Apontamentos das aulas

Método dos Mínimos Quadrados

Pedro R. S. Antunes (2005)

Em muitas aplicações estamos interessados em obter aproximações de uma dada função f por uma outra função g em geral mais simples. Uma possibilidade é considerar $g \equiv p_n$, o (único) polinómio de grau menor ou igual a n que interpola f em $n+1$ pontos distintos de um intervalo $[a, b]$. No entanto, em muitas situações não é conveniente que a função aproximadora seja uma função interpoladora. Suponhamos que a função f traduz uma relação física entre duas grandezas x e $f(x)$. Experimentalmente podemos efectuar medições de $f(x)$ em pontos $x_i,\ i = 0, 1, ..., n$ de forma a obter um conjunto de dados (x_0, f_0), (x_1, f_1),...,(x_n, f_n). No entanto, nesse caso apenas dispomos dos valores $f_i = f(x_i) + \epsilon_i$, $i = 0, 1, ..., n$, ou seja os valores f_i não correspondem exactamente aos valores $f(x_i)$, pois estão "contagiados" por um erro ϵ_i associado à imprecisão da medição. Ao tentar encontrar uma função g que constitua uma boa aproximação de f seria inadequado exigir que a função aproximadora interpolasse esses pontos.

Nesta secção vamos considerar outros tipos de funções aproximadoras para f, sem a exigência de que elas interpolem f. Uma solução mais adequada será definir uma base de funções

$$B = \{\phi_0(x), \phi_1(x), ..., \phi_m(x)\}$$

e tomar para função aproximadora uma combinação linear das funções da base B, ou seja, definir

$$g(x) = \alpha_0\phi_0(x) + \alpha_1\phi_1(x) + ... + \alpha_m\phi_m(x). \tag{1}$$

Pretendemos que a função g se aproxime "o mais possível" de f. Assim, vamos determinar os coeficientes $\alpha_0, \alpha_1, ..., \alpha_m$ de forma a que os desvios

$$d_i = f_i - g(x_i),\ i = 0, 1, ..., n$$

sejam tão pequenos quanto possível, segundo algum critério[1]. Para isso precisamos de uma noção de "distância", pelo que iremos considerar normas vectoriais. Designando por $\vec{d}$, $\vec{f}$ e $\vec{g}$ os vectores de $\mathbb{R}^{n+1}$ com componentes d_i, f_i e g_i e escolhendo uma norma vectorial $\|.\|$, os coeficientes $\alpha_0, \alpha_1, ..., \alpha_m$ deverão ser determinados de

[1]Se obtivermos $d_i = 0,\ i = 0, 1, ..., n$ obtemos um caso de interpolação.

modo a que $\left\| \vec{d} \right\| = \left\| \vec{f} - \vec{g} \right\|$ seja mínima.

Como exemplos de normas de $\mathbb{R}^{n+1}$ temos

$$\left\| \vec{d} \right\|_1 = \sum_{i=0}^{n} |d_i|, \qquad \left\| \vec{d} \right\|_\infty = max_{0\leq i\leq n} |d_i| \qquad ou \qquad \left\| \vec{d} \right\|_2 = \left[\sum_{i=0}^{n} \left(d_i^2\right) \right]^{\frac{1}{2}}$$

A norma que vulgarmente é escolhida neste contexto é a norma $\|.\|_2$, pelo que os parâmetros $\alpha_0, \alpha_1, ..., \alpha_m$ são determinados de modo a minimizar

$$\left\| \vec{d} \right\|_2 = \left[\sum_{i=0}^{n} (d_i)^2 \right]^{\frac{1}{2}} = \left[\sum_{i=0}^{n} (f_i - g_i)^2 \right]^{\frac{1}{2}}.$$

Isto equivale a minimizar a quantidade

$$D = \left\| \vec{d} \right\|_2^2 = \sum_{i=0}^{n} (f_i - g_i)^2 \tag{2}$$

o que justifica a designação de *método dos mínimos quadrados*.

Um caso em que se aplicam estes conceitos é o da teoria da regressão linear. Em muitas aplicações de estatística, dada uma lista de pontos tentamos determinar a recta que melhor se ajusta aos pontos, ou seja, a recta que minimiza a soma dos quadrados dos desvios d_i. Isto corresponde a considerar $g(x) = \alpha_0\phi_0(x) + \alpha_1\phi_1(x)$ com $\phi_0(x) = 1$ e $\phi_1(x) = x$ (Figura 1).

