**北京师范大学珠海分校**

**本科生毕业论文**

论文题目 **一种简单而重要的数学优化技术**

**— 最小二乘法及其应用**

学 院　　 应用数学学院

专 业 数学与应用数学

学 号 1617010009

学 生 姓 名 陈雪敏

指导教师姓名 李艳

指导教师单位 北师大珠海分校应用数学学院

2020年3月30日

**北京师范大学珠海分校学位论文写作声明和使用授权说明**

**学位论文写作声明**

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品或成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本声明的法律结果由本人承担。

论文作者签名： 日期： 年 月 日

**学位论文使用授权说明**

本人完全了解北京师范大学珠海分校关于收集、保存、使用学位论文的规定，即：按照学校要求提交学位论文的印刷本和电子版本；学校有权保存学位论文的印刷本和电子版，并提供目录检索与阅览服务；学校可以采用影印、缩印、数字化或其它复制手段保存论文；在不以赢利为目的的前提下，学校可以将学位论文编入有关数据库,提供网上服务。（保密论文在解密后遵守此规定）

论文作者签名： 导师签名：

日期： 年 月 日

一种简单而重要的数学优化技术

* 最小二乘法及其应用

摘要

最小二乘法是一种简单而重要的数学优化技术，主要应用于数据的曲线拟合。其方法是使所得拟合曲线上的各点的值与测量数据的值的差的平方和最小。拟合曲线不要求与测量的所有点都切合，只需要能够反映出测量数据的大致发展趋势。最小二乘法自发现以来就在科学实验、工程技术中发挥着重要的作用。计算机科学的快速发展，也给最小二乘法提供了不断发展和应用的动力。本文首先介绍了最小二乘法的发展历史和研究现状，之后归纳总结了最小二乘法曲线拟合的原理和几种常见的实现方式，包括一元线性拟合、多元线性拟合以及非线性拟合，并给出详细的求解过程。最后给出了最小二乘法在不同领域中拟合及预测的应用示例，探讨了针对不同复杂程度的拟合问题如何选择合适的拟合方法。

关键词：最小二乘法；曲线拟合；预测模型

A SIMPLE AND IMPORTANT MATHEMATICAL OPTIMZATION TECHNIQUE

—THE LEAST SQUARE METHOD AND ITS APPLICATIONS

**ABSTRACT**

Least squares is a simple and important mathematical optimization technique, which is mainly used for curve fitting of data. The method is to minimize the sum of squares of the difference between the value of each point on the obtained fitted curve and the value of the measurement data. The fitting curve does not need to be consistent with all points of the measurement, it only needs to be able to reflect the general development trend of the measurement data. The least square method has played an important role in scientific experiments and engineering technology since its discovery. The rapid development of computer science has also provided the impetus for the continuous development and application of the least square method. This article first introduces the development history and research status of the least squares method, and then summarizes the principles of least square curve fitting and several common implementation methods, including univariate linear fitting, multivariate linear fitting, and nonlinear fitting. The detailed solution process is given. Finally, the application examples of least squares fitting and prediction in different fields are given, and how to choose suitable fitting methods for fitting problems with different complexity is discussed.

**Key words:** Least squares; Curve Fitting; Predictive model

# 目 录

[1 绪论 1](#_Toc29592)

[1.1 最小二乘法的历史 1](#_Toc31727)

[1.2 研究意义 2](#_Toc2346)

[1.3 国内外研究现状 3](#_Toc4264)

[2 最小二乘法 4](#_Toc18124)

[2.1 最小二乘法的原理 4](#_Toc28664)

[2.2 曲线拟合 4](#_Toc15711)

[2.2.1 一元线性拟合 4](#_Toc14171)

[2.2.2 多元线性拟合 6](#_Toc11269)

[2.2.3 非线性拟合 7](#_Toc19099)

[2.3 曲线拟合模型评价 9](#_Toc19403)

[3 最小二乘法的应用示例 10](#_Toc31063)

[3.1 在温度传感器中的应用示例 10](#_Toc17806)

[3.2 在房地产销售价格中的应用示例 13](#_Toc27078)

[3.3 在销售预测中的应用示例 17](#_Toc3121)

[3.4 在传染病人数预测中的应用示例 21](#_Toc15013)

[4 结论 24](#_Toc25951)

[参考文献 26](#_Toc7526)

[致 谢 28](#_Toc32099)

# 绪论

## 最小二乘法的历史

最小二乘法是在天文学和测地学的发展中产生的。从古代到十八世纪，天文学都是应用数学中最发达的领域。在研究天文学的具体问题时，通常需要对观测的数据进行分析计算来确定结果。其中所研究的大多数问题可以归结为以下数学模型：假设观测值为，未知量为，它们有线性关系



通过观测多组数据，我们可以得到具有个未知量，个方程的线性方程组

，

由于存在观测误差，我们又将方程组表示为

，，

其中为随机误差。

我们需要通过观测数据得到未知量参数。在解决该问题时，若不存在误差，我们通过联立方程组很容易能够解得未知量。但正是因为误差的存在，如果我们仅仅从方程组中挑出其中个方程来求解，这样测量数据就不能充分利用，得到的结果也不够精确。因此如何处理观测数据中的误差，充分利用观测结果成为一个棘手的问题。18世纪数学家欧拉、梅耶、拉普拉斯都曾被这样的问题所困扰。

1749年，由于木星与土星相互吸引对各自的轨道产生影响，欧拉基于天文的观测数据对木星与土星的运行轨道进行研究，得到含75个方程，8个未知量的线性方程组。欧拉对方程求解进行了尝试，但后来证明欧拉的做法缺乏合理性，结果不太精确。

1750年，梅耶由确定地球上某定点的定位问题，得到含27个方程，3个未知数的线性方程组。梅耶在求解该问题时，按照方程的其中一个系数的大小，将27个方程分为3组，较大的9个，中间的9个，较小的9个分别为一组。将每组的9个方程相加，得到3个新的方程从而求解三个未知参数。

1787年，拉普拉斯同样也遇到这样的问题，他在天文学研究中得到含24个方程，4个未知数的一组方程。拉普拉斯在求解时也用他独特的做法将24个方程组合得出4个方程。

欧拉、梅耶和拉普拉斯在解决这样有个未知量，个方程的问题时，都设法将个方程组合成个方程再求解，但他们在求解的过程中没有一个固定的规律，忽视了整体的均衡性。

1805年勒让德在其论著《计算彗星轨道的新方法》中提出了最小二乘法。勒让德在解决问题时的想法不同于早期科学家们，他没有致力于找到与未知数个数相同的方程数来求解，而是尝试从误差在各个方程中的均衡性出发，最终提出要使误差的平方和最小。

