浙江工业大学

# 计算方法及实现期末论文

## 2022/2023(2)


论文题目　最小二乘法在线性回归中的应用

学生姓名　陈彦克

学生学号　202103150604

学生班级　计智 2101

任课教师　池凯凯

提交日期　2023 年 6 月 7 日

**计算机科学与技术学院、软件学院**

# 最小二乘法在线性回归中的应用


## 摘要

线性回归是一种常用的统计分析方法，用于研究自变量与因变量之间的线性关系。最小二乘法是线性回归中最常用的估计方法之一，通过最小化观测值与回归模型之间的残差平方和来确定最佳拟合直线。本文介绍了线性回归的原理，推导最小二乘法在线性回归中的应用公式，并给出了一个房价预测实例，说明了如何使用最小二乘法来求解线性回归模型的参数。



## 关键词

最小二乘法　线性回归　参数估计　模型评估


# 1 引言

线性回归是统计学和机器学习中非常重要的模型，其原理简单明了、易于理解，这使得线性回归成为许多人接触和学习机器学习的入门模型。同时它在机器学习和统计学中也起到了基础性的作用。许多复杂的机器学习算法，如逻辑回归、支持向量机、神经网络等，都可以由线性回归扩展得到。线性回归模型虽然原理简单，但在实际应用中非常广泛。尽管它假设数据之间存在线性关系，这在现实中可能并不总是成立，但在许多情况下，线性模型已经足够好，能够提供有效的预测。例如在经济学领域中，经济学家使用线性回归分析各种经济变量之间的关系，如预测未来的消费支出、GDP 增长等；在医学研究中，线性回归可以用来探究不同的生活习惯、环境因素如何影响人的健康状况；在金融领域，线性回归可以用来预测股票的价格、预测信用评分等。

最小二乘法是一种常见的优化算法，是求解线性回归模型的参数的核心估计方法。最小二乘法的主要思想是通过确定未知参数（通常是一个参数矩阵），来使得真实值和预测值的误差（也称残差）平方和最小。最小二乘法有很多优点，例如它可以利用矩阵运算，高效地计算出唯一的最优解（如果存在）；它还可以提供参数估计的置信区间和显著性检验等。

本文将介绍最小二乘法在线性回归中的应用，首先介绍线性回归模型的基本概念和假设，然后介绍最小二乘法的原理和求解方法，接着给出一个房价预测的多元线性回归的实例，最后总结最小二乘法的优缺点。

# 2 最小二乘法在线性回归中的应用

## 2.1 线性回归的定义

线性回归是一种用于建立自变量（输入变量）与因变量（输出变量）之间线性关系的统计分析方法。它通过寻找最佳拟合直线（或超平面）来描述自变量与因变量之间的关系。线性回归假设自变量和因变量之间存在一个线性关系，即因

变量可以通过自变量的线性组合来预测。

线性回归可分为：简单一元线性回归和多元线性回归，二者的主要区别也就是未知项的个数。

## 2.2 损失函数

为了找出回归中的最佳拟合直线(或超平面)，我们需要进行以下两个步骤：（1）想办法表示出这条直线到所有已知数据点的距离；（2）让这些距离最小。这样就能找出这种处于所有数据点中间的直线。

假设函数计算出的值$\hat{y}$与真实数据点$y_i$的间隔（差值）就是我们要找的点到直线的距离。在机器学习中这个差值叫误差（或者叫残差，符号表示：$\varepsilon$），其表达式为：

