MS211 - Cálculo Numérico

Ajuste de Curvas e o Método dos Quadrados Mínimos: Caso Discreto

(Slides Modificados de M. E. Valle)

O Método dos Quadrados Mínimos

O *Método dos Quadrados Mínimos* é um método de *otimização* que serve para encontrar um vetor $\mathbf{x}^* \in \mathbb{R}^n$ que minimiza

$$\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2, \text{ sendo } \mathbf{A} \in \mathbb{R}^{m \times n} \text{ e } \mathbf{b} \in \mathbb{R}^m.$$

Sob certas condições, existe um único $\mathbf{x}^* \in \mathbb{R}^n$ tal que

$$\|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|_2 = \min_{x \in \mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2.$$

Note que

$$\|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|_2 = \min_{x \in \mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \Leftrightarrow \|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|_2^2 = \min_{x \in \mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2.$$

Este método *pode* ser usado para *ajustar uma curva de uma certa forma a dados*.

Ajuste de Curvas – Caso Discreto

Suponha que temos uma tabela

x	x_1	x_2	...	x_m
y	y_1	y_2	...	y_m

com $x_1, x_2, \ldots, x_m$ em um intervalo $[a, b]$. Escolhidas funções $g_1, g_2, \ldots, g_n$, geralmente contínuas, nosso objetivo será encontrar coeficientes $\alpha_1, \alpha_2, \ldots, \alpha_n$ de modo que a função

$$g(x) = \alpha_1 g_1(x) + \alpha_2 g_2(x) + \ldots + \alpha_n g_n(x),$$

satisfaça

$$g(x_k) \approx y_k, \quad \forall k = 1, \ldots, m.$$

As funções $g_1, g_2, \ldots, g_n$ podem ser escolhidas observando o gráfico dos pontos tabelados ou baseando-se em conceitos teóricos do experimento que forneceu a tabela.

O modelo matemático

$$g(x) = \alpha_1 g_1(x) + \alpha_2 g_2(x) + \ldots + \alpha_n g_n(x),$$

é chamado **linear** porque os coeficientes $\alpha_1, \ldots, \alpha_n$ aparecem linearmente. As funções $g_1, \ldots, g_n$, porém, não precisam ser funções lineares; elas podem ser polinômios, funções trigonométricas, exponenciais, logaritmos, etc.

Podemos pensar que

$$y_k = f(x_k), \quad \forall k = 1, \ldots, m,$$

para alguma função desconhecida f. Nesse caso, a função g fornece uma aproximação para f com base nos pontos amostrados $(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$.

Exemplo 1

Considere a tabela

x	-1.00	-0.75	-0.60	-0.50	-0.30	0.00	0.20	0.40	0.50	0.70	1.00
y	2.05	1.15	0.45	0.40	0.50	0.00	0.20	0.60	0.51	1.20	2.05

Podemos colocar os pontos tabelados $(x_1, y_1), \ldots, (x_{11}, y_{11})$ em um gráfico cartesiano chamado **diagrama de dispersão.**

O diagrama de dispersão sugere que os pares (x_k, y_k) pertencem à uma parábola. Portanto, escolhermos

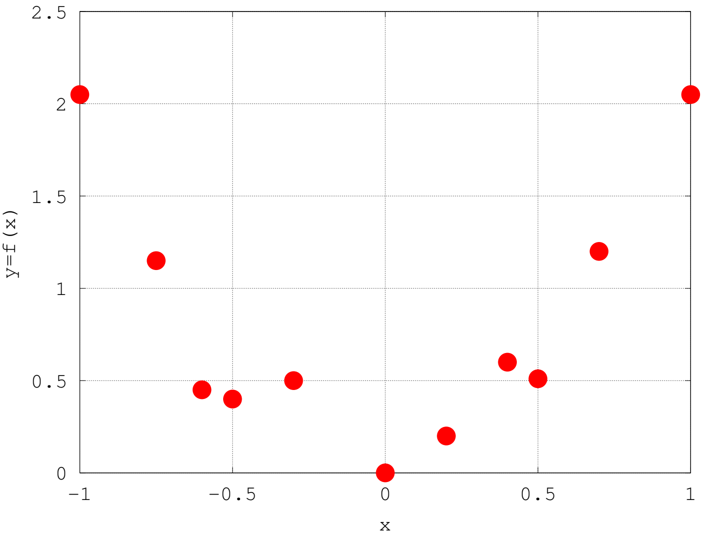
$$g_1(x) = x^2, \quad g_2(x) = x \quad \text{e} \quad g_3(x) = 1.$$