1809年高斯发表了著作《天体运动理论》，高斯在该著作中提出了与正态误差分布相结合的最小二乘法，并称自己在更早的时候就开始使用这个方法。近代学者认为勒让德与高斯是各自独立地发明了最小二乘法，他们发现的角度不同。勒让德提出的最小二乘法作为一种计算方法使得在数据分析时更为简便，而高斯由误差密度函数推导出最小二乘法，则给予了该方法更强有力的理论支撑。

可以看出，十八世纪，最小二乘法成功创立并广泛运用于天文学、测地学的研究中。而从那以后，最小二乘法在数理统计学的发展中发挥着巨大的作用。相关回归分析、方差分析和线性模型理论等数理统计学的几大分支都以最小二乘法为理论基础。有人称最小二乘法为数理统计学的灵魂，也有人认为“最小二乘法之于数理统计学犹如微积分之于数学”。

## 研究意义

随着信息化时代的到来，各个学科领域对于数据处理的需求日益增加，数学优化技术与各个领域相结合，也成为实验研究的重要组成部分。对于研究中模型参数的估计与曲线拟合、误差分析息息相关，往往要用平滑的曲线描绘出数据发展的趋势，最小二乘法的重要性也在其中显现。

随着科学技术的快速发展，计算机科学的进步也给最小二乘法的应用提供了计算支持。最小二乘法不仅广泛应用于物理、化学等自然学科，在工程技术、经济、农业等各个领域都发挥着重要作用。以最小二乘法为基础的线性模型在人工智能领域中的机器学习研究中也占有重要的地位。因此，无论是在理论上还是在应用上，最小二乘法都是很值得我们探讨的一种数学优化技术。相比其他复杂的优化方法，最小二乘法的原理很简单，也很容易求解，因此也成为应用最广泛的优化方法之一。本论文探讨最小二乘的不同形式，并给出应用实例和求解过程，进一步探讨不同的问题应如何选择不同形式的模型，这在方法上和应用上都有一定的意义和价值。

## 国内外研究现状

1905年Legendre最早提出最小二乘法。1809年Gauss提出最小二乘法，并将最小二乘法与正态误差理论相结合，推证了最小二乘法，丰富了最小二乘法的应用，提高了最小二乘法的重要性。1912年Markov系统阐述了最小二乘原理，得出测量平差中的高斯-马尔柯夫模型(Gauss-Markov Theorem)。

张东林在最小二乘曲线拟合的基础上提出了分段最小二乘拟合，给出了分段最小二乘多项式拟合的子程序[7]。侯超钧等提出了全局连续的多分段区间的最小二乘拟合[8]。丁克良等基于间接平差原理推导了正交最小二乘曲线拟合的方法，同时顾及了自变量与因变量的误差[9]。P. Lancaster 提出了移动最小二乘法[10]。Y. X. Mukherjee 等进行深入研究提出改进的移动最小二乘法[11]。陈美娟等利用带权正交函数作为基函数对移动最小二乘法进行了改进[12]。任红萍等提出了移动最小二乘插值法[13]。

此外最小二乘法曲线拟合应用广泛。曹可运用最小二乘法曲线拟合预测了传染病的发病人数[14]。王桂杰介绍了最小二乘法在三维坐标定位中的应用，并对最小二乘定位算法进行了优化[15]。李大鹏在液体火箭推进剂温度预测模型研究中，利用壁面实测温度数据样本，建立理论数学模型，采用最小二乘法进行拟合，消除了推进剂插入式测温方式带来的隐患[16]。王知雨等提出了一种采用非线性最小二乘法的等效电路模型参数辨识方法[17]。王辅臣等将最小二乘法应用到一种实时场地矫正方法中[18]。Ayse Gul Kaplan 等结合移动最小二乘法提出太阳辐射量的新模型[19]。Czaplewski K 等在开发定位电车轨道方法中，用带条件方程的最小二乘法提高了测定精度[20]。Asadpour 等用移动最小二乘法求解车道嵌入方程数值[21]。Gautam C等将最小二乘法应用于生物医学数据中[22]。

可见，最小二乘法自十八世纪提出以来，已经广泛应用于统计、医学、光学、传热、电路、定位等领域中。

# 最小二乘法

## 最小二乘法的原理

最小二乘法是一种数学优化方法，其原理很简单，就是给已知数据找出一条最优的拟合曲线，使拟合曲线上所对应点的值与测量值误差平方和最小。假设测量值为，待定参数为，对测量值进行拟合的曲线的关系式为：



记，其中是与实验点相对应的测量值，为实验点在的关系下得到的函数值，即

，

为了减小误差，在实验时通常会测量多组数据，使方程个数大于待定参数个数，即使得，构成矛盾方程组。通过最小二乘法使误差平方和，即最小，求出待定参数。

## 曲线拟合

对于实际需要拟合的数据，如果没有问题相关的领域知识或者专家所提供的信息，通常可以先画出数据的散点图，观察散点图用哪种与哪类函数图像更为接近，从而选取合适的曲线拟合方程。当然，通过画散点图进行可视化只适用于维数不大于3的数据，如果是高维数据的话，一般可视问题的复杂程度和问题性质（如变量的个数、因变量与自变量之间是线性还是非线性）对拟合方程进行选择，或者在无任何信息情况下，则通常采用不同拟合模型进行试错。

### 一元线性拟合

对组数据做一元线性拟合，假设变量与变量成线性关系。这里介绍两种本质上等价的求解方法。

**方法一：**求待定参数使误差平方和最小，由函数求极值的必要条件，应满足。计算如下：



即



化简得



解得

 (2.1)

 (2.2)

**方法二：矩阵形式的求解。**

将代入方程，得到矛盾方程组



令

，

方程组可写成



两边同时左乘，则有

****

可得



### 多元线性拟合

假设变量与个变量之间存在线性关系：。对进行次测量，的第次测量记作，对应的的测量值记作，由函数所得函数值记作，。则误差平方和为



同样要令误差平方和最小，则对求偏导



化简得方程组

， (2.3)

将测量值代入上式，即可得到未知参数。

以上线性模型的特点是事先假定输入变量与输出变量之间存在线性关系，根据输入变量的个数可分别采用一元或多元线性拟合，通过最小二乘法进行求解，可得到解析形式的最优解。但在现实中更多的问题中的变量可能是非线性关系，则要通过非线性模型进行拟合。下面我们介绍几种常用的非线性拟合形式，并详细说明其中如何利用最小二乘法进行求解。

