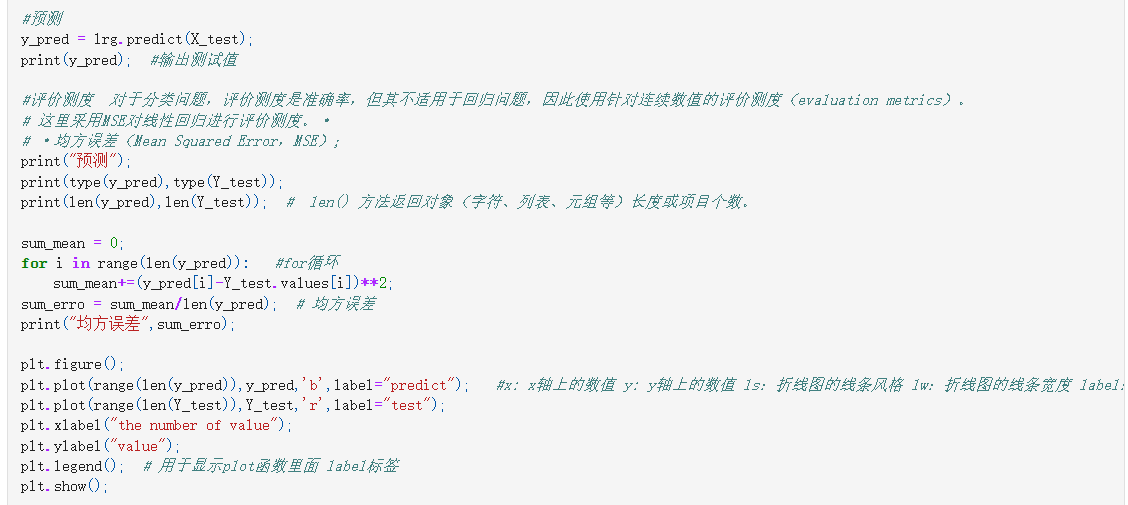


先是对训练集的数据进行线性回归的模型训练

lrg = LinearRegression();

model = lrg.fit(X\_train,Y\_train);  #训练

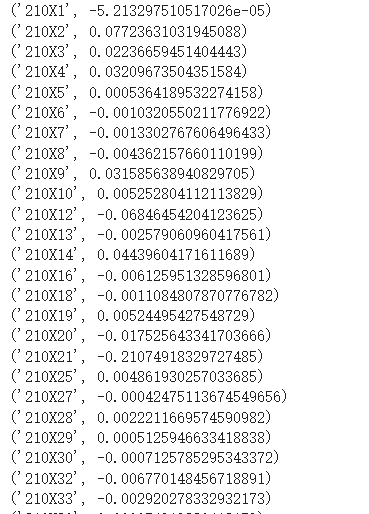
将训练集的多元数据与Value值采用最小二乘法，建立多元线性方程，得到各个变量的系数。



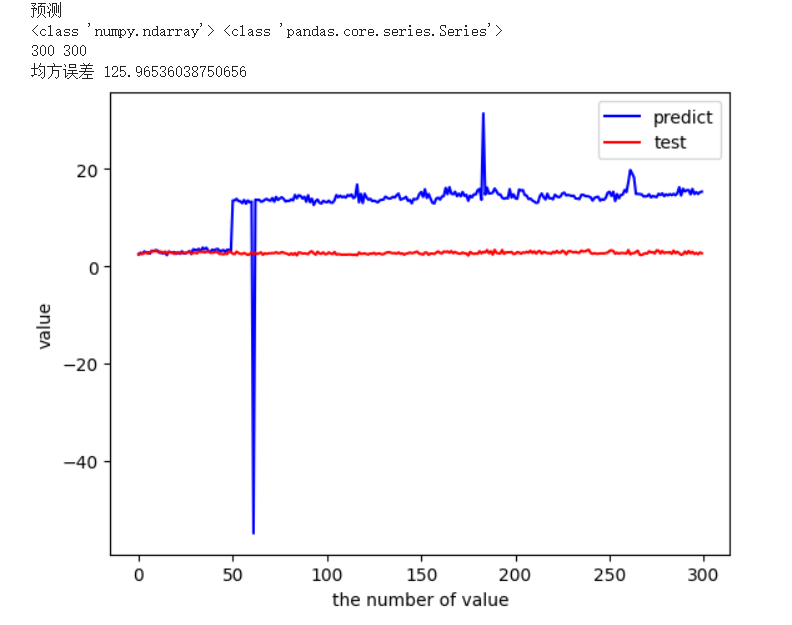
将测试集的数据代入到训练得到的多元线性方程中，并求出预测值与真实值的均方误差作为评价测度，画出预测值与实际值比较的折线图。

* 1. **运行结果**

①各个变量的系数：

****

②评价测度值以及各预测数据与实际数据比较的折线图：



* 1. **结果分析**

结果显示，线性回归算法得到结果的均方误差高达一百多，而通过折线图可以看出预测数据的前四十组数据与实际值相近，而后面的两百多组数据都与实际值有着相似较大的偏差。

显然，对训练集使用线性回归训练后的模型对测试集的预测结果不太理想，误差较大。

误差分析：

* 1. 训练集数据大小范围较为集中（即工艺相似，种类少），而测试集的数据大小跨度大（工艺种类多），导致测试结果与训练数据大小相近（工艺种类相似）的误差小（如前四五十个数据）；而后面预测的数据误差很大。
  2. 线性回归最小二乘的算法容易受“离群值”的影响，离群值影响被平方放大，拟合会“迁合”这些离群值，这是线性回归训练模型的不足。
  3. 线性回归缺乏因果推断能力，多元线性回归只能表明相关性，但无法推断出因果关系，而工艺制造的数据之间往往存在因果关系。
  4. **总结**

采用线性回归模型对该智能制造质量预测存在较大的不足，不是一个理想的预测模型。导致线性回归模型缺陷的可能有以下原因：

①对异常值和离群点敏感：

多元线性回归对异常值和离群点敏感，这些异常值可能会对回归系数产生较大影响，从而影响模型的稳健性和准确性。

②需要满足线性假设：

多元线性回归模型要求自变量与因变量之间的关系是线性的，如果实际关系存在非线性，就会导致模型拟合不佳。

③对多重共线性较为敏感：

当自变量之间存在高度相关性（多重共线性）时，多元线性回归模型的参数估计会变得不稳定，使得模型的解释能力下降。

④对数据分布要求较高：

多元线性回归模型通常要求自变量和因变量的分布符合正态分布，且误差项具有相同方差（同方差性），否则会影响参数估计的准确性。

⑤需要满足独立性假设：

多元线性回归模型假设误差项之间是相互独立的，如果存在自相关性（误差项之间相关）就会违反该假设。

1. **Lasso回归**

控

1. **Ridge回归**

理。

1. **小组分工**

从数字火炬

业做好准备。