Dessa forma,

$$\begin{aligned} g(x) &= \alpha_1 g_1(x) + \alpha_2 g_2(x) + \alpha_3 g_3(x) \\ &= \alpha_1 x^2 + \alpha_2 x + \alpha_3, \end{aligned}$$

descreve uma parábola.

Diagrama de Dispersão



Formulação Matemática

Escolhidas as funções $g_1, \ldots, g_n$, no problema de quadrados mínimos, queremos determinar $\alpha_1, \ldots, \alpha_n$ tais que

$$g(x) \approx f(x) \Leftrightarrow g(x) - f(x) \approx 0 \quad \text{para } x = x_1, \ldots, x_m.$$

Equivalentemente, obtemos

$$\begin{bmatrix} g(x_1) \\ g(x_2) \\ \vdots \\ g(x_m) \end{bmatrix} \approx \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_m) \end{bmatrix}$$

Queremos que a soma dos quadrados dos desvios $g(x_k) - y_k$ seja mínima, ou seja, queremos minimizar

$$E(\alpha_1, \ldots, \alpha_n) = \|\bar{g} - \bar{f}\|^2 = \sum_{k=1}^{m} \Big(g(x_k) - y_k\Big)^2;$$

Chegando a uma Forma Matricial

Seja $g(x) = \alpha_1 g_1(x) + \dots + \alpha_n g_n(x)$

$$\begin{pmatrix} g(x_1) \\ \vdots \\ g(x_m) \end{pmatrix} \approx \begin{pmatrix} f(x_1) \\ \vdots \\ f(x_m) \end{pmatrix}$$

$$\Leftrightarrow \begin{pmatrix} \underbrace{\begin{matrix} g_1(x_1) \\ \vdots \\ g_1(x_m) \end{matrix}}_{\bar{g}_1} & \cdots & \underbrace{\begin{matrix} g_n(x_1) \\ \vdots \\ g_n(x_m) \end{matrix}}_{\bar{g}_n} \end{pmatrix} \cdot \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \approx \begin{pmatrix} f(x_1) \\ \vdots \\ f(x_m) \end{pmatrix}$$

$$\Leftrightarrow \begin{pmatrix} | & & | \\ \bar{g}_1 & \cdots & \bar{g}_n \\ | & & | \end{pmatrix} \cdot \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \approx \begin{pmatrix} f(x_1) \\ \vdots \\ f(x_m) \end{pmatrix}$$

onde $\bar{g}_i = \begin{pmatrix} g_i(x_1) \\ \vdots \\ g_i(x_m) \end{pmatrix}$ $i = 1, \dots, n$

Resumo da Forma Matricial

Queremos minimizar a distância euclidiana entre Ax e b, onde

$$A = \begin{pmatrix} | & & | \\ g_1 & \cdots & g_n \\ | & & | \end{pmatrix}, \quad x = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$$

$$\text{e } b = \begin{pmatrix} f(x_1) \\ \vdots \\ f(x_m) \end{pmatrix}$$

Dica para lembrar:

$$Ax^* \approx b \iff A^t A x^* = A^t b$$

$$\boxed{\begin{array}{l} \text{Seja } m > n, \ \text{posto}(A) = n \\ \|Ax^* - b\|_2 = \min_{x \in \mathbb{R}^n} \|Ax - b\|_2 \\ \iff \quad A^t A x^* = A^t b \end{array}}$$

Justificativa

No curso de Cálculo II, vimos que o mínimo de $E(\alpha_1, \ldots, \alpha_n)$ deve satisfazer

$$\frac{\partial E}{\partial \alpha_j} = 0, \quad \forall j = 1, \ldots, n.$$

Pela regra da cadeia, a derivada parcial é

$$\frac{\partial E}{\partial \alpha_j} = 2 \sum_{k=1}^{m} \Big(\alpha_1 g_1(x_k) + \ldots + \alpha_n g_n(x_k) - y_k\Big) g_j(x_k).$$