### 非线性拟合

#### 多项式拟合

假定输入变量和输出变量之间的关系为多项式，则可通过多项式拟合求得待定系数，得到最优的拟合函数。

具体模型如下：对测量得到的组数据做多项式拟合，也就是求，使得最小。

模型求解：该问题同上述线性拟合一样转化为极值问题，则有

，

即

，

得到

，

上式是关于的线性方程组：

 (2.4)

用矩阵表示为：

 (2.5)

由此可以解出未知参数。

#### 可化为线性拟合的非线性拟合

除了上述介绍的拟合曲线外，还可以选取其他的与数据更为接近的曲线。一些非线性拟合的问题可以通过适当的变量替换转化为线性拟合问题，按照线性拟合的方式求解后还原为原变量。这样做可以简化模型求解的过程。

以指数函数拟合为例。

假设对于测量数据我们选取拟合函数，其中为待定参数。

对两边同时取自然对数有



令，

则有



此时就将拟合函数转化成了线性形式，利用2.2.1中一元线性拟合的方式即可求出，再由，可得。则求得拟合函数为。

另外几例可以化为线性拟合的函数及其变量替换替换如下表所示

表 1 非线性函数化为线性函数变量替换表

|  |  |  |
| --- | --- | --- |
| 曲线拟合函数 | 变量替换 | 变换后的线性形式 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

## 曲线拟合模型评价

在对测量数据进行曲线拟合后，我们可以依据自变量的取值来估计或预测因变量的取值。估计和预测的精准程度取决于测量数据紧密围绕拟合曲线的程度，测量数据越是接近拟合曲线，则拟合的效果越好。下面介绍反映拟合曲线对于观测数据拟合程度的判断定系数。

因变量取值的波动称为变差，变差的大小可以用实际测量值与其均值之差来表示，次测量值的总变差可以由离差的平方和来表示，称为总平方和，记为，即。

另外，估计值与均值的离差平方和记为，称为回归平方和，即；测量值与估计值的残差的平方和记为，称为残差平方和，即

总平方和可以分解为回归平方和与残差平方和相加，即。曲线拟合的效果取决于回归平方和占总平方和的比例大小，测量数据越接近拟合曲线，越大，则拟合得越好。回归平方和占总平方和的比例称为判定系数，记为，其计算公式为：

 (2.6)

越接近1，则曲线拟合效果越好。

# 最小二乘法的应用示例

前文中提到最小二乘法的应用十分广泛，下面给出了最小二乘法的四个具有代表性的应用示例，分别说明了在不同的问题中如何利用最小二乘法建立一元、多元线性模型，以及非线性模型，在各个模型中怎样进行求解，从而对输出变量进行拟合和预测。

## 在温度传感器中的应用示例

温度传感器是温度测量仪的重要部分，某温度传感器利用电流与温度的关系，将温度信号转换为电流信号输出。但由于温度传感器测量误差受到温度的影响，需要对测量的数据进行修正，再进行温度的精确标定，从而消除传感器的非线性误差。

试利用最小二乘法拟合实验数据，实验数据如下表。（下表数据来源于[19]）

表 2 传感器温度与电流数据表

| 温度传感器温度 | 输出电流 | 温度传感器温度 | 输出电流 |
| --- | --- | --- | --- |
| -15.0 | 260.6 | 40.0 | 315.5 |
| -10.0 | 265.8 | 45.0 | 319.6 |
| -5.0 | 266.3 | 50.0 | 324.8 |
| 0.0 | 272.8 | 55.0 | 327.7 |
| 5.0 | 278.6 | 60.0 | 335.6 |
| 10.0 | 285.8 | 65.0 | 341.2 |
| 15.0 | 287.4 | 70.0 | 345.2 |
| 20.0 | 296.7 | 80.0 | 353.9 |
| 25.0 | 298.6 | 90.0 | 362.6 |
| 30.0 | 302.7 | 100.0 | 373.4 |
| 35.0 | 306.7 |  |  |

**第一步：**绘制散点图。

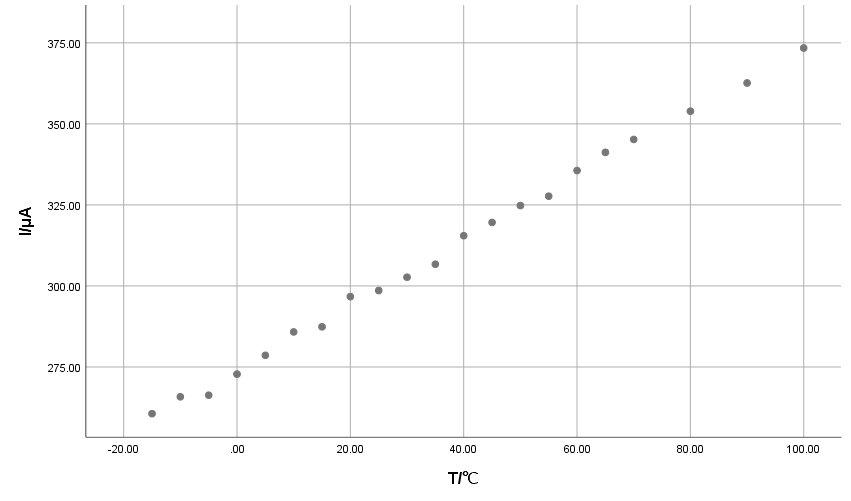


图 1 传感器温度与电流散点图

**第二步：**选择拟合函数。这里根据散点图与题中要求，选择一元线性拟合。

**第三步：**处理数据进行计算。

假设传感器温度为，输出电流为，用进行拟合，根据测量数据得到下表：

表 3 传感器数据计算表

|  |  |  |  |
| --- | --- | --- | --- |
| -15.0 | 260.6 | -3909 | 225 |
| -10.0 | 265.8 | -2658 | 100 |
| -5.0 | 266.3 | -1331.5 | 25 |
| 0.0 | 272.8 | 0 | 0 |
| 5.0 | 278.6 | 1393 | 25 |
| 10.0 | 285.8 | 2858 | 100 |
| 15.0 | 287.4 | 4311 | 225 |
| 20.0 | 296.7 | 5934 | 400 |
| 25.0 | 298.6 | 7465 | 625 |
| 30.0 | 302.7 | 9081 | 900 |
| 35.0 | 306.7 | 10734.5 | 1225 |
| 40.0 | 315.5 | 12620 | 1600 |
| 45.0 | 319.6 | 14382 | 2025 |
| 50.0 | 324.8 | 16240 | 2500 |
| 55.0 | 327.7 | 18023.5 | 3025 |
| 60.0 | 335.6 | 20136 | 3600 |
| 65.0 | 341.2 | 22178 | 4225 |
| 70.0 | 345.2 | 24164 | 4900 |
| 80.0 | 353.9 | 28312 | 6400 |
| 90.0 | 362.6 | 32634 | 8100 |
| 100.0 | 373.4 | 37340 | 10000 |

