Ajuste de Curvas pelo Método dos Quadrados Mínimos

Marcone Jamilson Freitas Souza, Departamento de Computação, Instituto de Ciências Exatas e Biológicas, Universidade Federal de Ouro Preto, 35400-000 Ouro Preto, MG, Brasil. Homepage: http://www.decom.ufop.br/prof/marcone, E-mail: marcone@iceb.ufop.br

1 Introdução

Em muitas situações, conhece-se uma tabela de pontos (x_i, y_i), onde cada y_i é obtido experimentalmente, e deseja-se obter a expressão analítica de uma dada curva $y = f(x)$ que melhor se ajusta a esse conjunto de pontos. Por exemplo, sabe-se que o número y de bactérias, por unidade de volume, existente em uma cultura após um determinado número x de horas, cresce exponencialmente com o aumento de x. Neste caso, o número de bactérias cresce com o decorrer das horas na forma $y = \alpha e^{\beta x}$. O problema consiste, então, em determinar os valores mais apropriados dos parâmetros α e β desta exponencial.

2 Ajuste a uma reta

Mostremos, inicialmente, como ajustar um conjunto de pontos a uma reta $y = a + bx$, onde a e b são parâmetros a serem determinados.

Neste caso, estamos interessados em minimizar a distância de cada ponto (x_i, y_i) da tabela à cada ponto $(x_i,\ a + bx_i)$ da reta, conforme ilustra a figura 1.

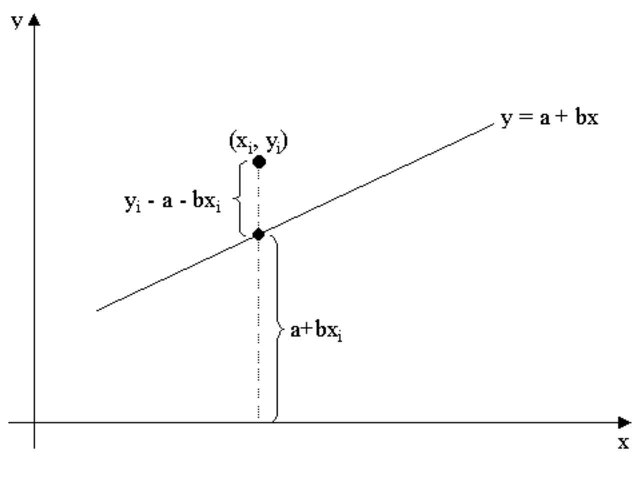

Figura 1: Distância de um ponto (x_i, y_i) à reta $y = a + bx$

A distância entre esses pontos é $|y_i - a - bx_i|$ e a soma dos quadrados dessas distâncias é:

$$q = \sum_{i=1}^{n} (y_i - a - bx_i)^2 \tag{2.1}$$

Os candidatos a ponto de mínimo da função 2.1 são aqueles para os quais são nulos as derivadas parciais de q em relação a cada um de seus parâmetros, isto é:

$$\frac{\partial q}{\partial a} = -2 \sum_{i=1}^{n} (y_i - a - bx_i) = 0 \tag{2.2}$$

$$\frac{\partial q}{\partial b} = -2 \sum_{i=1}^{n} x_i (y_i - a - bx_i) = 0 \tag{2.3}$$

Tendo em vista que:

$$\begin{array}{rcl} \sum\limits_{i=1}^{n} (y_i - a - bx_i) & = & \sum\limits_{i=1}^{n} y_i - \sum\limits_{i=1}^{n} a - \sum\limits_{i=1}^{n} bx_i \\ & = & \sum\limits_{i=1}^{n} y_i - na - \left(\sum\limits_{i=1}^{n} x_i \right) b \end{array}$$

e que:

$$\sum_{i=1}^{n} x_i (y_i - a - bx_i) = \sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i \right) a - \left(\sum_{i=1}^{n} x_i^2 \right) b$$

obtemos o seguinte sistema de equações, denominado "equações normais" do problema, cujas incógnitas são os parâmetros a e b da equação $y = a + bx$:

$$\left\{ \begin{array}{rcrcl} na & + & \left(\sum\limits_{i=1}^{n} x_i \right) b & = & \sum\limits_{i=1}^{n} y_i \\ \left(\sum\limits_{i=1}^{n} x_i \right) a & + & \left(\sum\limits_{i=1}^{n} x_i^2 \right) b & = & \sum\limits_{i=1}^{n} x_i y_i \end{array} \right. \tag{2.4}$$