Dessa forma, devemos ter

$$\sum_{k=1}^{m} \Big(\alpha_1 g_1(x_k) + \ldots + \alpha_n g_n(x_k) - y_k\Big) g_j(x_k) = 0, \quad \forall j = 1, \ldots, n.$$

Equações Normais

Alternativamente, podemos escrever

$$\begin{cases} \left(\sum_{k=1}^{m} g_1(x_k)g_1(x_k)\right)\alpha_1 + \ldots + \left(\sum_{k=1}^{m} g_n(x_k)g_1(x_k)\right)\alpha_n & = \sum_{k=1}^{m} y_k g_1(x_k), \\ \left(\sum_{k=1}^{m} g_1(x_k)g_2(x_k)\right)\alpha_1 + \ldots + \left(\sum_{k=1}^{m} g_n(x_k)g_2(x_k)\right)\alpha_n & = \sum_{k=1}^{m} y_k g_2(x_k), \\ & \vdots \\ \left(\sum_{k=1}^{m} g_1(x_k)g_n(x_k)\right)\alpha_1 + \ldots + \left(\sum_{k=1}^{m} g_n(x_k)g_n(x_k)\right)\alpha_n & = \sum_{k=1}^{m} y_k g_n(x_k), \end{cases}$$

que é um sistema linear com n equações e incógnitas $\alpha_1, \ldots, \alpha_n$.

Em termos matriciais, essas equações podem ser escritas como

$$\mathbf{B}\boldsymbol{\alpha} = \mathbf{b},$$

em que $\mathbf{B} = (b_{ij}) \in \mathbb{R}^{n \times n}$, $\boldsymbol{\alpha} = (\alpha_j) \in \mathbb{R}^n$ e $\mathbf{b} = (b_i) \in \mathbb{R}^n$, com

$$b_{ij} = \sum_{k=1}^{m} g_i(x_k) g_j(x_k) \quad \text{e} \quad b_i = \sum_{k=1}^{m} y_k g_i(x_k), \quad \forall i, j = 1, \ldots, n.$$

O sistema linear acima é chamado **equação normal.**

Note que **B** e **b** do slide anterior são dados por:

- $\mathbf{B} = \mathbf{A}^T\mathbf{A}$,
- $\mathbf{b} = \mathbf{A}^T\mathbf{y}$,

sendo

$$\mathbf{A} = \begin{bmatrix} g_1(x_1) & g_2(x_1) & \dots & g_n(x_1) \\ g_1(x_2) & g_2(x_2) & \dots & g_n(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ g_1(x_m) & g_2(x_m) & \dots & g_n(x_m) \end{bmatrix} \in \mathbb{R}^{m \times n} \quad \text{e} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}.$$

Pode-se mostrar que o sistema linear

$$\mathbf{B}\boldsymbol{\alpha} = \mathbf{b},$$

possui uma **única solução** $\boldsymbol{\alpha}^* = [\alpha_1^*, \ldots, \alpha_n^*]^T$ se os vetores $\bar{\mathbf{g}}_1, \ldots, \bar{\mathbf{g}}_n$ forem **linearmente independentes**, ou - equivalentemente - se posto($\mathbf{B}$) $= n$.

Sobretudo, os coeficientes $\alpha_1^*, \ldots, \alpha_n^*$ obtidos fornecem o valor mínimo de

$$E(\alpha_1, \ldots, \alpha_n) = \sum_{k=1}^{m} \Big(g(x_k) - y_k \Big)^2.$$

Exemplo 2

Considere a tabela

x	-1.00	-0.75	-0.60	-0.50	-0.30	0.00	0.20	0.40	0.50	0.70	1.00
y	2.05	1.15	0.45	0.40	0.50	0.00	0.20	0.60	0.51	1.20	2.05

e as funções

$$g_1(x) = x^2, \quad g_2(x) = x \quad \text{e} \quad g_3(x) = 1.$$

Nesse caso, temos os vetores

$\mathbf{g}_1$	1.00	0.56	0.36	0.25	0.09	0.00	0.04	0.16	0.25	0.49	1.00
$\mathbf{g}_2$	-1.00	-0.75	-0.60	-0.50	-0.30	0.00	0.20	0.40	0.50	0.70	1.00
$\mathbf{g}_3$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Além disso, temos

$$\mathbf{A}^T = \begin{bmatrix} 1.00 & 0.56 & 0.36 & \dots & 1.00 \\ -1.00 & -0.75 & -0.60 & \dots & 1.00 \\ 1.00 & 1.00 & 1.00 & \dots & 1.00 \end{bmatrix}$$