根据上表有

，，，，

由2.2.1一元线性拟合求得的式子(2.1)(2.2)，代入以上数据可得





得到电流与温度拟合的线性关系：。

由2.3中判定系数的公式(2.6)计算出，表明模型的误差很小，数据拟合程度很好，则可以利用最小二乘法拟合后的数据进行温度传感器的温度的标定。

## 在房地产销售价格中的应用示例

一家房地产评估公司想对某城市的房地产销售价格与地产估价、房地产估价和使用面积建立一个模型，以便对销售价格作出合理预测。为此，房地产评估公司收集了20栋住宅的房地产评估数据。试帮该公司建立模型。（下表数据来源于[20]）

表 4 房地产评估数据表

| 房地产编号 | 销售价格  （元/平方米） | 地产估价  （万元） | 房产估价  （万元） | 使用面积  （平方米） |
| --- | --- | --- | --- | --- |
| 1 | 6890 | 596 | 4497 | 18730 |
| 2 | 4850 | 900 | 2780 | 9280 |
| 3 | 5550 | 950 | 3144 | 11260 |
| 4 | 6200 | 1000 | 3959 | 12650 |
| 5 | 11650 | 1800 | 7283 | 22140 |
| 6 | 4500 | 850 | 2732 | 9120 |
| 7 | 3800 | 800 | 2986 | 8990 |
| 8 | 8300 | 2300 | 4775 | 18030 |
| 9 | 5900 | 810 | 3912 | 12040 |
| 10 | 4750 | 900 | 2935 | 17250 |
| 11 | 4050 | 730 | 4012 | 10800 |
| 12 | 4000 | 800 | 3168 | 15290 |
| 13 | 9700 | 2000 | 5851 | 24500 |
| 14 | 4550 | 800 | 2345 | 11510 |
| 15 | 4090 | 800 | 2089 | 11730 |
| 16 | 8000 | 1050 | 5625 | 19600 |
| 17 | 5600 | 400 | 2086 | 13440 |
| 18 | 3700 | 450 | 2261 | 9880 |
| 19 | 5000 | 340 | 3595 | 10760 |
| 20 | 2240 | 150 | 578 | 9620 |

第一步：绘制散点图。

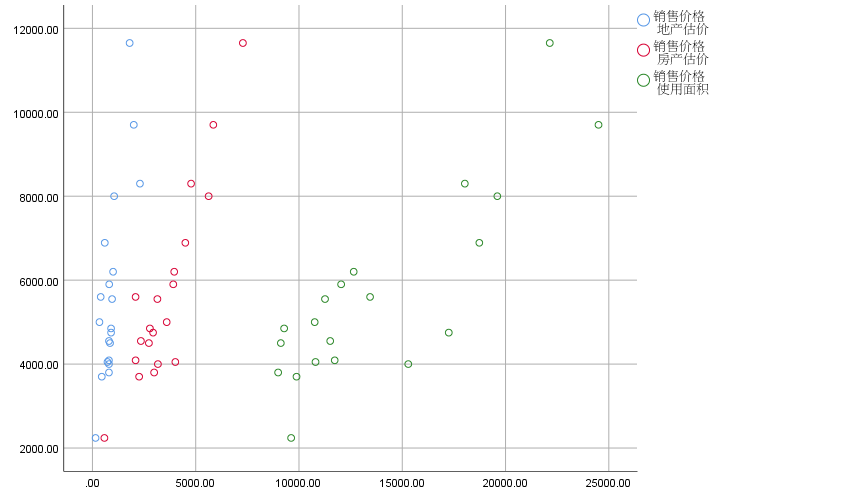


图 2 房地产销售价格与其他因素散点图

第二步：选择拟合函数。散点图大致呈线性关系，同时根据题意，这里选择多元线性拟合。

第三步：处理数据进行计算。

假设销售价格为，地产估价为，房产估价为，使用面积为，用进行拟合。

表 5 房地产数据处理表

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 6890 | 596 | 4497 | 18730 | 4106440 | 30984330 | 129049700 |
| 2 | 4850 | 900 | 2780 | 9280 | 4365000 | 13483000 | 45008000 |
| 3 | 5550 | 950 | 3144 | 11260 | 5272500 | 17449200 | 62493000 |
| 4 | 6200 | 1000 | 3959 | 12650 | 6200000 | 24545800 | 78430000 |
| 5 | 11650 | 1800 | 7283 | 22140 | 20970000 | 84846950 | 257931000 |
| 6 | 4500 | 850 | 2732 | 9120 | 3825000 | 12294000 | 41040000 |
| 7 | 3800 | 800 | 2986 | 8990 | 3040000 | 11346800 | 34162000 |
| 8 | 8300 | 2300 | 4775 | 18030 | 19090000 | 39632500 | 149649000 |
| 9 | 5900 | 810 | 3912 | 12040 | 4779000 | 23080800 | 71036000 |
| 10 | 4750 | 900 | 2935 | 17250 | 4275000 | 13941250 | 81937500 |
| 11 | 4050 | 730 | 4012 | 10800 | 2956500 | 16248600 | 43740000 |
| 12 | 4000 | 800 | 3168 | 15290 | 3200000 | 12672000 | 61160000 |
| 13 | 9700 | 2000 | 5851 | 24500 | 19400000 | 56754700 | 237650000 |
| 14 | 4550 | 800 | 2345 | 11510 | 3640000 | 10669750 | 52370500 |
| 15 | 4090 | 800 | 2089 | 11730 | 3272000 | 8544010 | 47975700 |
| 16 | 8000 | 1050 | 5625 | 19600 | 8400000 | 45000000 | 156800000 |
| 17 | 5600 | 400 | 2086 | 13440 | 2240000 | 11681600 | 75264000 |
| 18 | 3700 | 450 | 2261 | 9880 | 1665000 | 8365700 | 36556000 |
| 19 | 5000 | 340 | 3595 | 10760 | 1700000 | 17975000 | 53800000 |
| 20 | 2240 | 150 | 578 | 9620 | 336000 | 1294720 | 21548800 |
|  | 113320 | 18426 | 70613 | 276620 | 122732440 | 460810710 | 1737601200 |