$$\varepsilon = y_i - \hat{y}_i。$$

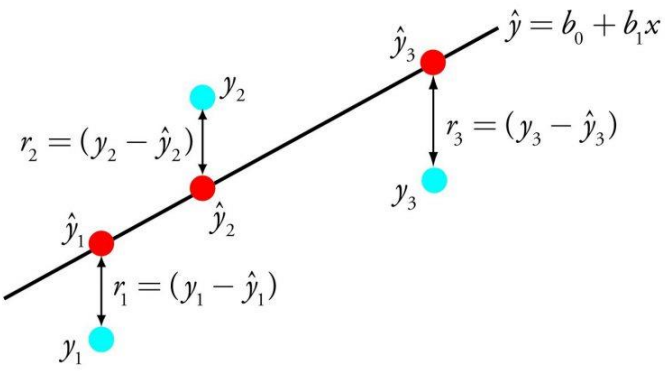


**图 1　真实值与预测值之间的误差$\varepsilon$**

由于误差的值有正负之分，为了处理方便，我们考虑将误差平方，称其为平方误差，表达式为：

$$\sum \varepsilon^2 = (y_i - \hat{y}_i)^2。$$

这个公式就是残差平方和，即 SSE（Sum of Squares for Error）。在机器学习中，人们也称误差为损失，所以这种求误差的方法也可以说是求损失的方法。而 SSE 也就是线性回归中最常用的损失函数了。

残差平方和 SSE 的矩阵表示形式如下：

$$SSE = (y - Xw)^T (y - Xw)$$

现在我们知道了损失函数是衡量回归模型误差的函数，也就是我们要找出的最佳“直线”的评价标准。这个函数的值越小，说明直线越能够拟合我们的数据。

## 2.3 使用最小二乘法求解最优值

现在问题就转换成了求解一组参数向量，使损失函数最小化。在线性回归中，

这种通过最小化真实值和预测值之间的损失函数值来求解参数的方法叫做最小二乘法。

## 一元线性回归参数求解

我们先以一元线性模型为例来说明。

假设有一组数据$X=\{(x_1,y_1),\ldots,(x_m,y_m)\}$，我们希望求出对应的一元线性模型来拟合这一组数据：$y=\beta_0+\beta_1 x$。

既然要拟合，我们需要一个评判拟合程度好坏的度量方式。上文说到，最小二乘法中使用的拟合度量方式就是误差平方和方法。所以，这时候的损失函数，或者说我们的目标函数就是：

$$J(\beta)=\sum_{i=0}^{m}(y_i-\beta_1 x_i-\beta_0)^2。$$

有了这个目标函数，我们要做的就是求出一组最优的参数值$\beta_0$和$\beta_1$，得$J(\beta)$最小，在这里就是极小值。

求极值的一个很好的方法就是求导。这里由于有多个参数，所以我们分别对$\beta_0$和$\beta_1$求偏导：

$$\frac{\partial J(\beta)}{\partial \beta_1}=\sum_{i=0}^{m}2(y_i-\beta_1 x_i-\beta_0)(-x_i)=2\sum_{i=0}^{m}(\beta_1 x_i^2+\beta_0 x_i-x_i y_i)$$

$$\frac{\partial J(\beta)}{\partial \beta_0}=\sum_{i=0}^{m}2(y_i-\beta_1 x_i-\beta_0)(-1)=2\sum_{i=0}^{m}(\beta_1 x_i^2+\beta_0-y_i)$$

因为：

$$\overline{x}=\frac{\sum_1^m x_i}{m},\overline{y}=\frac{\sum_1^m y_i}{m}$$

所以，对$\beta_0$的偏导可以转化为：

$$\frac{\partial J(\beta)}{\partial \beta_0}=2(m\beta_1\overline{x}+m\beta_0-m\overline{y})$$

我们知道，当目标函数取得极值时，偏导等于0。所以我们令等式左边式子$\frac{\partial J(\beta)}{\partial \beta_0}$为0，于是有：

$$\beta_0=\overline{y}-\beta_1\overline{x}$$

接着，我们继续回到上面对$\beta_1$的偏导。令$\frac{\partial J(\beta)}{\partial \beta_1}=0$，并将$\beta_0=\overline{y}-\beta_1\overline{x}$代入，得：

$$2\sum_{i=0}^{m}(\beta_1 x_i^2+(\overline{y}-\beta_1\overline{x})x_i-x_i y_i)=0$$

$$\beta_1=\frac{\sum_{i=1}^m x_i y_i-\overline{y}\sum_{i=1}^m x_i}{\sum_{i=1}^m x_i^2-\overline{x}\sum_{i=1}^m x_i}$$

