

我是8位的

I am 8 bits, what about you?

随笔 - 205, 文章 - 0, 评论 - 103, 阅读 - 101万

导航

博客园

首页

新随笔

联系

订阅 

管理

<

2022年3月

>

日	一	二	三	四	五	六
27	28	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31	1	2
3	4	5	6	7	8	9

公告

你的支持是我的动力

欢迎关注微信公众号“我是8位的”



昵称：我是8位的

园龄：4年7个月

粉丝：288

关注：5

+加关注

盖楼抽奖

#她的梦想在发光#

HWD科技女性故事有奖征集

分享最打动的科技女性故事

活动时间：2022年3月8日-3月18日

马上参与



搜索

找找看

谷歌搜索

常用链接

我的随笔

我的评论

我的参与

最新评论

我的标签

积分与排名

积分 - 457097

排名 - 1198

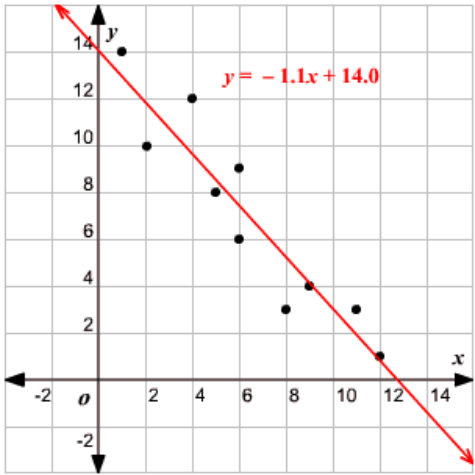
多变量微积分笔记2——最小二乘法

回归分析

回归分析是一种预测性的建模技术，用于连续性数据的机器学习（在监督学习中，如果预测的变量是离散的，称其为分类；如果预测的变量是连续的，称其为回归），它研究的是因变量（目标）和自变量（预测器）之间的关系。这种技术通常用于预测分析，时间序列模型以及发现变量之间的因果关系。

一元线性回归

在机器学习中，一元线性回归是回归分析中最简单的一种，它有一个自变量和一个因变量，它是这样描述的：如果能够一系列训练数据 $(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$ 寻找出一条最佳拟合直线 $y = ax + b$ ，用这条直线近似地表示 x 和 y 的关系，那么这种分析就称为一元线性回归分析。



多元线性回归

多元线性回归和一元线性回归本质上是一样的，只不过数据样本有 n 个维度（或称为 n 个特征向量），即：

$$x^{(i)} = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

这里 x 的上标表示第 i 个样本集。线性回归的目的是找到 x 的系数 θ ：

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix}$$

通过求解 θ 得到最终的拟合曲线，也就是预测模型：

随笔分类 (211)

- ★★资源下载★★(1)
- Java并发编程(1)
- 程序员的数学(24)
- 单变量微积分(31)
- 多变量微积分(24)
- 概率(24)
- 机器学习(27)
- 软件设计(1)
- 数据分析(6)
- 数据结构与算法(27)
- 随笔(5)
- 线性代数(34)
- 项目管理(2)
- 转载(4)

随笔档案 (205)

- 2021年2月(1)
- 2020年3月(2)
- 2020年2月(6)
- 2020年1月(4)
- 2019年12月(7)
- 2019年11月(15)
- 2019年9月(3)
- 2019年8月(6)
- 2019年7月(1)
- 2019年6月(8)
- 2019年5月(3)
- 2019年4月(5)
- 2019年3月(7)
- 2019年2月(3)
- 2019年1月(7)
- 更多

阅读排行榜

- 1. 使用Apriori进行关联分析（一）(29768)
- 2. 线性代数笔记12——列空间和零空间(28772)
- 3. FP-growth算法发现频繁项集（一）——构建FP树(24430)
- 4. 寻找“最好”（2）——欧拉-拉格朗日方程(23099)
- 5. 多变量微积分笔记3——二元函数的极值(22772)

评论排行榜

- 1. 隐马尔可夫模型（一）(8)
- 2. 线性代数笔记12——列空间和零空间(7)
- 3. 线性代数笔记3——向量2（点积）(7)
- 4. FP-growth算法发现频繁项集（一）——构建FP树(5)
- 5. 寻找“最好”（2）——欧拉-拉格朗日方程(4)

推荐排行榜

- 1. 寻找“最好”（2）——欧拉-拉格朗日方程(7)
- 2. FP-growth算法发现频繁项集（一）——构建FP树(7)
- 3. 线性代数笔记3——向量2（点积）(6)
- 4. FP-growth算法发现频繁项集（二）——发现频繁项集(5)
- 5. 隐马尔可夫模型（一）(5)