表 6 房地产数据处理表

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | 355216 | 20223009 | 350812900 | 2680212 | 11163080 | 84228810 |
| 2 | 810000 | 7728400 | 86118400 | 2502000 | 8352000 | 25798400 |
| 3 | 902500 | 9884736 | 126787600 | 2986800 | 10697000 | 35401440 |
| 4 | 1000000 | 15673681 | 160022500 | 3959000 | 12650000 | 50081350 |
| 5 | 3240000 | 53042089 | 490179600 | 13109400 | 39852000 | 161245620 |
| 6 | 722500 | 7463824 | 83174400 | 2322200 | 7752000 | 24915840 |
| 7 | 640000 | 8916196 | 80820100 | 2388800 | 7192000 | 26844140 |
| 8 | 5290000 | 22800625 | 325080900 | 10982500 | 41469000 | 86093250 |
| 9 | 656100 | 15303744 | 144961600 | 3168720 | 9752400 | 47100480 |
| 10 | 810000 | 8614225 | 297562500 | 2641500 | 15525000 | 50628750 |
| 11 | 532900 | 16096144 | 116640000 | 2928760 | 7884000 | 43329600 |
| 12 | 640000 | 10036224 | 233784100 | 2534400 | 12232000 | 48438720 |
| 13 | 4000000 | 34234201 | 600250000 | 11702000 | 49000000 | 143349500 |
| 14 | 640000 | 5499025 | 132480100 | 1876000 | 9208000 | 26990950 |
| 15 | 640000 | 4363921 | 137592900 | 1671200 | 9384000 | 24503970 |
| 16 | 1102500 | 31640625 | 384160000 | 5906250 | 20580000 | 110250000 |
| 17 | 160000 | 4351396 | 180633600 | 834400 | 5376000 | 28035840 |
| 18 | 202500 | 5112121 | 97614400 | 1017450 | 4446000 | 22338680 |
| 19 | 115600 | 12924025 | 115777600 | 1222300 | 3658400 | 38682200 |
| 20 | 22500 | 334084 | 92544400 | 86700 | 1443000 | 5560360 |
|  | 22482316 | 294242295 | 4236997600 | 76520592 | 287615880 | 1083817900 |

计算得到：，，，，，，，，，，，，

由2.2.2多元线性拟合求得的方程组(2.3)，有

，

即



代入数据得到方程组



解得



由此建立了该城市的房地产销售价格与地产估价、房产估价和使用面积的模型：，从而可以对该城市房地产销售价格进行预测。

由2.3中判定系数的公式(2.6)计算出，表明拟合的效果较好，基本可以用该多元线性模型解释值的变差，因此可以利用该模型作为该城市房地产销售价格预测的参考。

## 在销售预测中的应用示例

某品牌一畅销的紧凑型汽车近十年销售量如下表所示，试预测该汽车2020年的销售量。（本数据来源于车主之家网站：https://xl.16888.com/s/57509/）

表 7 汽车近十年销量数据表

|  |  |
| --- | --- |
| 年份 | 销量（辆） |
| 2010 | 172053 |
| 2011 | 170117 |
| 2012 | 151887 |
| 2013 | 146480 |
| 2014 | 171486 |
| 2015 | 254301 |
| 2016 | 306541 |
| 2017 | 333488 |
| 2018 | 376077 |
| 2019 | 348840 |

**第一步：**画散点图。

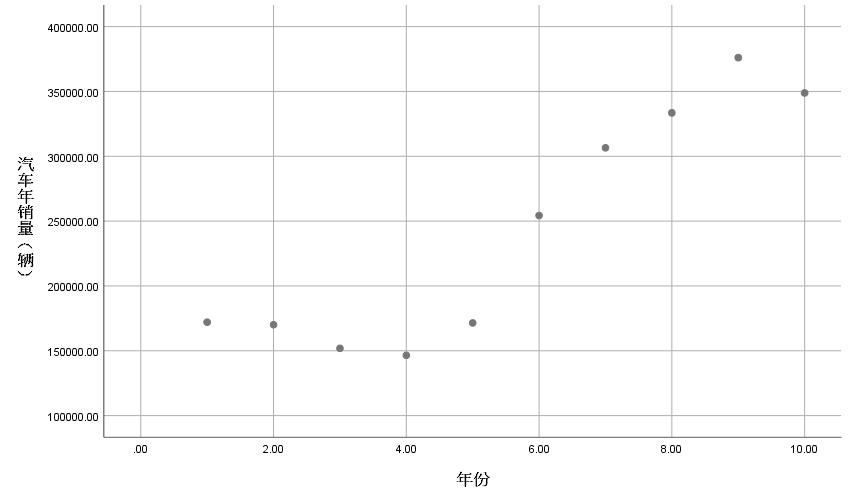


图 3 汽车近十年销量散点图

**第二步：**选择拟合函数。

根据散点图，我们选取多项式拟合。由于不确定用几次多项式拟合的效果更好，因此对数据进行二次和三次的多项式拟合，再对二次和三次的多项式拟合图像和进行对比，最后选取更为合适的拟合曲线。

**第三步：**处理数据进行计算。

下面以三次多项式拟合的计算为例。

以2010年为第一年，则2019年为第十年，用数字1-10表示年份。假设年份为，销量为，拟合曲线，需要求出参数。

表 8 汽车销量数据处理表

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 1 | 1 | 1 | 1 | 1 | 1 | 172053 | 172053 | 172053 | 172053 |
| 1 | 2 | 4 | 8 | 16 | 32 | 64 | 170117 | 340234 | 680468 | 1360936 |
| 2 | 3 | 9 | 27 | 81 | 243 | 729 | 151887 | 455661 | 1366983 | 4100949 |
| 3 | 4 | 16 | 64 | 256 | 1024 | 4096 | 146480 | 585920 | 2343680 | 9374720 |
| 4 | 5 | 25 | 125 | 625 | 3125 | 15625 | 171486 | 857430 | 4287150 | 21435750 |
| 5 | 6 | 36 | 216 | 1296 | 7776 | 46656 | 254301 | 1525806 | 9154836 | 54929016 |
| 6 | 7 | 49 | 343 | 2401 | 16807 | 117649 | 306541 | 2145787 | 15020509 | 105143563 |
| 7 | 8 | 64 | 512 | 4096 | 32768 | 262144 | 333488 | 2667904 | 21343232 | 170745856 |
| 8 | 9 | 81 | 729 | 6561 | 59049 | 531441 | 376077 | 3384693 | 30462237 | 274160133 |
| 9 | 10 | 100 | 1000 | 10000 | 100000 | 1000000 | 348840 | 3488400 | 34884000 | 348840000 |