求和后化简可得：

$$\beta_1=\frac{\sum_{i=1}^m (x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^m (x_i-\overline{x})^2}$$

因此，$y=\beta_0+\beta_1 x$的最小二乘法的解为：

$$\begin{cases}\beta_1=\dfrac{\sum_{i=1}^{m}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{m}(x_i-\overline{x})^2}\\ \beta_0=\overline{y}-\beta_1\overline{x}\end{cases}$$

### 多元线性回归参数求解

对于多元线性回归的情况，我们需要使用矩阵运算来求解。这里先用矩阵表示原式：

$$X\beta=y$$

其中，

$$X=\begin{bmatrix}1 & x_{12} & \dots & x_{1n}\\ 1 & x_{22} & \dots & x_{2n}\\ \dots & \dots & \dots & \dots\\ 1 & x_{m2} & \dots & x_{mn}\end{bmatrix},\ \beta=\begin{bmatrix}\beta_0\\ \beta_1\\ \dots\\ \beta_n\end{bmatrix},y=\begin{bmatrix}y_1\\ \dots\\ y_m\end{bmatrix}$$

而我们的目标函数为：

$$J(\beta)=\sum_{i=1}^{m}|y_i-\sum_{j=1}^{n}x_{ij}\beta_j|^2=\|y-X\beta^T\|^2$$

如果要使上述目标函数最小，显然其最小值为 0，即：

$$y-X\beta^T=0$$

也就是说：

$$X\beta^T=y$$
$$X^TX\beta^T=X^Ty$$
$$(X^TX)^{-1}X^TX\beta^T=(X^TX)^{-1}X^Ty$$

最终获得解：

$$\beta^T=(X^TX)^{-1}X^Ty$$

可以看出，对于一般的最小二乘法求解多元线性回归的参数最优值时，使用矩阵运算即可，不需要使用迭代法。

## 2.4 线性拟合效果评估指标

在线性回归中，我们常采用均方误差和 R2 系数作为线性拟合的评估指标。

### 均方误差（Mean Squared Error，MSE）

均方误差 MSE 是最常见的评估指标之一，它衡量的是实际观测值与模型预测值之间的误差平方和的平均值。计算均方误差的公式如下：

$$MSE=\frac{1}{n}\sum_{i=1}^{m}(y_i-\hat{y}_i)^2$$

### R2 系数（Coefficient of Determination）

在统计学中，R2 系数又称决定系数，它反映因变量的全部变异能通过回归关系被自变量解释的比例。R2 系数的取值范围在 0 到 1 之间，值越接近 1 表示

模型拟合得越好。计算 R2 系数的公式如下：

$$R2\_score = 1 - \frac{\sum_{i=1}^{m} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{m} (y_i - \overline{y})^2}$$

## 2.5 最小二乘法的局限性以及梯度下降法

在线性回归中，最小二乘法是一种解析解方法，它可以直接计算出最优的模型参数。但是，在非线性回归中，模型的参数无法通过解析方式获得，因此需要使用数值优化算法来找到最优解。梯度下降法是一种常用的数值优化算法，适用于寻找非线性回归模型的最优参数。

梯度下降法的基本思想是通过迭代更新模型参数，以使损失函数最小化。具体步骤如下：

1. 初始化模型参数。
2. 计算当前参数下的损失函数值。
3. 计算损失函数关于参数的梯度。
4. 更新参数：将当前参数减去梯度乘以学习率。
5. 重复步骤 2-4，直到满足停止条件（例如达到最大迭代次数或损失函数的变化很小）。

**Algorithm1：梯度下降法**

Require: 步长$\alpha$，初始参数$x_0$

repeat:

梯度计算：$\nabla f(x_i)$

参数更新：$x_{i+1} = x_i - \alpha \nabla f(x_i)$

until: 达到收敛条件

**图 2 梯度下降法的基本步骤**

梯度下降法的核心思想是通过沿着损失函数下降最快的方向更新参数，从而逐步接近最优解。这种迭代的过程可以在非线性回归中寻找最优参数。

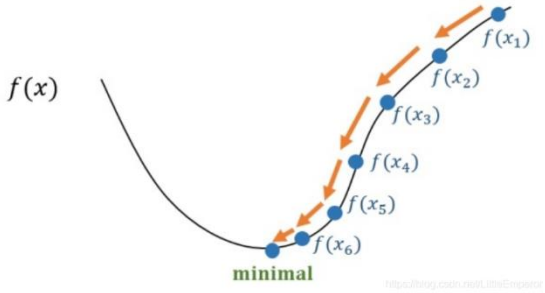


**图 3 梯度下降法的示意图**

相比之下，最小二乘法在非线性回归中的应用受到限制，因为无法直接求解最优参数。非线性回归的损失函数通常具有多个局部最小值，而最小二乘法只能找到其中的一个最小值，可能无法得到全局最优解。因此我们在复杂的回归模型

中通常使用梯度下降法来求解参数的最佳值。

需要注意的是，常用的梯度下降法有很多，例如批量梯度下降法、小批量梯度下降法和随机梯度下降法，等等。这些梯度下降法的选择取决于具体的问题和数据集的大小。我们需要根据具体问题和数据集的特点选择适合的梯度下降法，并进行调节参数值，以获得最佳结果。

# 3 实验及结果

## 3.1 实验数据集

本实验选取机器学习中线性回归的经典数据集——波士顿房价数据集。这份数据集共 506 行，每行包含了波士顿郊区的一类房屋的相关信息及该类房屋价格的中位数。其各维属性的意义如下：

**表 1　数据集属性介绍**

| 属性名 | 解释 | 类型 |
|---|---|---|
| CRIM | 该镇的人均犯罪率 | |
| ZN | 占地面积超过 25,000 平方呎的住宅用地比例 | 连续值 |
| INDUS | 非零售商业用地比例 | 连续值 |
| CHAS | 是否邻近 Charles River（查尔斯河） | 离散值，1=邻近；0=不邻近 |
| NOX | 一氧化氮浓度 | 连续值 |
| RM | 每栋房屋的平均客房数 | 连续值 |
| AGE | 1940 年之前建成的自用单位比例 | 连续值 |
| DIS | 到波士顿 5 个就业中心的加权距离 | 连续值 |
| RAD | 到径向公路的可达性指数 | 连续值 |
| TAX | 全值财产税率 | 连续值 |
| PTRATIO | 学生与教师的比例 | 连续值 |
| B | 1000(BK - 0.63)^2，其中 BK 为黑人占比 | 连续值 |
| LSTAT | 低收入人群占比 | 连续值 |
| MEDV | 同类房屋价格的中位数 | 连续值 |

## 3.2 实验内容

模型假设：在波士顿房价数据集中，和房屋相关的值共有 14 个，前 13 个用来描述房屋相关的各种相关信息，即模型中的$x_i$；最后一个值为我们要预测的该类房屋价格的中位数，即模型中的$y_i$。

我们假设这些特性信息与房价之间是线性相关的。因此，我们的模型就可以表示成：

$$\hat{y} = w_1x_1 + w_2x_2 + \ldots + w_{13}x_{13} + b$$

实验中我们使用 python 编程来求解问题。步骤如下：

1. 导入房价数据集，分为训练集和测试集两部分；
2. 对数据进行一些预处理工作（包括数据归一化等等）；
3. 让线性回归模型在训练集上用最小二乘法做参数拟合；
4. 让训练完的模型和拟合的参数在测试集上做预测；
5. 画出预测值与实际值的房价曲线，求出均方误差的值；
6. 评估线性回归模型的拟合效果。