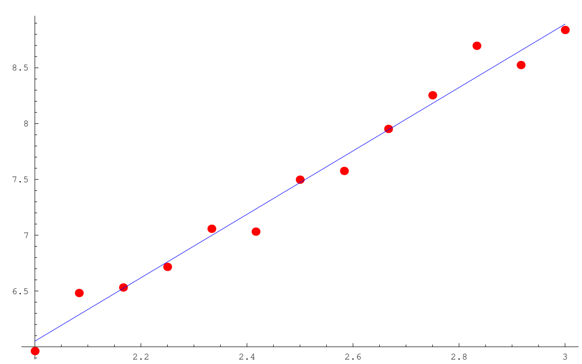

Figura 1: *A recta que, no sentido dos mínimos quadrados, melhor se ajusta à lista de pontos.*

Exemplo

Consideramos a tabela de valores de uma função f

x_i	0	1	2	3
$f(x_i)$	0	0.2	1	5

Vamos determinar a recta $g(x) = \alpha_0 + \alpha_1 x$ que melhor se ajusta aos pontos da tabela segundo o critério dos mínimos quadrados. Consideramos

$$D = \left\| \vec{d} \right\|_2^2 = \sum_{i=0}^{3} (f_i - g_i)^2 \tag{3}$$

como função de α_0 e α_1. Pretende-se os valores α_0 e α_1 que tornam D mínima. Assim, vamos impor que

$$\frac{\partial D}{\partial \alpha_0} = 0 \quad e \quad \frac{\partial D}{\partial \alpha_1} = 0$$

Calculando directamente as derivadas parciais a partir da expressão (3) obtemos

$$\frac{\partial D}{\partial \alpha_0} = \sum_{i=0}^{3} \frac{\partial}{\partial \alpha_0} \left[(f_i - \alpha_0 - \alpha_1 x_i)^2 \right] = \sum_{i=0}^{3} 2\,(f_i - \alpha_0 - \alpha_1 x_i) \times (-1) = 0$$

Assim

$$\frac{\partial D}{\partial \alpha_0} = 0 \Longleftrightarrow \sum_{i=0}^{3} (\alpha_0 + \alpha_1 x_i) = \sum_{i=0}^{3} f_i \tag{4}$$

Da mesma forma

$$\frac{\partial D}{\partial \alpha_1} = \sum_{i=0}^{3} \frac{\partial}{\partial \alpha_1} \left[(f_i - \alpha_0 - \alpha_1 x_i)^2 \right] = \sum_{i=0}^{3} 2\,(f_i - \alpha_0 - \alpha_1 x_i) \times (-x_i) = 0$$

pelo que

$$\frac{\partial D}{\partial \alpha_1} = 0 \Longleftrightarrow \sum_{i=0}^{3} \left(\alpha_0 x_i + \alpha_1 x_i^2 \right) = \sum_{i=0}^{3} x_i f_i \tag{5}$$

As relações (4) e (5) conduzem a um sistema linear de duas equações nas duas incógnitas α_0 e α_1. Matricialmente temos

$$\begin{pmatrix} \sum_{i=0}^{3} 1 & \sum_{i=0}^{3} x_i \\ \sum_{i=0}^{3} x_i & \sum_{i=0}^{3} x_i^2 \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} = \begin{pmatrix} \sum_{i=0}^{3} f_i \\ \sum_{i=0}^{3} x_i f_i \end{pmatrix}$$

e substituindo os valores obtemos

$$\begin{pmatrix} 4 & 6 \\ 6 & 14 \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} = \begin{pmatrix} 6.2 \\ 17.2 \end{pmatrix} \Rightarrow \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} = \begin{pmatrix} -0.82 \\ 1.58 \end{pmatrix}$$

pelo que a recta que melhor se ajusta aos pontos da tabela, segundo o critério dos mínimos quadrados é $g(x) = -0.82 + 1.58x$.

Na Figura 2 apresentamos dois possíveis ajustamentos de uma recta aos pontos da tabela. No primeiro caso consideramos uma recta arbitrária; no segundo caso temos o melhor ajustamento, no sentido dos mínimos quadrados. No primeiro caso obtivémos $D \approx 6157.1$, no segundo $D \approx 3.95$ um valor inferior como seria de esperar.