由上表可得

，，，，，

，，，，



根据2.2.3.1多项式拟合求得的方程组(2.4)(2.5)，代入数据，得到正规方程组



解得



得到拟合曲线：

同样的方法，可以得到二次拟合曲线：

**第四步：**拟合曲线对比。

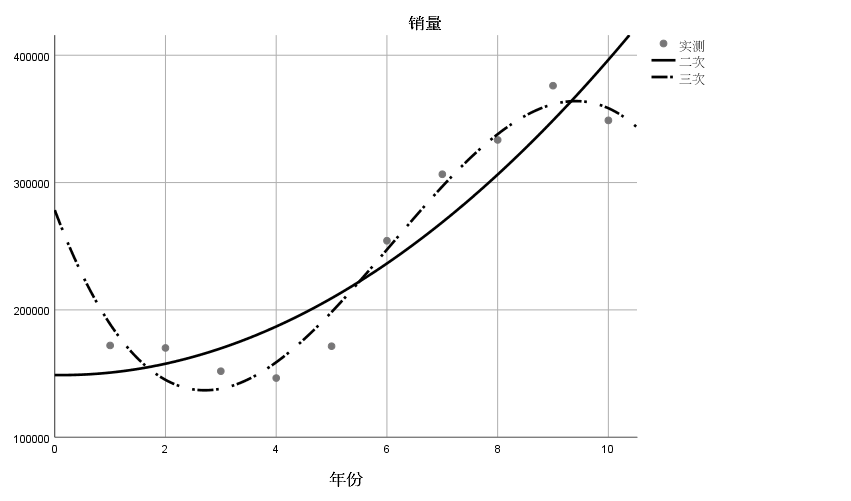


图 4 拟合曲线对比图

对数据进行二次和三次的多项式拟合得到拟合曲线对比图。由2.3中判定系数的公式(2.6)计算出二次多项式拟合的，三次多项式拟合的。从对比图和判定系数都可以得到三次多项式拟合更贴合汽车的销售数据，误差更小。因此选择三次多项式拟合。

**第五步：**代入数据预测。

将代入三次多项式拟合曲线，得，因此预测该汽车在2020年的销量为319599辆。

此例也说明了在不确定哪种模型拟合效果更优时，可以通过采用不同的模型拟合进行试错。

## 在传染病人数预测中的应用示例

2020年我国爆发了新冠肺炎疫情，其中湖北省情况最为严重。试根据湖北省每日现存确诊人数，大致预测湖北省疫情结束时间。（下表的数据来源于丁香园网站：ncov.dxy.cn每日公布的数据。）

表 9 湖北省每日现存确诊人数数据表

|  |  |  |  |
| --- | --- | --- | --- |
| 日期 | 现存确诊人数 | 日期 | 现存确诊人数 |
| 1月20日 | 239 | 2月17日 | 50338 |
| 1月21日 | 338 | 2月18日 | 50633 |
| 1月22日 | 399 | 2月19日 | 49665 |
| 1月23日 | 494 | 2月20日 | 48730 |
| 1月24日 | 658 | 2月21日 | 47647 |
| 1月25日 | 958 | 2月22日 | 46439 |
| 1月26日 | 1303 | 2月23日 | 45054 |
| 1月27日 | 2567 | 2月24日 | 43369 |
| 1月28日 | 3349 | 2月25日 | 41660 |
| 1月29日 | 4334 | 2月26日 | 39755 |
| 1月30日 | 5486 | 2月27日 | 36829 |
| 1月31日 | 6738 | 2月28日 | 34715 |
| 2月1日 | 8565 | 2月29日 | 32959 |
| 2月2日 | 10532 | 3月1日 | 30543 |
| 2月3日 | 12712 | 3月2日 | 28216 |
| 2月4日 | 15679 | 3月3日 | 25905 |
| 2月5日 | 18483 | 3月4日 | 24085 |
| 2月6日 | 20677 | 3月5日 | 22695 |
| 2月7日 | 23139 | 3月6日 | 21239 |
| 2月8日 | 24881 | 3月7日 | 19710 |
| 2月9日 | 26965 | 3月8日 | 18303 |
| 2月10日 | 28532 | 3月9日 | 17151 |
| 2月11日 | 29659 | 3月10日 | 15671 |
| 2月12日 | 43455 | 3月11日 | 14427 |
| 2月13日 | 46429 | 3月12日 | 13171 |
| 2月14日 | 48175 | 3月13日 | 11772 |
| 2月15日 | 49030 | 3月14日 | 10431 |
| 2月16日 | 49847 | 3月15日 | 9605 |