<u>Exemplo 1</u>:

Dada a tabela de pontos (x_i, y_i) a seguir, determine pelo Método dos Quadrados Mínimos a equação da reta que melhor se ajusta a esses pontos.

x_i	-1.0	-0.1	0.2	1.0
y_i	1.000	1.099	0.808	1.000

<u>Solução</u>:

Como são $n = 4$ pontos, $\sum\limits_{i=1}^{n} x_i = 0.1$, $\sum\limits_{i=1}^{n} x_i^2 = 2.05$, $\sum\limits_{i=1}^{n} y_i = 3.907$ e $\sum\limits_{i=1}^{n} x_i y_i = 0.0517$, as equações normais do problema são, de acordo com 2.4:

$$\left\{ \begin{array}{rcrcl} 4a & + & 0.10b & = & 3.9070 \\ 0.1a & + & 2.05b & = & 0.0517 \end{array} \right.$$

A solução deste sistema é $a = 0.9773$ e $b = -0.0224$. Assim, a reta que melhor se ajusta à tabela de pontos dada é:

$$y = 0.9773 - 0.0224x$$

3 Ajuste a uma exponencial

Mostremos, agora, como ajustar um conjunto de pontos (x_i, y_i) a uma exponencial do tipo $y = \alpha e^{bx}$.

Esta função exponencial pode ser ajustada através da seguinte transformação:

$\ln y = \ln\left(\alpha e^{bx}\right) = \ln \alpha + bx$.

Fazendo $Y = \ln y$ e $a = \ln \alpha$, reduzimos o problema de ajustar a tabela de pontos (x_i, y_i) referente a uma exponencial ao problema de ajustar a tabela de pontos (x_i, Y_i), onde $Y_i = \ln y_i$, à equação de uma reta $Y = a + bx$.

Exemplo 2:

Suponhamos que em um laboratório obtivemos experimentalmente os seguintes valores para $f(x_i)$ sobre os pontos x_i:

x_i	-1.0	-0.7	-0.4	-0.1	0.2	0.5	0.8	1.0
y_i	36.547	17.264	8.155	3.852	1.820	0.860	0.406	0.246

Solução:

Fazendo o diagrama de dispersão dos dados acima, verifica-se que um ajuste do tipo $y = \alpha e^{bx}$ é o mais indicado. Efetuando-se as transformações $Y = \ln y_i$, obtemos a tabela $(x_i, \ln y_i)$ a seguir:

x_i	-1.0	-0.7	-0.4	-0.1	0.2	0.5	0.8	1.0
$\ln y_i$	3.599	2.849	2.099	1.349	0.599	-0.151	-0.901	-1.402

Como $n = 8$ pontos, $\sum_{i=1}^{n} x_i = 0.3$, $\sum_{i=1}^{n} x_i^2 = 3.59$, $\sum_{i=1}^{n} y_i = 0.041$ e $\sum_{i=1}^{n} x_i y_i = -8.646$, as equações normais do problema são, de acordo com 2.4:

$$\begin{cases} 8a & + & 0.30b & = & 0.041 \\ 0.30a & + & 3.59b & = & -8.646 \end{cases}$$

A solução deste sistema é $a = 1.099$ e $b = -2.5$. Como $a = \ln \alpha$ então $\alpha = e^a = e^{1.099} = 3.001$. Assim, a exponencial que melhor se ajusta à tabela de pontos dada é:

$$y = 3.001 e^{-2.5x}$$

4 Ajuste a uma hipérbole

Para ajustar uma tabela de pontos (x_i, y_i), onde:

$$y = \frac{1}{\alpha_1 + \alpha_2 x} \tag{4.5}$$

basta fazer $z = \frac{1}{y} = \alpha_1 + \alpha_2 x$.