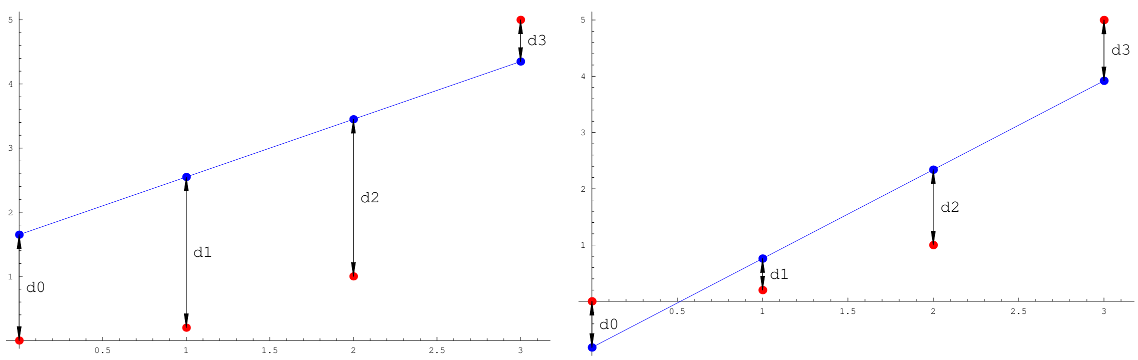

Figura 2: *dois possíveis ajustamentos por uma recta*

Caso Geral - As equações normais

Consideramos agora o caso geral de obter uma função da forma (1) onde os coeficientes $\alpha_0, \alpha_1, ..., \alpha_m$ são obtidos de forma a minimizar a função

$$D = \left\| \vec{d} \right\|_2^2 = \sum_{i=0}^{n} (f_i - g_i)^2 \tag{6}$$

pelo que é natural impor as condições

$$\frac{\partial D}{\partial \alpha_j} = 0, \ j = 0, 1, ..., n.$$

Assim obtemos

$$\frac{\partial D}{\partial \alpha_0} = \sum_{i=0}^{n} 2\left(f_i - \alpha_0\phi_0(x_i) - ... - \alpha_m\phi_m(x_i)\right) \times \left(-\phi_0(x_i)\right) = 0$$

$$\frac{\partial D}{\partial \alpha_1} = \sum_{i=0}^{n} 2\left(f_i - \alpha_0\phi_0(x_i) - ... - \alpha_m\phi_m(x_i)\right) \times \left(-\phi_1(x_i)\right) = 0$$

$$\begin{array}{cc} \vdots & \vdots \end{array}$$

$$\frac{\partial D}{\partial \alpha_m} = \sum_{i=0}^{n} 2\left(f_i - \alpha_0\phi_0(x_i) - ... - \alpha_m\phi_m(x_i)\right) \times \left(-\phi_m(x_i)\right) = 0$$

e passando para o membro direito os termos que dependem dos valores f_i obtemos

$$\alpha_0 \sum_{i=0}^{n} \phi_0(x_i)\phi_0(x_i) + \alpha_1 \sum_{i=0}^{n} \phi_0(x_i)\phi_1(x_i) + ... + \alpha_m \sum_{i=0}^{n} \phi_0(x_i)\phi_m(x_i) = \sum_{i=0}^{n} f_i\phi_0(x_i)$$

$$\alpha_0 \sum_{i=0}^{n} \phi_1(x_i)\phi_0(x_i) + \alpha_1 \sum_{i=0}^{n} \phi_1(x_i)\phi_1(x_i) + ... + \alpha_m \sum_{i=0}^{n} \phi_1(x_i)\phi_m(x_i) = \sum_{i=0}^{n} f_i\phi_1(x_i)$$

$$\begin{array}{cc} \vdots & \vdots \end{array}$$

$$\alpha_0 \sum_{i=0}^{n} \phi_m(x_i)\phi_0(x_i) + \alpha_1 \sum_{i=0}^{n} \phi_m(x_i)\phi_1(x_i) + ... + \alpha_m \sum_{i=0}^{n} \phi_m(x_i)\phi_m(x_i) = \sum_{i=0}^{n} f_i\phi_m(x_i)$$

Chegamos assim a um sistema linear de $m+1$ equações nas $m+1$ incógnitas $\alpha_0, \alpha_1, ..., \alpha_m$ a que se chama sistema de equações normais ou sistema normal.