**第一步：**画散点图。

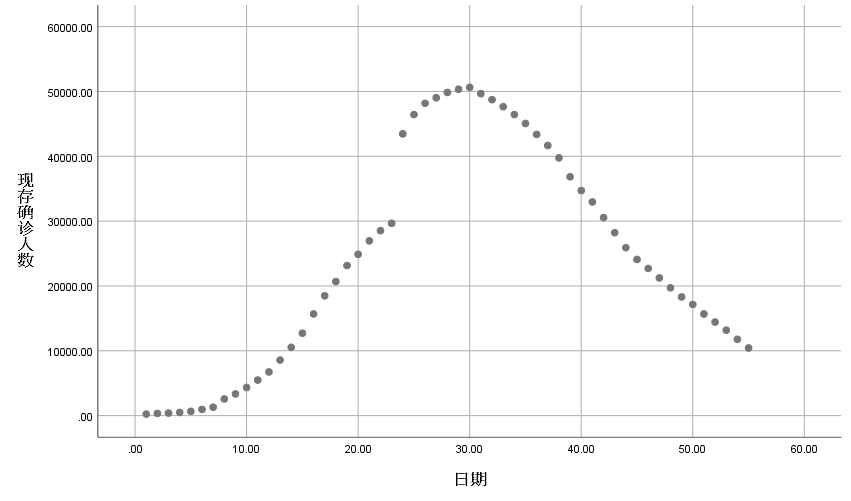


图 5 湖北省每日现存确诊人数散点图

**第二步：**选择拟合函数。

从散点图可以看到，湖北省现存确诊人数从数据的第三十天，即2月18日开始呈下降趋势。因此我选取后半部分数据进行拟合。从趋势上看，选取指数函数拟合。

**第三步：**处理数据进行计算。

从2月18日开始算第一天。设日期为，现存确诊人数为，拟合函数为。

表 10 湖北省每日现存确诊人数数据处理表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1 | 50633 | 10.83235882 | 10.83235882 | 1 |
| 2 | 49665 | 10.81305574 | 21.62611148 | 4 |
| 3 | 48730 | 10.79405014 | 32.38215041 | 9 |
| 4 | 47647 | 10.77157495 | 43.08629979 | 16 |
| 5 | 46439 | 10.7458949 | 53.72947451 | 25 |
| 6 | 45054 | 10.71561705 | 64.2937023 | 36 |
| 7 | 43369 | 10.67750018 | 74.74250125 | 49 |
| 8 | 41660 | 10.63729671 | 85.09837372 | 64 |
| 9 | 39755 | 10.5904909 | 95.31441809 | 81 |
| 10 | 36829 | 10.51404086 | 105.1404086 | 100 |
| 11 | 34715 | 10.45492715 | 115.0041986 | 121 |
| 12 | 32959 | 10.40301964 | 124.8362357 | 144 |
| 13 | 30543 | 10.32689081 | 134.2495805 | 169 |
| 14 | 28216 | 10.24764447 | 143.4670226 | 196 |
| 15 | 25905 | 10.16219128 | 152.4328692 | 225 |
| 16 | 24085 | 10.08934452 | 161.4295123 | 256 |
| 17 | 22695 | 10.02989991 | 170.5082986 | 289 |
| 18 | 21239 | 9.963594393 | 179.3446991 | 324 |
| 19 | 19710 | 9.8888814 | 187.8887466 | 361 |
| 20 | 18303 | 9.81482026 | 196.2964052 | 400 |
| 21 | 17151 | 9.74981176 | 204.746047 | 441 |
| 22 | 15671 | 9.65956715 | 212.5104773 | 484 |
| 23 | 14427 | 9.57685673 | 220.2677048 | 529 |
| 24 | 13171 | 9.485772722 | 227.6585453 | 576 |
| 25 | 11772 | 9.373479109 | 234.3369777 | 625 |
| 26 | 10431 | 9.252537421 | 240.5659729 | 676 |
| 27 | 9605 | 9.170039075 | 3739.380147 | 729 |

计算可得

，，，

，

根据2.2.3.2中的方法将指数函数转化为一元线性函数求解，代入2.2.1一元线性拟合中求得的公式中，得





还原为原变量



从而得到现存确诊人数的拟合曲线

接着我们绘制出拟合函数的图像

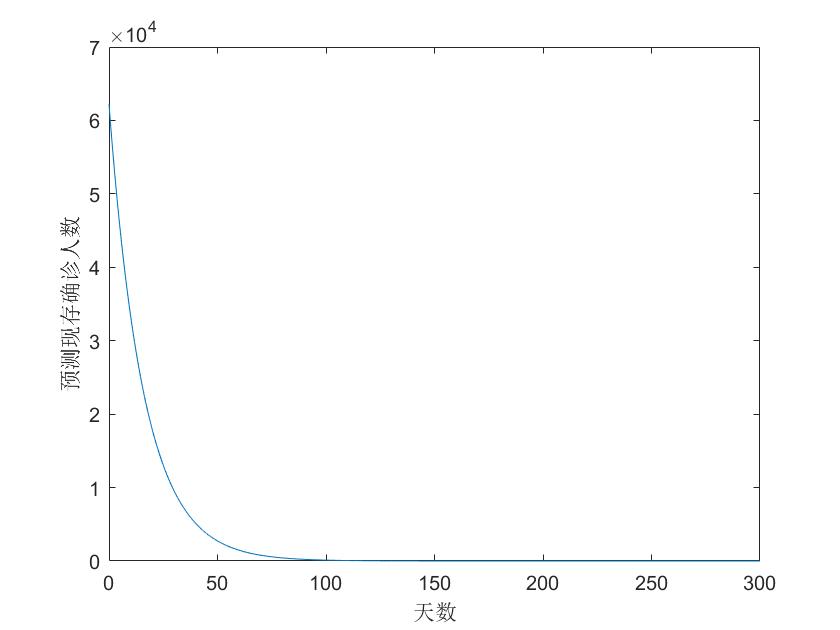


图 6 拟合曲线图像

从拟合图像可以观察到，到第70天时，湖北省仍患病人数接近为零，因此预测第70天左右（即四月底）湖北省疫情能够基本结束。

可见本模型的预测结果基本上是合理的，这是由于模型所依赖的数据隐含了病毒自身的发展规律，以及政府的强力措施下的人为干预产生的效果的叠加，当我们选择了合适的拟合形式，利用最小二乘法这样简单的工具就可以得到大致的发展趋势。当然，实际情况往往更为复杂，比如，我们尚未考虑境外输入病例对于确诊人数的影响。因此，在建立模型的过程中，要实时更新信息和数据是至关重要的。

# 结论

本文首先回顾了最小二乘法发现的历史，阐述了最小二乘法的基本原理并给出了最小二乘法的常见拟合形式，最后进一步给出了最小二乘法在不同方面的应用示例。具体地，我们进行了以下工作：

1. 较为全面地归纳了最小二乘曲线拟合的不同形式，包括一元线性拟合、多元线性拟合、多项式拟合以及一些可以化成线性形式的非线性拟合。详细给出了计算的过程，得出了计算结果。探讨了不同形式的特点以及它们之间的联系。
2. 给出了最小二乘法在拟合函数及预测中的应用示例。详细地说明了解决方式，利用不同形式曲线拟合的过程。在不同方面的应用示例，使大家更为清晰地了解到最小二乘法的具体应用，以及在模型建立和求解过程中应该注意的问题。
3. 在研究曲线拟合的过程中发现，在进行曲线拟合时，应当根据已有数据的散点图的分布趋势，结合不同的函数图像，选择最合适的拟合曲线，否则拟合结果会与实际相差较大。其次，在进行多项式拟合时，并不是越高次数的多项式拟合效果越好，有时较低次的多项式更能拟合到预期效果。因此，在实际拟合过程中，可以尝试不同的拟合方式，选取拟合最优的形式。

