WIKIPEDIA

曲线拟合

维基百科，自由的百科全书

曲线拟合（fit theory），俗称拉曲线，是一种把现有数据透过数学方法来代入一条数式的表示方式。科学和工程问题可以通过诸如采样、实验等方法获得若干离散的数据，根据这些数据，我们往往希望得到一个连续的函数（也就是曲线）或者更加密集的离散方程与已知数据相吻合，这过程就叫做拟合 **(fitting)**。

本条目讲述如何透过拉曲线的方法来进行插值法运算及递归分析的基础。

目录

类型

将函数拟合至数据点 (Fitting functions to data points)

最常见的是，一函数符合 $y=f(x)$ 的形式。

(Most commonly, one fits a function of the form $y=f(x)$.)

拟合直线或多项式曲线

方程$y = ax + b$ 在笛卡尔平面上是一条直线，而这条直线的斜率是a。因为任何两点可以决定一条直线，因此总能找到次数不多于1的多项式来串起任何两个x值相异的点。

如果把多次式的次数增加到2

$$y = ax^2 + bx + c$$

那么只要给定x值各异的3点，总会有次数不多于2的多项式可以把它们串起。

如果把多次式的次数再增加到3

$$y = ax^3 + bx^2 + cx + d$$

那么只要给定x值各异的4点，总会有次数不多于3的多项式可以把它们串起。

对于这条多项式，更正确的描述是这条多项式附合任何4个限制。限制可以是一点（x,y）、角度或曲率（即半径的倒数 $1/R$）。角度和曲率的限制通常在曲线的终端，因此称为终端条件 。为了样条(spline) 的交接平滑，通常会用到全等的终端条件。 也可以增加如曲率变化等高阶约束。例如，在高速公路立体交叉点苜蓿叶型的设计中，可以用来理解当汽车绕着交叉点运动时作用在汽车上的力，并依此设定合理的限定时速。

一次多项式也可以拟合一个单点和一个角度，三次多项式则可以拟合两点，一个角度约束以及一个曲率约束。许多其它类型的约束组合也同样可以用低阶或者高阶多项式来拟合。

如果有超过n+1个约束（n是多项式的阶次），仍然可以使用多项式拟合。通常一个满足所有约束的精确拟合不一定能够得到（但是有可能得到，例如，用一次多项式拟合共线的三点三点共线）。通常，我们需要使用一些方法来评价拟合的好坏。最小平方法就是用来评价差别的一种常用的方法。

不通过提高多项式的次数来更好的拟合曲线的原因有下：

- 即使存在精确的拟合，也不意味着必须得到这样的拟合。根据使用的算法不同，我们可能遇到分歧，要么精确的拟合无法得到，要么需要太多的计算机时去得到精确的拟合。不管哪种情况，最终都会以得到近似拟合而结束。
- 通常人们会希望得到一个近似的拟合，而不愿为了精确拟合数据而使拟合的曲线产生扭曲。
- 高次多项式往往有高度波动的特性。如果我们通过两点“A”和“B”作一条曲线，我们希望这条曲线也能通过“A”和“B”的中点。但是对于高次多项式，情况就不是这样了，高次多项式曲线往往可能有很大或者很小的幅值。对于低次多项式，曲线将没有很大波动，而能通过中点（对于一次多项式，甚至能保证肯定通过中点）。

将其他函数拟合至数据点 (Fitting other functions to data points)

其他类型的曲线在特定条件下也可使用，诸如三角函数 (如正弦或是余弦函数)。

(Other types of curves, such as trigonometric functions (such as sine and cosine), may also be used, in certain cases.)

在光谱学中，数据可以用高斯，罗伦兹，Voigt 和相关函数来拟合。

(In spectroscopy, data may be fitted with Gaussian, Lorentzian, Voigt and related functions.)

在农业中，反向逻辑 S函数 (S曲线）用于描述作物产量和生长因子之间的关系。 蓝色图形是将农田中测量数据进行S形回归得出的。 可以看出在初始情况，即土壤低盐度的状况下，作物产量在土壤盐度增加时缓慢降低，而之后降得更快。

(In agriculture the inverted logistic sigmoid function (S-curve) is used to describe the relation between crop yield and growth factors. The blue figure was made by a sigmoid regression of data measured in farm lands. It can be seen that initially, i.e. at low soil salinity, the crop yield reduces slowly at increasing soil salinity, while thereafter the decrease progresses faster.)