Dessa forma, temos as equações normais

$$\underbrace{\begin{bmatrix} 2.85 & -0.25 & 4.20 \\ -0.25 & 4.20 & -0.35 \\ 4.20 & -0.35 & 11.00 \end{bmatrix}}_{\mathbf{A}^T\mathbf{A}} \underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}}_{\boldsymbol{\alpha}} = \underbrace{\begin{bmatrix} 5.87 \\ -0.11 \\ 9.11 \end{bmatrix}}_{\mathbf{A}^T\mathbf{y}},$$

cuja solução é

$$\boldsymbol{\alpha}^* = \begin{bmatrix} 1.94 & 0.10 & 0.09 \end{bmatrix}^T.$$

Concluindo, a parábola que melhor se ajusta aos dados é

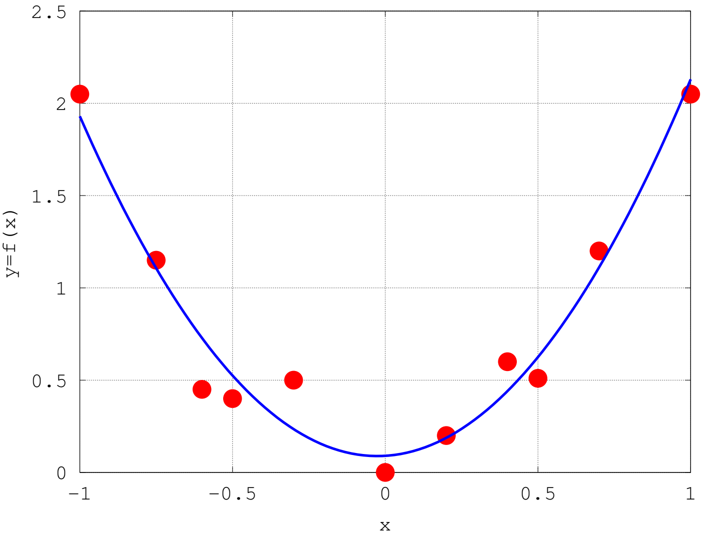
$$g(x) = 1.94x^2 + 0.10x + 0.09.$$

O mínimo da soma dos quadrados dos desvios, também chamado **resíduo**, é

$$E(\alpha_1^*, \alpha_2^*, \alpha_3^*) = \sum_{k=1}^{m} \left((1.94x_k^2 + 0.10x_k + 0.09) - y_k \right)^2 = 0.24.$$

O gráfico a seguir ilustra os dados tabelados e a parábola obtida:

Ajuste de curva



O Caso Não Linear: Um Exemplo de Linearização

Em alguns casos, o método dos quadrados mínimos linear pode ser usado para ajustar uma função g **não linear** nos coeficientes. Suponha que queremos ajustar uma função exponencial

$$g(x) = \beta_1 e^{\beta_2 x}.$$

Nesse caso, podemos linearizar o problema usando uma transformação conveniente:

$$y \approx \beta_1 e^{\beta_2 x} \quad \Longrightarrow \quad z = \ln(y) \approx \ln(\beta_1) + \beta_2 x.$$

Dessa forma, temos um problema linear

$$z \approx \alpha_1 + \alpha_2 x,$$

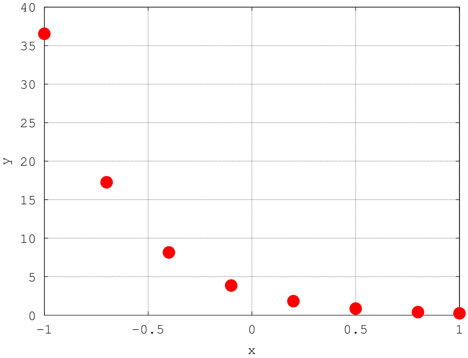
em que $\alpha_1 = ln(\beta_1)$ e $\alpha_2 = \beta_2$.