最新评论

- 1. Re:线性代数笔记3——向量2（点积）
如果点积小于0，即夹角小于90°，这个写错了吧。应该是夹角大于90°
--猫猫猫猫大人

$$h_{\theta}(x) = \theta_0x_0 + \theta_1x_1 + \theta_2x_2 + \cdots + \theta_nx_n$$
$$= [\theta_0, \theta_1, \theta_2, \cdots, \theta_n] \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$
$$= \theta^T \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

简写为：

$$h_{\theta}(X) = \theta^T X$$

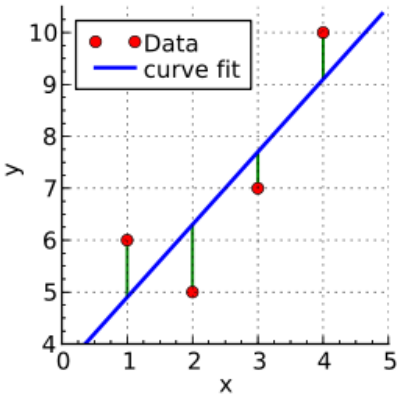
这正是y = ax的形式。如果令x₀ = 1，则多元线性回归就可看作一条直线y = ax + b，实际上，多元线性回归是n维空间的一个超平面。

误差的计算

一元线性回归的目标是找到最佳的a和b，这里“最佳”是说我们找出的这组a和b能使得所有训练数据误差的总和达到最小化。

计算误差的办法有很多种，其中两种典型的做法，一是计算数据到直线的距离，二是数据和直线所对应的y值。根据计算方法的不同，得到的最终训练模型也不同。

实际应用中，如果计算点到直线的距离，就需使用两个维度的数据进行计算，而实际上两个维度大多数时候都不存在直接的计算关系，比如时间和房价，二者并不能直接进行加减运算，想要运算必须通过成本函数转换。基于上述原因，通常使用第二种方法，也就是计算数据和直线所对应的y值：



最小二乘

找到最佳拟合直线的方法有很多，其中一种就是最小二乘法。

最小二乘法（又称最小平方方法）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，并使得这些求得的数据与实际数据之间误差的平方和为最小。最小二乘法还可用于曲线拟合。其他一些优化问题也可通过最小化能量或最大化熵用最小二乘法来表达。最小二乘法的计算公式：

2. Re:线性代数笔记10——矩阵的LU分解写的很好，不过LU分解的前提是错的，LU分解只需要第三个条件，如果允许行置换就是下面写到的PLU，可以分解所有矩阵

--wiki3D

3. Re:单变量微积分笔记20——三角替换1 (sin和cos) 很nice

--尹保棕

4. Re:线性代数笔记24——微分方程和exp(At) 有些图片挂了呢

--ccchendada

5. Re:寻找“最好” (2) ——欧拉-拉格朗日方程

提个issue，最速降线中

$v = \{2gh\}^{1/2}$ 与配图不一致，建议以起点为原点，向右伸出x轴，向下伸出y轴建立坐标系

--trustInU

$$E = \sum_{i=1}^n (y_i - h(x_i))^2$$

其中h是预期模型， $h(x_i)$ 是模型预测的结果， y_i 是实际结果。对于一元线性回归来说：

$$D(a, b) = \sum_{i=1}^n (y_i - (ax_i + b))^2$$

$D(a, b)$ 是关于a和b的函数，把 x_i 和 y_i 代入到最小二乘法的式子中，就得到了一个关于a和b的方程。在这个方程中，a和b是未知数，x和y是已知数。使 $D(a, b)$ 最小的方法就是找到临界点，也就是D关于a的偏导和D关于b的偏导同时等于0的点：

$$\begin{aligned} \begin{cases} \frac{\partial D}{\partial a} = \sum_{i=1}^n 2(y_i - (ax_i + b))(-x_i) = 0 \\ \frac{\partial D}{\partial b} = \sum_{i=1}^n 2(y_i - (ax_i + b))(-1) = 0 \end{cases} \\ \Rightarrow \begin{cases} \sum_{i=1}^n (y_i - (ax_i + b))(-x_i) = 0 \\ \sum_{i=1}^n (y_i - (ax_i + b))(-1) = 0 \end{cases} \\ \Rightarrow \begin{cases} \sum_{i=1}^n (x_i^2 a + x_i b - x_i y_i) = 0 \\ \sum_{i=1}^n (x_i a + b - y_i) = 0 \end{cases} \\ \Rightarrow \begin{cases} \left(\sum_{i=1}^n x_i^2 \right) a + \left(\sum_{i=1}^n x_i \right) b = \sum_{i=1}^n x_i y_i \\ \left(\sum_{i=1}^n x_i \right) a + nb = \sum_{i=1}^n y_i \end{cases} \end{aligned}$$