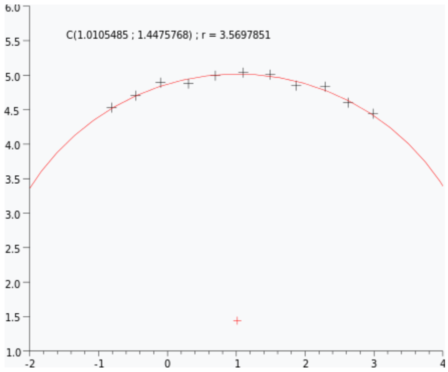
将平面曲线拟合至数据点 (Fitting plane curves to data points)

If a function of the form $y = f(x)$ cannot be postulated, one can still try to fit a plane curve.

Other types of curves, such as conic sections (circular, elliptical, parabolic, and hyperbolic arcs) or trigonometric functions (such as sine and cosine), may also be used, in certain cases. For example, trajectories of objects under the influence of gravity follow a parabolic path, when air resistance is ignored. Hence, matching trajectory data points to a parabolic curve would make sense. Tides follow sinusoidal patterns, hence tidal data points should be matched to a sine wave, or the sum of two sine waves of different periods, if the effects of the Moon and Sun are both considered.

For a parametric curve, it is effective to fit each of its coordinates as a separate function of arc length; assuming that data points can be ordered, the chord distance may be used.[1]

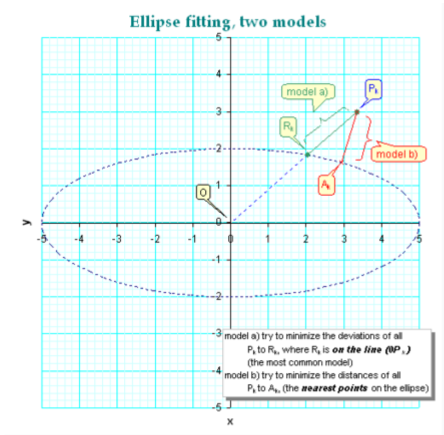
Fitting a circle by geometric fit



Circle fitting with the Coope method, the points describing a circle arc, centre (1 ; 1), radius 4.

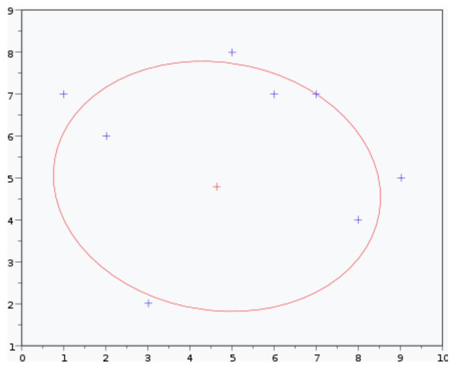
Coope[2] approaches the problem of trying to find the best visual fit of circle to a set of 2D data points. The method elegantly transforms the ordinarily non-linear problem into a linear problem that can be solved without using iterative numerical methods, and is hence an order of magnitude faster than previous techniques.

Fitting an ellipse by geometric fit



different models of ellipse fitting

The above technique is extended to general ellipses[3] by adding a non-linear step, resulting in a method that is fast, yet finds visually pleasing ellipses of arbitrary orientation and displacement.


Ellipse fitting minimising the algebraic distance (Fitzgibbon method).

参见

- 方差分析
- 安斯库姆四重奏
- 横截面回归
- 曲线拟合
- 经验贝叶斯方法
- 逻辑斯蒂回归
- M估计
- 非线性回归
- 非参数回归
- 多元自适应回归样条
- Lack-of-fit sum of squares
- 截断回归模型
- 删失回归模型
- 简单线性回归
- 分段线性回归

外部链接

- Curve Expert (shareware) (https://web.archive.org/web/20060507151227/http://www.ebicom.net/~dhyams/cftp.htm) fits functions to data (limited to one dependant and one independent variable.)
- Zunzun.com (http://zunzun.com) Online curve and surface fitting
- TableCurve2D and TableCurve3D by Systat (http://www.systat.com) automates curve fitting
- another choice (http://www.softintegration.com/chhtml/lang/lib/libch/numeric/CGI_Curvefit.html)
- online curve-fitting textbook (http://curvefit.com/)
- Interactive curve fitting using Least Squares with Weights on savetman.com (https://web.archive.org/web/20060325083538/http://www.savetman.com/curvefit/index.html)

参考资料

1. p.51 in Ahlberg & Nilson (1967) *The theory of splines and their applications*, Academic Press, 1967 [1] (https://books.google.com/books?id=S7d1pjJHsRgC&pg=PA51)
2. Coope, I.D. Circle fitting by linear and nonlinear least squares. Journal of Optimization Theory and Applications. 1993, **76** (2): 381–388. doi:10.1007/BF00939613.
3. Paul Sheer, A software assistant for manual stereo photometrology, M.Sc. thesis, 1997