## 3.3 实验结果与分析

程序运行结果如下图：

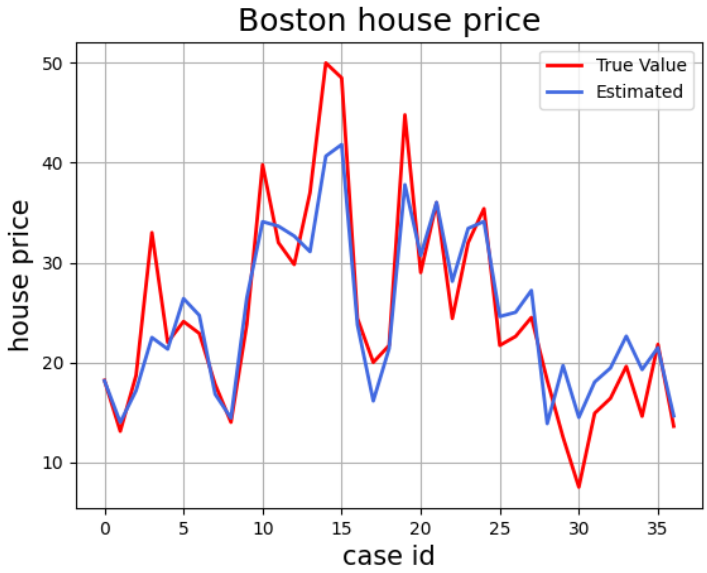


**图 4　房价的真实值（红色）与预测值（蓝色）的对比**

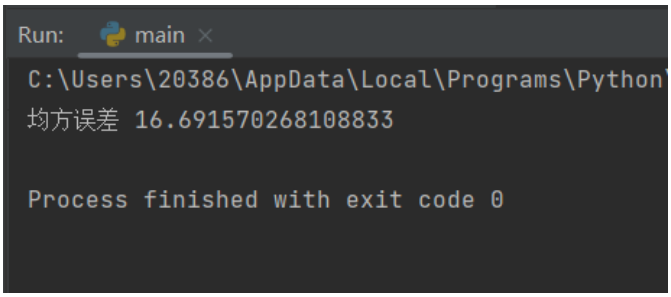


**图 5　均方误差值**

从图 8 中我们可以看出，使用多元线性回归去拟合，房价的真实值（红色）和预测值（蓝色）在大体上拟合效果较好，这说明我们的多元线性回归模型能够挖掘出房价与特征之间的规律，预测出房价的真实价格。

从图 9 我们可以看出，在测试集上的均方误差值较低，约为 16.70，这说明拟合误差较小。

# 4 结语

本论文研究了最小二乘法在线性回归中的应用，并对其在实际问题中的重要性和优势进行了探讨。最小二乘法作为一种常见的统计方法，通过最小化误差的平方和来估计线性回归模型的参数，具有简单、稳定和有效的特点。在本文中，我们首先介绍了线性回归模型的基本原理和假设，然后详细阐述了最小二乘法的数学推导和求解过程。通过应用最小二乘法，我们可以得到对线性回归模型参数的最优估计，从而实现对观测数据的预测和分析。

在研究过程中，我们通过对房价数据集的实例分析和数值实验验证了最小二乘法在线性回归中的可行性和有效性。结果表明，最小二乘法能够准确地估计出线性回归模型的参数，并且对于存在噪声的数据集也具有一定的鲁棒性。

最小二乘法的最大优点是直观和简单。对于简单的线性回归问题，最小二乘法存在解析解，可以直接计算出参数的数值，这使得计算效率较高。

而最小二乘法的缺点是对数据中的异常值敏感。并且最小二乘法仅能处理线性关系，对于非线性问题的拟合能力有限。在实际应用中，如果数据存在非线性关系，最小二乘法可能无法提供准确的模型拟合，因此延伸出了梯度下降这种通过迭代求解参数最佳值的方法。

近年来，随着计算机算力的增强、数据规模的爆炸式增长，人工神经网络在各领域兴起。人工神经网络凭借其强大的数据处理能力和关系建模能力，在处理复杂的回归问题时往往能得到更好的结果。但使用最小二乘法的拟合参数的思想在整个机器学习领域中仍具有十分重要的借鉴意义。

# 参考文献

[1]朱建新，李有法.数值计算方法[M].北京：高等教育出版社，2020

[2]李俊岳,胡典顺.概率与统计的知识理解之最小二乘法[J].数学通讯,2023,No.895(02):12-15+37.

[3]李小宇.线性回归分析在电商店铺指数预测中的应用[J].中国市场,2022,No.1127(28):185-187.DOI:10.13939/j.cnki.zgsc.2022.28.185.