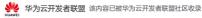
**渝 胤风** ( 关注

置顶 胤风 🗧 已于 2022-05-27 15:57:32 修改 💿 阅读量10w+ 🏚 收藏 3.2k 💧 点赞数 986 分类专栏: 数学知识 文章标签: 线性代数 矩阵 机器学习 深度学习 概率论





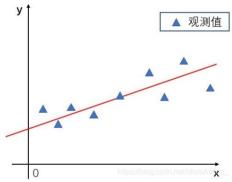
# 188 订阅 13 篇文章

#### 要解决的问题

在工程应用中,我们经常会用一组观测数据去估计模型的参数,模型是我们根据先验知识定下的。比如我们有一组观测数据 $(x_i, y_i)$ (一维),通过一 析我们猜测y和x之间存在线性关系,那么我们的模型就可以定为: f(x) = kx + b

这个模型只有两个参数,所以理论上,我们只需要观测两组数据建立两个方程,即可解出两个未知数。类似的,假如模型有n个参数,我们只需要观测 可求出参数,换句话说,在这种情况下,模型的参数是唯一确定解。

但是在实际应用中,由于我们的观测会存在误差(偶然误差、系统误差等),所以我们总会做多余观测。比如在上述例子中,尽管只有两个参数,但 会观测n组数据 $(x_1,y_1)...,(x_n,y_n)$ ,这会导致我们无法找到一条直线经过所有的点,也就是说,方程无确定解。



于是这就是我们要解决的问题:虽然没有确定解,但是我们能不能求出近似解,使得模型能在各个观测点上达到"最佳"拟合。那么"最佳"的准则是什么 所有观测点到直线的距离和最小,也可以是所有观测点到直线的误差(真实值-理论值)绝对值和最小,也可以是其它,如果是你面临这个问题你会怎

早在19世纪,勒让德就认为让"误差的平方和最小"估计出来的模型是最接近真实情形的。

为什么就是误差平方而不是其它的,这个问题连欧拉、拉普拉斯都未能成功回答,后来是高斯建立了一套误差分析理论,从而证明了确实是使误差平, 情况下系统是最优的。理论的证明也并不难,我写在了另外一篇博客 最小二乘法的原理理解,相信你了解后会对最小二乘法有更深刻的认识。

按照勒让德的最佳原则,于是就是求:

$$L = \sum_{i=1}^n \left(y_i - f(x)\right)^2$$

这个目标函数取得最小值时的函数参数,这就是最小二乘法的思想,所谓"二乘"就是平方的意思。从这里我们可以看到,**最小二乘法其实就是用来做**图 一种思想。

至于怎么求出具体的参数那就是另外一个问题了,理论上可以用导数法、几何法,工程上可以用梯度下降法。下面以最常用的线性回归为例进行推导

#### 线性回归

**4** 986 **4** 

线性回归因为比较简单,可以直接推导出解析解,而且许多非线性的问题也可以转化为线性问题来解决,所以得到了广泛的应用。甚至许多人认为最小 的就是线性回归,其实并不是,最小二乘法就是一种思想,它可以拟合任意函数,线性回归只是其中一个比较简单而且也很常用的函数,所以讲最小 都会以它为例。







$$\begin{split} h_1 &= \theta_0 + \theta_1 \, x_{1,1} + \theta_2 \, x_{1,2} + \ldots + \theta_{n-1} \, x_{1,n-1} \\ h_2 &= \theta_0 + \theta_1 \, x_{2,1} + \theta_2 \, x_{2,2} + \ldots + \theta_{n-1} \, x_{2,n-1} \\ \vdots \\ h_m &= \theta_0 + \theta_1 \, x_{m,1} + \theta_2 \, x_{m,2} + \ldots + \theta_{n-1} \, x_{m,n-1} \end{split}$$