Vamos definir os vectores de $\mathbb{R}^{n+1}$

$\vec{\phi_0} = [\phi_0(x_0), ..., \phi_0(x_n)]^T$; $\vec{\phi_1} = [\phi_1(x_0), ..., \phi_1(x_n)]^T$; ... ; $\vec{\phi_m} = [\phi_m(x_0), ..., \phi_m(x_n)]^T$.

Podemos agora reparar que, por exemplo o termo $\sum_{i=0}^{n} \phi_0(x_i)\phi_1(x_i)$ não é mais do que o produto interno[2] (usual) de $\mathbb{R}^{n+1}$ dos vectores $\vec{\phi_0}$ e $\vec{\phi_1}$. Assim o sistema de

[2]dados dois vectores $\vec{u},\ \vec{v} \in \mathbb{R}^{n+1}$ definimos o produto interno $\langle \vec{u}, \vec{v} \rangle = \sum_{i=0}^{n} u_i v_i$

equações normais toma a forma

$$\begin{pmatrix} \left\langle \vec{\phi_0}, \vec{\phi_0} \right\rangle & \left\langle \vec{\phi_0}, \vec{\phi_1} \right\rangle & ... & \left\langle \vec{\phi_0}, \vec{\phi_m} \right\rangle \\ \left\langle \vec{\phi_1}, \vec{\phi_0} \right\rangle & \left\langle \vec{\phi_1}, \vec{\phi_1} \right\rangle & ... & \left\langle \vec{\phi_1}, \vec{\phi_m} \right\rangle \\ .... & ... & ... & ... \\ \left\langle \vec{\phi_m}, \vec{\phi_0} \right\rangle & \left\langle \vec{\phi_m}, \vec{\phi_1} \right\rangle & ... & \left\langle \vec{\phi_m}, \vec{\phi_m} \right\rangle \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ ... \\ ... \\ \alpha_m \end{pmatrix} = \begin{pmatrix} \left\langle \vec{f}, \vec{\phi_0} \right\rangle \\ \left\langle \vec{f}, \vec{\phi_1} \right\rangle \\ ... \\ \left\langle \vec{f}, \vec{\phi_m} \right\rangle \end{pmatrix} \tag{7}$$

Podemos facilmente ver que a matriz do sistema é simétrica, pois por comutatividade do produto em $\mathbb{R}$

$$\left\langle \vec{\phi_k}, \vec{\phi_j} \right\rangle = \sum_{i=0}^{n} \phi_k(x_i)\phi_j(x_i) = \sum_{i=0}^{n} \phi_j(x_i)\phi_k(x_i) = \left\langle \vec{\phi_j}, \vec{\phi_k} \right\rangle .$$

Prova-se que se a funções $\phi_0(x), \phi_1(x), ..., \phi_m(x)$ forem linearmente independentes, então o sistema tem uma solução única, precisamente a função g que minimiza D.

Exemplo

Se considerarmos a aproximação através de funções polinomiais temos como funções base $\phi_0(x) = 1;\ \phi_1(x) = x;\ ...\ ;\ \phi_m(x) = x^m$ e o sistema de equações normais (7) toma a forma

$$\begin{pmatrix} n+1 & \sum_{i=0}^n x_i & ... & \sum_{i=0}^n x_i^m \\ \sum_{i=0}^n x_i & \sum_{i=0}^n x_i^2 & ... & \sum_{i=0}^n x_i^{m+1} \\ .... & ... & ... & ... \\ \sum_{i=0}^n x_i^m & \sum_{i=0}^n x_i^{m+1} & ... & \sum_{i=0}^n x_i^{2m} \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ ... \\ \alpha_m \end{pmatrix} = \begin{pmatrix} \sum_{i=0}^n f_i \\ \sum_{i=0}^n x_i f_i \\ ... \\ \sum_{i=0}^n x_i^m f_i \end{pmatrix}. \tag{8}$$