这就是最终的方程组，可以根据该方程组求出a和b。可以看到，如果数据集的不同，求出的a和b也不同，也就是最终的预测模型不同。

示例

用最小二乘法把下列数据拟合成直线和曲线，(0, 1), (2, 1), (3, 4)

$$D(a, b) = \sum_{i=1}^3 (y_i - (x_i a + b))^2$$

$$\begin{cases} \frac{\partial D}{\partial a} = \sum_{i=1}^3 2(y_i - (x_i a + b))(-x_i) = 0 \\ \frac{\partial D}{\partial b} = \sum_{i=1}^3 2(y_i - (x_i a + b))(-1) = 0 \end{cases}$$

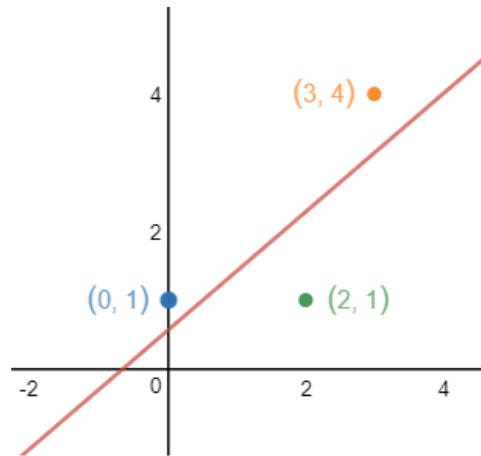
$$\Rightarrow \begin{cases} \left(\sum_{i=1}^3 x_i^2\right)a + \left(\sum_{i=1}^3 x_i\right)b = \sum_{i=1}^3 x_i y_i \\ \left(\sum_{i=1}^3 x_i\right)a + 3b = \sum_{i=1}^3 y_i \end{cases}$$

$$\Rightarrow \begin{cases} (0^2 + 2^2 + 3^2)a + (0 + 2 + 3)b = 0 \times 1 + 2 \times 1 + 3 \times 4 \\ (0 + 2 + 3)a + 3b = 1 + 1 + 4 \end{cases}$$

$$\Rightarrow \begin{cases} 13a + 5b = 14 \\ 5a + 3b = 6 \end{cases}$$

$$\Rightarrow \begin{cases} a = \frac{6}{7} \\ b = \frac{4}{7} \end{cases}$$

最佳拟合直线是 $y = 6x/7 + 4/7$



最佳拟合直线

如果拟合为曲线, 则 $y = ax^2 + bx + c$,

$$D(a, b, c) = \sum_{i=1}^n (y_i - (x_i^2 a + x_i b + c))^2$$

$$\begin{cases} \frac{\partial D}{\partial a} = \sum_{i=1}^n 2(y_i - (x_i^2 a + x_i b + c))(-x_i^2) = 0 \\ \frac{\partial D}{\partial b} = \sum_{i=1}^n 2(y_i - (x_i^2 a + x_i b + c))(-x_i) = 0 \\ \frac{\partial D}{\partial c} = \sum_{i=1}^n 2(y_i - (x_i^2 a + x_i b + c))(-1) = 0 \end{cases}$$

$$\begin{aligned}
& \Rightarrow \begin{cases} \sum_{i=1}^n (y_i - (x_i^2 a + x_i b + c))(-x_i^2) = 0 \\ \sum_{i=1}^n (y_i - (x_i^2 a + x_i b + c))(-x_i) = 0 \\ \sum_{i=1}^n (y_i - (x_i^2 a + x_i b + c))(-1) = 0 \end{cases} \\
& \Rightarrow \begin{cases} a \sum_{i=1}^n x_i^4 + b \sum_{i=1}^n x_i^3 = \sum_{i=1}^n x_i^2 y_i - c \sum_{i=1}^n x_i^2 \\ a \sum_{i=1}^n x_i^3 + b \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i - c \sum_{i=1}^n x_i \\ a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i = \sum_{i=1}^n y_i - nc \end{cases}
\end{aligned}$$

将(0, 1), (2, 1), (3, 4)代入方程:

$$\begin{aligned}
& \begin{cases} (0 + 2^4 + 3^4)a + (0 + 2^3 + 3^3)b = (0 + 2^2 \times 1 + 3^2 \times 4) - c(0 + 2^2 + 3^2) \\ (0 + 2^3 + 3^3)a + (0 + 2^2 + 3^2)b = (0 + 2 \times 1 + 3 \times 4) - c(0 + 2 + 3) \\ (0 + 2^2 + 3^2)a + (0 + 2 + 3)b = (1 + 1 + 4) - 3c \end{cases} \\
& \Rightarrow \begin{cases} 97a + 35b = 40 - 13c \\ 35a + 13b = 14 - 5c \\ 13a + 5b = 6 - 3c \end{cases} \\
& \Rightarrow \begin{cases} 97a + 35b + 13c = 40 \\ 35a + 13b + 5c = 14 \\ 13a + 5b + 3c = 6 \end{cases}
\end{aligned}$$