为方便用矩阵表示,我们令 $x_0 = 1$ ,于是上述方程可以用矩阵表示为

$$\mathbf{h} = \mathbf{X}\mathbf{\theta}$$

其中,h为mx1的向量,代表模型的理论值,创为nx1的向量,X为mxn维的矩阵,m代表样本的个数n代表样本的特征数,于是目标损失函数用矩阵表

$$J(\theta) = \|\mathbf{h} - \mathbf{Y}\|^2 = \|\mathbf{X}\theta - \mathbf{Y}\|^2 = (\mathbf{X}\theta - \mathbf{Y})^{\mathrm{T}}(\mathbf{X}\theta - \mathbf{Y})$$

其中Y是样本的输出向量, 维度为mx1。

根据高数知识我们知道函数取得极值就是导数为0的地方,所以我们只需要对损失函数求导令其等于0就可以解出 $\theta$ 。矩阵求导属于矩阵微积分的内容 学的(.... 这里先介绍两个用到的公式:

$$\frac{\partial \mathbf{x}^{\mathrm{T}} \mathbf{a}}{\partial \mathbf{x}} = \frac{\partial \mathbf{a}^{\mathrm{T}} \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}$$

$$\frac{\partial x^T A x}{\partial x} = A x + A^T x$$

如果矩阵A是对称的:

$$Ax + A^Tx = 2Ax$$

对目标函数化简:

$$J(\theta) = \theta^T X^T X \theta - \theta^T X^T Y - Y^T X \theta + Y^T Y$$

求导令其等于0:

$$\frac{\partial}{\partial \theta} J(\theta) = 2X^T X \theta - 2X^T Y = 0$$

解得  $\theta = \left(X^TX\right)^{-1}X^TY$ ,经过推导我们得到了 $\theta$ 的解析解,现在只要给了数据,我们就可以带入解析解中直接算出 $\theta$ 。

## 几何意义

几何意义会直观的帮助你理解最小二乘法究竟在干什么。首先先来解释一下矩阵乘法的几何意义,对于一个方程组Ax,我们可以看做是x对矩阵A的 性组合,比如:

$$\left\{\begin{array}{ll} 1\times x_1+x_2=3 \\ -1\times x_1+x_2=1 \end{array} \Leftrightarrow \left[\begin{array}{cc} 1 & 1 \\ -1 & 1 \end{array}\right] \left[\begin{array}{c} x_1 \\ x_2 \end{array}\right] = \left[\begin{array}{c} 3 \\ 1 \end{array}\right] \Leftrightarrow A\times x=b$$

可以看作:

$$\left[\begin{array}{c}1\\-1\end{array}\right]\times x_1+\left[\begin{array}{c}1\\1\end{array}\right]\times x_2=\left[\begin{array}{c}3\\1\end{array}\right]\Leftrightarrow a_1\times x_1+a_2\times x_2=b$$

画在坐标轴上可以看到,向量b其实就是向量 $a_1$ 与 $a_2$ 的线性组合,因为他们都是在一个平面上,显然是有解的。



986

♠ CSDN 博客 下载 学习新 社区 C知道 GitCode InsCode 会议





2 https://bl3q.csdn.n知初4@Eureka

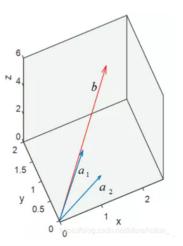
但是如文章开头所说,由于存在观测误差,我们往往会做多余观测,比如要拟合一次方程 y = kx + b,我们可能观测了三个点(0.2),(1.2),(2.4) 矩阵形式如下(为表述方便,用x1代替k, x2代替b):

$$\left\{\begin{array}{ll} 1\times x_1+x_2=2\\ 0\times x_1+x_2=2\\ 2\times x_1+x_2=3 \end{array} \right. \Leftrightarrow \left[\begin{array}{c} 1 & 1\\ 0 & 1\\ 2 & 1 \end{array}\right] \left[\begin{array}{c} x_1\\ x_2 \end{array}\right] = \left[\begin{array}{c} 2\\ 2\\ 3 \end{array}\right] \Leftrightarrow A\times x=b$$

表示成线性组合的方式::

$$\left[\begin{array}{c}1\\0\\2\end{array}\right]\times x_1+\left[\begin{array}{c}1\\1\\1\end{array}\right]\times x_2=\left[\begin{array}{c}2\\2\\3\end{array}\right]\Leftrightarrow a_1\times x_1+a_2\times x_2=b$$