Voltemos ao exemplo ilustrado na Figura 1. Para uma dada lista de pontos pretendemos adequar um polinómio. Na Figura 1 considerámos um polinómio de grau 1. Vamos agora considerar ajustamentos por polinómios de grau superior. Na Figura 3 calculámos pelo método dos mínimos quadrados os polinómios de grau 8 e de grau 12 que melhor se ajustam aos pontos apresentados. Para cada um dos exemplos calculámos D e obtivémos respectivamente no caso de polinómios de grau 1, 8 e 12 os valores $D_1 \approx 0.21$, $D_8 \approx 0.1$ e $D_{12} = 0$. Observamos que quando consideramos polinómios de grau superior, na verdade o que estamos a fazer é aumentar o número de funções base ϕ_j e consequentemente "alargar" o conjunto de funções que podem ser geradas pela base de funções. Sobre cada conjunto de funções base, a aproximação dos mínimos quadrados é aquela que minimiza o valor de D. Assim basta pensar que se tivermos dois conjuntos $A, B \subset \mathbb{R}$, então

$$A \subseteq B \Rightarrow min(A) \geq min(B)$$

para concluir que aumentar o número de funções de base, em geral diminui o valor de D. No caso em que consideramos polinómios de grau $n+1$, obtemos exactamente o polinómio interpolador e nesse caso $D = 0$ (Figura 3). No entanto como é visível no segundo gráfico da Figura 3 o polinónimo interpolador apresenta grandes oscilações nos extremos do intervalo o que indicia que considerar ajustamento por polinómios de grau elevado, em geral, não será uma boa estratégia.

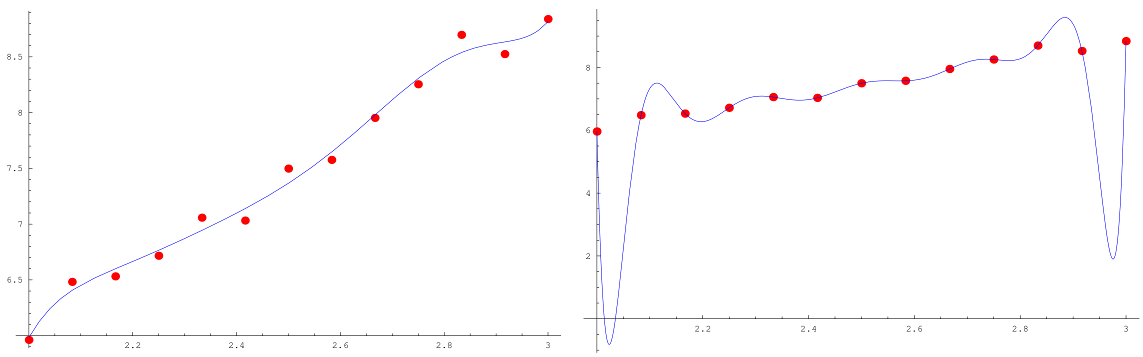

Figura 3: *dois possíveis ajustamentos (respect.) por polinómios de grau 8 e grau 12.*

Exercício

Consideremos a tabela de valores obtidos numa dada experiência

x_i	-2	-1	0	1	2
f_i	0	0	1	3	7

Determine a função da forma $g(x) = \alpha_0 x + \alpha_1 x^2$ que melhor se ajusta aos pontos da tabela segundo o critério dos mínimos quadrados.

Resolução

Neste caso temos as funções de base $\phi_0(x) = x$ e $\phi_1(x) = x^2$ e o sistema de equações normais (7) toma a forma

$$\begin{pmatrix} \sum_{i=0}^4 x_i^2 & \sum_{i=0}^4 x_i^3 \\ \sum_{i=0}^4 x_i^3 & \sum_{i=0}^4 x_i^4 \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} = \begin{pmatrix} \sum_{i=0}^4 x_i f_i \\ \sum_{i=0}^4 x_i^2 f_i \end{pmatrix}$$

e substituindo os valores obtemos

$$\begin{pmatrix} 10 & 0 \\ 0 & 34 \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} = \begin{pmatrix} 17 \\ 31 \end{pmatrix} \Rightarrow \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} = \begin{pmatrix} \frac{17}{10} \\ \frac{31}{34} \end{pmatrix}$$

pelo que a função que melhor se ajusta aos pontos da tabela, segundo o critério dos mínimos quadrados é $g(x) = \frac{17}{10}x + \frac{31}{34}x^2$. Na Figura 4 representamos os pontos da tabela e função g obtida.