Exemplo 3

Considere a tabela

x	-1.00	-0.70	-0.40	-0.10	0.20	0.50	0.80	1.00
y	36.54	17.26	8.15	3.85	1.82	0.86	0.40	0.24

cujo diagrama de dispersão é



Linearizando do Problema

O diagrama de dispersão sugere um ajuste

$$y \approx \beta_1 e^{\beta_2 x}.$$

Fazendo a linearização $z = \ln(y)$, obtemos

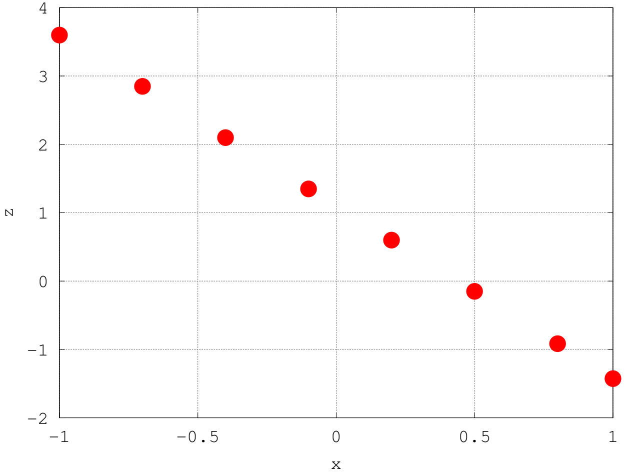
$$z \approx \alpha_1 + \alpha_2 x,$$

em que $\beta_1 = e^{\alpha_1}$ e $\beta_2 = \alpha_2$.

Portanto, podemos considerar a tabela seguinte referente o problema linearizado:

x	-1.00	-0.70	-0.40	-0.10	0.20	0.50	0.80	1.00
z	3.60	2.85	2.10	1.35	0.60	-0.15	-0.92	-1.43

Diagrama de Dispersão do Problema Linearizado



Teste de Alinhamento

Teste de Alinhamento

Supor que vc escolheu $g(x)$ com parametros α_1, ~~$\ldots, \alpha_n$~~ e α_2 para fazer ajuste de curva.

teste de alinhamento para ver se a escolha foi boa:

(a) "Linearização" de $g(x)$

($y \approx g(x) \Rightarrow z \approx$ combinação fn linear)

(b) "Diagrama de dispersão" dos z's

(c) Se os pts no diagrama em (b) estiveram alinhados (~~numa reta~~)

$\Rightarrow$ boa escolha

Resolução do Problema Linearizado

Considerando a seguinte matriz **A** e o seguinte vetor **z**:

$$\mathbf{A} = \begin{bmatrix} 1 & -1.00 \\ 1 & -0.70 \\ \vdots & \vdots \\ 1 & 1.00 \end{bmatrix}, \quad \mathbf{z} = ln(\mathbf{y}) = \begin{bmatrix} 3.6 \\ 2.85 \\ \vdots \\ -1.43 \end{bmatrix},$$

as equações normais do problema linearizado podem ser expressados como segue:

$$\underbrace{\begin{bmatrix} 8.00 & 0.30 \\ 0.30 & 3.59 \end{bmatrix}}_{\mathbf{B}=\mathbf{A}^T\mathbf{A}} \underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}}_{\boldsymbol{\alpha}} = \underbrace{\begin{bmatrix} 8.00 \\ -8.68 \end{bmatrix}}_{\mathbf{b}=\mathbf{A}^T\mathbf{z}},$$

cuja solução é

$$\boldsymbol{\alpha}^* = \begin{bmatrix} 1.09 & -2.51 \end{bmatrix}^T.$$

Concluindo, o problema linearizado fornece

$$z \approx 1.09 - 2.51x.$$

$$E_{\text{linearizado}}(\alpha_1^*, \alpha_2^*) = \sum_{k=1}^{m} \Big((1.09 - 2.51x_k) - z_k \Big)^2 = 3.2 \times 10^{-4}.$$