画在图中如下:



从图中我们可以看到,无论  $a_1$  和  $a_2$  怎么线性组



换句话说,方程组Ax = b虽然无解,也就是b不在A的列空间中,但是我们可以在A的列空间中找到一个和b最接近的向量p,p就是b在A的列空间中 诵讨求Ax = p的解,就是原方程的最小二乘解。

由几何意义可知垂线e = b - p = b - Ax正交于平面 S,也就是 $a_1^T e = 0$ , $a_2^T e = 0$ ,写成矩阵形式:

$$A^{T}e = A^{T}(b - Ax) = A^{T}b - A^{T}Ax = 0$$

解得  $\mathbf{x} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{b}$ ,可以看到推导结果和矩阵法一样。从上面可以看到,最小二乘法的几何意义就是**求解 \mathbf{b} 在A的列向量空间中的投影**。 到这里最小一乘法的推导已经完成了,但是我们忽略了一个问题,就是假如A<sup>T</sup>A不可逆怎么办?这个问题我写在了另外一篇文章 详解岭回归与L2下I 以上就是全部内容。

#### 相关推荐阅读:

最小二乘法的原理理解 详解岭回归与L2正则化

另外推荐一篇关于主成成分分析PCA写得也很不错的文章: 通俗易懂的PCA原理及代码实现(超详细推导)

#### Reference

https://www.cnblogs.com/pinard/p/5976811.html

https://zhuanlan.zhihu.com/p/38128785

https://www.zhihu.com/question/304164814/answer/549972357

如果对你有帮助, 请点个赞计我知道:-D

#### 文章知识点与官方知识档案匹配,可进一步学习相关知识

算法技能树 首页 概览 61685 人正在系统学习中

#### ES一站式日志分析训练营

鹅厂大牛带你30分钟玩转ES,精美好礼、证书等你来拿!

53 条评论



最小二乘法,残差,线性模型-线性回归 最小二乘法与残差

大佬姓同山市 <del>里小一条注</del>辞目进图探到\_\_\_ 冬古佬 体和方径未到古佬 L 的硕庄呢窗今: 基于均方误差最小化来进行模型求解的方法称为"最小二季







♠ CSDN 博客 下载 学习新 社区 C知道 GitCode InsCode 会议

极摇奥卡姆利力原理.我们选择一次多项式作为拟台图数是最台语的。 4、<u>最小一架法</u>与止念分布.我们对朝止德的猜测.即最小<u>来法</u>.仍然把有外辈.力一这个猜测是销民的

半小时学习最小 一乘法

这里是我的个人网站: https://endlesslethe.com/easy-to-learn-ols.html 有更多总结分享,最新更新也只会发布在我的个人网站上。排版也可能会更好看一点=v= 前言最小

#### 最小 = 乘法的基本原理

最小二乘法的基本原理和多项式拟合 — 最小二乘法的基本原理□ 从整体上考虑近似函数 同所给数据点 (i=0.1....m)误差......

#### 一文看懂最小一乘法

最小二乘法的基本原理介绍.doc 09-05 最小二乘法(又称最小平方法)是一种数学优化技术。它通过最小化误差的平方和寻找数据的最佳函数匹配。利用最小二乘法可以简便

#### 最小二乘法 最小二乘法公式

最常用的是普通最小二乘法(Ordinary Least Square, OLS):所选择的回归模型应该使所有观察值的残差平方和达到最小。(Q为残差平方和)- 即采用平方损失函数。 样本回归

最小二乘法原理理解 热门推荐

小楼吹彻玉

概念:最小二乘法是一种熟悉而优化的方法。主要是通过最小化误差的平方以及最合适数据的匹配函数。 作用: (1) 利用最小二乘法可以得到位置数据(这些数据与实际数据

最小二乘法笔记 最新发布

weixin 41652700f5

资料: https://zhuanlan.zhihu.com/p/653406161。

#### 【机器学习】9、最小二乘法和岭回归 最小二乘回归

一、最小二乘法(线性回归) 1.1 公式最小二乘法的本质 公式推导过程 1.2 矩阵 假设我们有n个样本数据.每个数据有p个特征值.然后p个特征值是线性关系。即对应的线性核