为了把这个无聊的方程交给计算机处理, 将其变为矩阵方程:

$$\begin{bmatrix} 97 & 35 & 13 \\ 35 & 13 & 5 \\ 13 & 5 & 3 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 40 \\ 14 \\ 6 \end{bmatrix} \Rightarrow \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 97 & 35 & 13 \\ 35 & 13 & 5 \\ 13 & 5 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 40 \\ 14 \\ 6 \end{bmatrix}$$

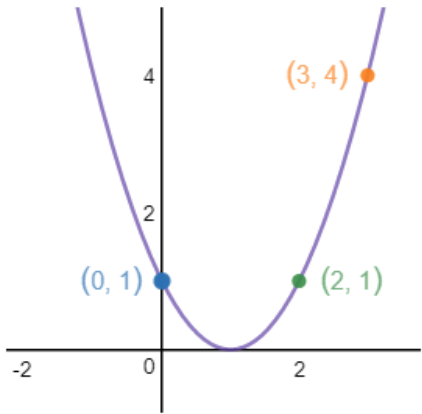
```

1 import numpy as np
2
3 a = np.matrix([[97,35,13],[35,13,5],[13,5,3]]) **
4 b = np.matrix([[40],[14],[6]])
5 print(a * b)

```

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

最终的拟合曲线是 $y = x^2 - 2x + 1$



最佳拟合曲线

可以看出，三个点全部在曲线上，最小误差是0。这实际上正是期望的结果，因为训练数据只有三个点，所以根据这三个点得到的三元一次方程组：

$$\begin{cases} 0^2a + 0b + c = 1 \\ 2^2a + b + c = 1 \\ 3^2a + 3b + c = 4 \end{cases} \Rightarrow \begin{cases} a = 1 \\ b = -2 \\ c = 1 \end{cases}$$

这也从侧面反映出最小二乘法的正确性。

出处：微信公众号“我是8位的”

本文以学习、研究和分享为主，如需转载，请联系本人，标明作者和出处，非商业用途！

扫描二维码关注作者公众号“我是8位的”



随笔

分类: [多变量微积分](#)

标签: [最小二乘法](#), [线性回归](#)

[好文要顶](#)[关注我](#)[收藏该文](#)



[我是8位的](#)
关注 - 5
粉丝 - 288
[+加关注](#)

0

 推荐

0

 反对

« 上一篇: [水乡苏州](#)
» 下一篇: [多变量微积分笔记3——二元函数的极值](#)

posted on 2018-01-17 20:37 [我是8位的](#) 阅读(3678) 评论(2) [编辑](#) [收藏](#) [举报](#)

 登录后才能查看或发表评论，立即 [登录](#) 或者 [逛逛](#) 博客园首页

【推荐】华为 HWD 2022 故事征集，分享最打动你的科技女性故事

【推荐】华为开发者专区，与开发者一起构建万物互联的智能世界



The advertisement banner for ASP.NET Core features a dark background with a grid of charts and graphs. The text 'WebForms升级到ASP.NET Core' is prominently displayed in white, with 'Core' in a larger font. Below it, 'fineui.com' is written in a smaller font. A large white arrow points to the right. The banner also includes a small '广告' (Advertisement) label in the top right corner.

编辑推荐:

- 革命性创新，动画杀手铜 @scroll-timeline
- 戏说领域驱动设计（十二）—— 服务
- ASP.NET Core 6框架揭秘实例演示[16]: 内存缓存与分布式缓存的使用
- .Net Core 中无处不在的 Async/Await 是如何提升性能的?
- 分布式系统改造方案 —— 老旧系统改造篇



The banner for the HWD Women's Story Contest has a pink and purple background. It features the text '#她的梦想在发光#' and 'HWD科技女性故事有奖征集' in bold white and pink letters. Below this, it says '活动时间: 2022年3月6日~3月18日'. On the right side, there is a colorful illustration of a woman's head and shoulders, composed of geometric shapes.

最新新闻:

- 乔布斯的创业搭档：他缺乏工程师才能，不得不锻炼营销能力来弥补
 - 美国大厂码农薪资曝光：年薪18万美元，够养家，不够买海景房
 - 两张照片就能转视频！Google提出FLIM帧插值模型
 - Android 再推“杀手级”功能，可回收 60% 存储空间
 - 溺在理财暴雷潮的投資人：本金63万，月兑25元不够买菜
- » 更多新闻...

Powered by:
[博客园](#)

Copyright © 2022 我是8位的
Powered by .NET 6 on Kubernetes