5 Ajuste a uma curva exponencial $y = \alpha_1 \alpha_2^x$

Para ajustar uma tabela de pontos (x_i, y_i), onde:

$$y = \alpha_1 \alpha_2^x \tag{5.6}$$

basta fazer as seguintes transformações, considerando $y > 0$:

$z = \ln y = \underbrace{\ln\alpha_1}_{a} + x\underbrace{\ln\alpha_2}_{b} = a + bx$

6 Ajuste a uma curva geométrica $y = \alpha_1 x^{\alpha_2}$

Para ajustar uma tabela de pontos (x_i, y_i), onde:

$$y = \alpha_1 x^{\alpha_2} \tag{6.7}$$

basta fazer as seguintes transformações, considerando $y > 0$ e $x > 0$:

$z = \ln y = \underbrace{\ln\alpha_1}_{a} + \underbrace{\alpha_2}_{b}\underbrace{\ln x}_{t} = a + bt$

Neste caso, estamos minimizando as somas dos quadrados dos desvios nos logaritmos de y, para os logaritmos dos desvios de x.

7 Ajuste a um polinômio

O objetivo, agora, é mostrar como ajustar os pontos de uma tabela com n pontos a uma função polinomial de grau m:

$$P(x) = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m \tag{7.8}$$

onde $m \leq n - 1$. Neste caso, a soma dos quadrados das distâncias de y_i à $P(x_i)$ é dada por:

$$q = \sum (y_i - P(x_i))^2 \tag{7.9}$$

e depende de $m + 1$ parâmetros $a_0, a_1, \cdots, a_m$. Para minimizar essa função, temos que satisfazer às $m + 1$ condições a seguir:

$$\frac{\partial q}{\partial a_i} = 0 \quad \forall i = 0, 1, \cdots, m \tag{7.10}$$

a qual fornece um sistema de $m + 1$ equações normais.

No caso de a função polinomial ser quadrática, isto é:

$$P(x) = a_0 + a_1x + a_2x^2 \tag{7.11}$$

as equações normais são:

$$\left\{ \begin{array}{rcrcrcl} na_0 & + & \left(\sum\limits_{i=1}^{n} x_i\right) a_1 & + & \left(\sum\limits_{i=1}^{n} x_i^2\right) a_2 & = & \sum\limits_{i=1}^{n} y_i \\ \left(\sum\limits_{i=1}^{n} x_i\right) a_0 & + & \left(\sum\limits_{i=1}^{n} x_i^2\right) a_1 & + & \left(\sum\limits_{i=1}^{n} x_i^3\right) a_2 & = & \sum\limits_{i=1}^{n} x_iy_i \\ \left(\sum\limits_{i=1}^{n} x_i^2\right) a_0 & + & \left(\sum\limits_{i=1}^{n} x_i^3\right) a_1 & + & \left(\sum\limits_{i=1}^{n} x_i^4\right) a_2 & = & \sum\limits_{i=1}^{n} x_i^2y_i \end{array} \right. \tag{7.12}$$

Observe que este sistema é simétrico. Para resolvê-lo, isto é, para encontrar as incógnitas $a_0, a_1, \cdots, a_m$, podemos aplicar qualquer um dos métodos numéricos apresentados anteriormente.

8 Qualidade do ajuste

A qualidade de um ajuste linear pode ser verificada em função do coeficiente de determinação r^2, dado por:

$$r^2 = \frac{\sum_{i=1}^{n} (a + bx_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{8.13}$$

sendo $\bar{y} = \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)$. Quanto mais próximo da unidade r^2 estiver, melhor é o ajuste.

Observe que o coeficiente de determinação é uma medida da proporção da variação total dos dados em torno da média. De fato, o numerador desta expressão representa a soma dos quadrados dos desvios de cada ponto da reta de ajuste ao ponto médio $\bar{y}$ dos pontos dados. Já o denominador representa a soma dos quadrados dos desvios de cada ponto dado ao ponto médio $\bar{y}$.

Tendo em vista que:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - a - bx_i)^2 + \sum_{i=1}^{n}(a + bx_i - \bar{y})^2$$

a expressão 8.13 pode ser reescrita como:

$$r^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2 - \sum_{i=1}^{n} (y_i - a - bx_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

Como:

$$\begin{aligned} \sum_{i=1}^{n} (y_i - \bar{y})^2 &= \sum_{i=1}^{n} y_i^2 - 2\bar{y}\sum_{i=1}^{n} y_i + n\sum_{i=1}^{n} \bar{y}^2 \\ &= \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2 \end{aligned}$$

a expressão para determinação do coeficiente de determinação r^2 pode ser simplificada para:

$$r^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - a - bx_i)^2}{\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2} \tag{8.14}$$