#### 最优化方法——最小二乘法与梯度下降法 最小二乘法优化

1.最小二乘法 1.1 线性回归定义与算法步骤 1.2 最小二乘法的应用 2.梯度下降法 2.1 BP神经网络 2.2 梯度下降法的应用 七、实验小结 1.最小二乘问题求解总结 2.参考资料

#### 最小二乘法(LSM)入门详解(原理及公式推导),MATLAB实现及应用

weixin 40857506的

1 概念 定义:最小二乘法(又称最小平方法)是一种数学优化技术。它通过最小化误差的平方和寻找数据的最佳函数匹配。 利用最小二乘法可 以简便地求得未知的数据

#### 最小 = 乘法计管里孩身高

计算与预测男孩女孩身高模型,在本次综合实验中我计算的是男孩身高模型,男孩模型为: a0+a1x1+a2x2=y1; 其中a0、a1、a2均为常数项,而x1为设为父亲身高, x2设

#### 最小二乘 最小二乘法范数

如果将最小二乘法中向量的维数设置为2.即(xi,1),拟合函数为y=ax+b.参数theta为2维.化简后的公式就是我们在书本上常见的直线拟合公式。 4.2曲线拟合 如果将最小二乘污

#### 递归最小二乘算法(原理篇) 递归最小二乘法

最小二乘法.也称最小平方法.即计算误差平方和最小.得到的最佳估计。 核心问题.最小二乘估计的合理性证明是什么?数学王子高斯(1777-1855)也像我们一样心存怀疑。高

## 最小二乘法的思路及推导过程

weixin 40255714的

本文介绍了最小二乘法的基本思想,以及如何推导出最小二乘法的拟合公式

#### 最小二乘公式推导(步骤详细,一看就会)

上网找了关于最小二乘公式的<mark>推导</mark>,发现有的文章写的过于简略,有的<mark>推导</mark>虽然步骤<del>详细</del>但是其中<mark>矩阵</mark>、向量啥的命名不按常理出牌(比如把元素都是观测量x的<mark>矩阵</mark>叫做

#### 最小二乘法原理的几何解释

本文从代数几何的角度来解析分析在线性回归中使用的最小二乘法原理。

## 最佳的最小二乘解

小裘的

关于最佳的最小二乘解的一些理解

#### 最小二乘法较全面的证明过程

普通最小二乘法 (ordinary least squares, OLS) 是线性回归预测问题中一个很重要的概念,在 Introductory Econometrics A Modern Approach (Fourth Edition) 第2章 简单

#### 最小二乘法推导.pdf

拟合的常用算法之最小二乘法的推导,其中包含了一些重要的矩阵求导公式证明,通过对最小二乘法的数学矩阵推导,我们对其算法的原理应用有更深的理解

## 一文让你彻底了解大数据实时计算引擎Flink

随着这些年大数据的飞速发展,也出现了不少计算的框架(Hadoop、Storm、Spark、Flink)。在网上有人将大数据计算引擎的发展分为四个阶段。第一代:Hadoop承载

#### 一文让你彻底了解卷积神经网络

本文来自csdn,文章先来卷积神经网络和全连接神经网络对比,接着让我们直观理解卷积,卷积计算流程,结合案例进行相关的介绍。卷积神经网络(ConvolutionalNeur

#### 一文让你彻底理解JavaHashMap和ConcurrentHashMap

Map这样的KeyValue在软件开发中是非常经典的结构,常用于在内存中存放数据。本篇主要想讨论ConcurrentHashMap这样一个并发容器,在正式开始之前我觉得有必要

#### 一文让你彻底了解大数据实时计算引擎 Flink

随着这些年大数据的飞速发展,也出现了不少计算的框架(Hadoop、Storm、Spark、Flink)。在网上有人将大数据计算引擎的发展分为四个阶段。

#### 一文让你读懂工业4.0-智能制造与工业互联网.pdf

一文让你读懂工业4.0-智能制造与工业互联网.一文让你读 【工程师学算法】工程常用算法(一)——最小二