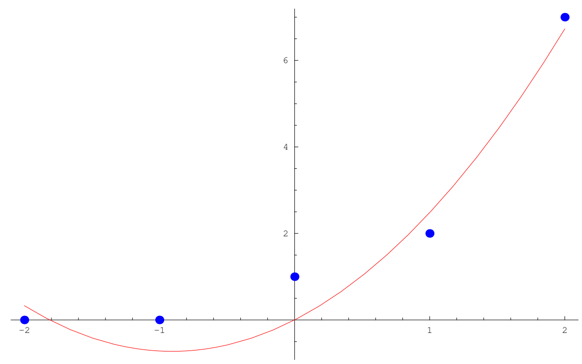

Figura 4: *A função da forma $g(x) = \alpha_0 x + \alpha_1 x^2$ que se melhor se ajusta aos pontos.*

Exercício

Consideremos o conjunto de pontos $\{x_0, x_1, x_2, x_3, x_4\} = \{-2, -1, 0, 1, 2\}$. Calcule

$$min_{\alpha_0,\alpha_1 \in \mathbb{R}} \left(\sum_{i=0}^{4} \left(x_i^2 - \alpha_0 - \alpha_1 x_i \right)^2 \right) \tag{9}$$

Resolução

Podemos resolver este problema usando a teoria do método dos mínimos quadrados. De facto se definirmos $f(x) = x^2$ e $g(x) = \alpha_0 + \alpha_1 x$ temos

$$min_{\alpha_0,\alpha_1 \in \mathbb{R}} \left(\sum_{i=0}^{4} \left(x_i^2 - \alpha_0 - \alpha_1 x_i \right)^2 \right) = min_{\alpha_0,\alpha_1 \in \mathbb{R}} \left(\sum_{i=0}^{4} \left(f(x_i) - g(x_i) \right)^2 \right) \tag{10}$$

Assim, para uma dada função g o somatório

$$D = \sum_{i=0}^{4} \left(f(x_i) - g(x_i) \right)^2$$

é a soma dos quadrados dos desvios de g à função f nos pontos x_i. Já vimos que D tem o valor mínimo quando g é aproximação dada pelo método dos mínimos quadrados. Assim

$$min_{\alpha_0,\alpha_1 \in \mathbb{R}} \left(\sum_{i=0}^{4} \left(f(x_i) - g(x_i)\right)^2 \right)$$

é atingido nos valores α_0 e α_1 que são solução do sistema de equações normais.

Definimos a tabela de valores

x_i	-2	-1	0	1	2
f_i	4	1	0	1	4

Pretendemos agora determinar a função da forma $g(x) = \alpha_0 + \alpha_1 x$ que melhor se ajusta aos pontos da tabela segundo o critério dos mínimos quadrados. Neste caso o sistema de equações normais (8) toma a forma

$$\begin{pmatrix} 5 & 0 \\ 0 & 10 \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} = \begin{pmatrix} 10 \\ 0 \end{pmatrix} \Rightarrow \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

e portanto $g(x) = 2$. Assim

$$min_{\alpha_0,\alpha_1 \in \mathbb{R}} \left(\sum_{i=0}^{4} \left(f(x_i) - g(x_i)\right)^2 \right) = \sum_{i=0}^{4} \left(f_i - 2\right)^2 = 14$$

Na Figura 5 apresentamos a função g e os pontos da tabela.

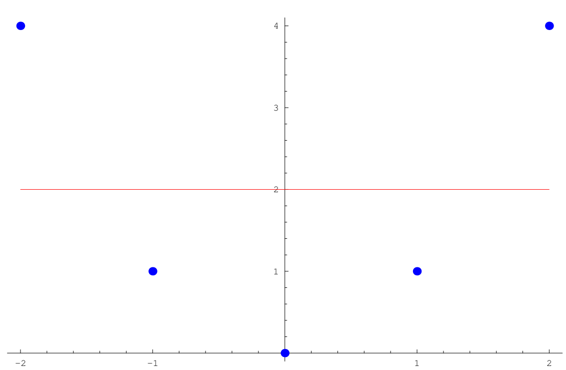

Figura 5: *A função da forma $g(x) = \alpha_0 + \alpha_1 x$ que se melhor se ajusta aos pontos.*