Retornando ao problema original, temos

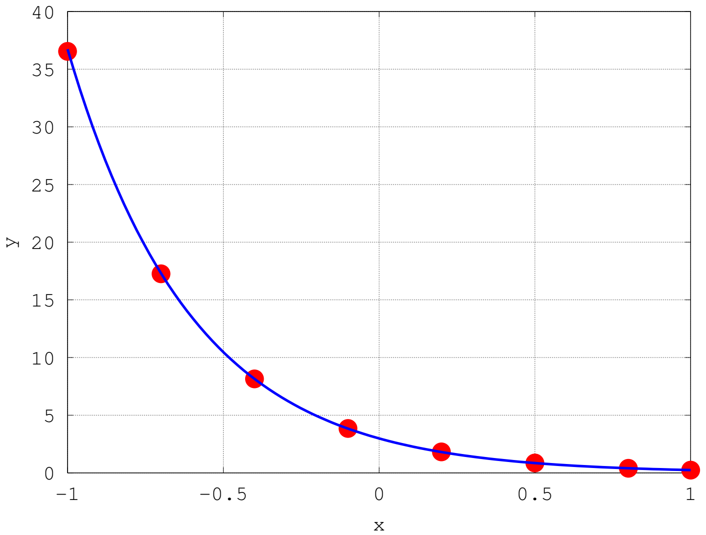
$$\beta_1^* = e^{\alpha_1} = 2.99 \quad \text{e} \quad \beta_2^* = \alpha_2 = -2.51.$$

$$\Rightarrow \quad y \approx 2.99e^{-2.51x}.$$

O resíduo do problema original, quer dizer, a soma dos quadrados dos desvios do problema original, é

$$E(\beta_1^*, \beta_2^*) = \sum_{k=1}^{m} \Big(2.99e^{-2.51x_k} - y_k \Big)^2 = 0.038.$$

Visualização do Ajuste de Curva Obtido



Observação

É importante observar que os parâmetros β_1^* e β_2^* não minimizam necessariamente

$$E(\beta_1, \beta_2) = \sum_{k=1}^{m} \left(\beta_1 e^{\beta_2 x_k} - y_k\right)^2,$$

pois eles foram obtidos através do problema linearizado, não do problema original!

Outros Exemplos de Linearização

Outros exemplos:

(1) Hipérbole

$$y \approx \frac{1}{\alpha_1 + \alpha_2 x} = g(x)$$

$$\Rightarrow z = \frac{1}{y} \approx \alpha_1 + \alpha_2 x$$

(2) Curva exponencial

$$y \approx \alpha_1 \alpha_2^x = g(x)$$

onde $y > 0$

$$\Rightarrow z = \ln(y) \approx \underbrace{\ln(\alpha_1)}_{a_1} + x \underbrace{\ln(\alpha_2)}_{a_2}$$

(3) Curva geométrica

$$y \approx \alpha_1 x^{\alpha_2} = g(x)$$

onde $x, y > 0$

$$\Rightarrow z = \ln(y) \approx \underbrace{\ln(\alpha_1)}_{a_1} + \underbrace{\alpha_2}_{a_2} \underbrace{\ln(x)}_{t}$$

(4) Curva trigonomét.

$y \approx \alpha_1 + \alpha_2 \underbrace{\cos(\omega x)}_{t}$, onde ω é constante

Considerações Finais

O método dos quadrados mínimos linear é usado para encontrar

$$g(x) = \alpha_1 g_1(x) + \alpha_2 g_2(x) + \ldots + \alpha_n g_n(x),$$

que melhor se ajusta a uma tabela

x	x_1	x_2	$\ldots$	x_m
y	y_1	y_2	$\ldots$	y_m

com $x_1, x_2, \ldots, x_m$ em um intervalo $[a, b]$.

Os coeficientes $\alpha_1^*, \ldots, \alpha_n^*$ são obtidos resolvendo o sistema

$$(\mathbf{A}^T\mathbf{A})\boldsymbol{\alpha} = \mathbf{A}^T\mathbf{y},$$

conhecido como **equação normal.**

Muito grato pela atenção!