综上，本文不仅详细的介绍了最小乘法的原理及曲线拟合的形式，还具体阐释了所列举的相关示例，以便于人们对最小二乘法有一个全面的理解，从而可以在实际生活中将其进行更好运用。通过前文介绍可以发现，最小二乘法在数理统计学中占据着至关重要的地位，它已经深入到科学研究与社会生活的许多个方面。在科学研究中，我们会遇到许多求解变量之间的关系问题，也会遇到许多需要预测的问题，最小二乘法曲线拟合恰好能很好地解决这些问题，若是能够对曲线拟合进行灵活的运用，那么不管是对科学研究还是对我们的日常生活，都将带来很大的便利。事实上，在物理、工程技术、测绘、经济、营销等各领域，最小二乘法都得到了广泛的应用。本文仅讲述了最小二乘法的一些简单示例。今后我们将更加深入地研究它的原理以及改进方式，以探索最小二乘法更为广泛及精准的应用。

# 参考文献

1. 陆健. 最小二乘法及其应用[J]. 中国西部科技, 2007, 000(019).
2. 陈希孺. 最小二乘法的历史回顾与现状[J]. 中国科学院研究生院学报, 1998, 015(001):4-11.
3. 贾小勇, 徐传胜, 白欣. 最小二乘法的创立及其思想方法[J]. 西北大学学报（自然科学版）, 2006, 036(003):507-511.
4. 邓亮章. 最小二乘法的拟合及其应用[J]. 兰州教育学院学报, 2012(08):113-114+135.
5. 刘佳. 最小二乘法基本思想及其应用[J]. 科技视界, 2016, 000(022):186-187.
6. 李蓓蕾. 多次自适应最小二乘曲线拟合方法及其应用[D]. 长江大学, 2014.
7. 张东林. 分段最小二乘曲线拟合[J]. 沈阳大学学报：自然科学版(2期):80-83.
8. 侯超钧, 曾艳姗, 吴东庆, et al. 全局连续的分段最小二乘曲线拟合方法[J]. 重庆师范大学学报（自然科学版）(6):50-54.
9. 丁克良,欧吉坤,赵春梅.正交最小二乘曲线拟合法[J].测绘科学, 2007(03): 18-19+17+192.
10. P. Lancaster and K. Salkauskas. Surfaces Generated by Moving Least Squares Methods[J]. Mathematics of Computation, 1981, 37(155):141-158.
11. Mukherjee Y X , Mukherjee S . On boundary conditions in the element-free Galerkin method[J]. Computational Mechanics, 1997, 19(4):264-270.
12. 陈美娟, 程玉民. 改进的移动最小二乘法[J]. 力学季刊, 2003(02):120-126.
13. 任红萍,程玉民,张武.改进的移动最小二乘插值法研究[J].工程数学学报, 2010, 27(06):1021-1029.
14. 曹可,吕继续.基于最小二乘法预测传染病的发病人数[J].科技资讯, 2019, 17(33):228-229.
15. 王桂杰,焦良葆,曹雪虹.基于最小二乘法的室内三维定位算法研究[J/OL].计算机技术与发展,2020(04):1-7[2020-03-17].
16. 李大鹏,朱平平,陈士强,潘辉,李若全.线性回归分析在液体火箭推进剂温度预测中的应用[J].导弹与航天运载技术,2020(01):43-47.
17. 王知雨,王斌,王朝晖.采用非线性最小二乘法的超级电容等效电路模型参数辨识[J/OL].西安交通大学学报, 2020(04): 1-9[2020-03-17]. http://kns.cnki.net/kcms/detail/61.1069.T.20200111.1717.002.html.
18. 王辅臣,马强,陶冬旺,刘赫奕.一种实时场地校正方法及其在地震预警中的应用[J].地震工程与工程振动,2020,40(01):171-184.
19. Ayse Gul Kaplan,Yusuf Alper Kaplan. Developing of the new models in solar radiation estimation with curve fitting based on moving least-squares approximation[J]. Renewable Energy,2020,146.
20. Czaplewski K, Specht C, Dąbrowski P, et al. Use of a Least Squares with Conditional Equations Method in Positioning a Tramway Track in the Gdansk Agglomeration[J]. TransNav: International Journal on Marine Navigation & Safety of Sea Transportation, 2019, 13(4):895-900. doi:10.12716/1001.13.04.25.
21. Asadpour, S., Hosseinzadeh, H., & Yazdani, A. Numerical Solution of the Lane-Emden Equations with Moving Least Squares Method[J]. Applications & Applied Mathematics, 2019, 14(2):762–776.
22. Gautam C, Mishra PK, Tiwari A, et al. Minimum variance-embedded deep kernel regularized least squares method for one-class classification and its applications to biomedical data[J]. Neural networks : the official journal of the International Neural Network Society. 2020,123:191-216. doi:10.1016/j.neunet.2019.12.001.
23. 莫小琴.基于最小二乘法的线性与非线性拟合[J].无线互联科技, 2019, 16(04):128-129.
24. 商继敏,王海燕,蒋逢春,魏茂才,陈鹏.最小二乘法对温度传感器测温数据线性拟合及其应用[J].大学物理实验,2019,32(02):81-84.
25. 贾俊平, 何晓群, 金勇进. 统计学[M]. 第七版. 北京: 中国人民大学出版社, 2018:248-249+283

# 致 谢

首先要感谢我的指导老师李艳老师，感谢她对我论文撰写过程中无微不至的关心以及的悉心指导。从论文的选题，写作框架到最后的定稿都离不开导师的细心指导。李艳老师以她渊博的学识，严谨的治学态度，诲人不倦的崇高师德，平易近人的人格魅力给我树立了一个很好的榜样，无论是在学习和生活中都深深地感染和激励着我。

感谢应用数学学院的老师和同学们，这四年的本科生涯离不开各位老师的教诲和同学们的帮助，在学习方式和为人处世上我都收获颇丰。

数院的每个老师都拥有他们独特的人格魅力，都很关心和爱护学生。感谢我的专业课老师，让我能够扎实地学好专业知识。感谢班主任谭洋老师，从大一开始就一直督促我们做好生涯规划，多方面地关心我们，让我们在一个很优秀的班级里度过了愉快的大学生活。

感谢我的室友刘永银、周彤、关晓桐，在生活上给予了我很多的关心，同时也在宿舍里营造了很好的学习氛围，四年里，相互督促，共同进步。感谢陈琛经常拉我一起吃饭。感谢社团组织的学长学姐和同学们使我的大学生活更加充实快乐。

最后非常感谢我的家人，感谢爸爸妈妈爷爷奶奶对我的关心和支持，让我无忧无虑地度过本科四年美好的时光。

衷心感谢所有在我人生道路上给予我鼓励和帮助的人，您们的帮助是我继续不断努力向前的动力。