Métodos dos quadrados mínimos no caso continuo e no caso não linear

MS211 – Cálculo Numérico

Giuseppe Romanazzi

30 Julho 2020

Conteúdo

Problemas dos quadrados mínimos

Analisaremos nesta aula os problemas e métodos de quadrados mínimos no caso contínuo. Estes problemas aparecem quando se quer aproximar ou ajustar uma curva disponibilizada $f(x)$ dentro um padrão de curvas possíveis minimizando a distancia da curva dada do padrão de curvas considerado. A melhor curva neste padrão será obtida através um método dos quadrados mínimos no caso contínuo.
Além disso vamos ver como se podem aproximar uma curva, ou dados discretos observados, usando um modelo não linear através um método dos quadrados mínimos.

Ajuste de uma curva no caso contínuo

Suponhamos de ter uma função contínua real $f(x) : [a, b] \to \mathbb{R}$ que queremos aproximar ou ajustar usando uma combinação linear de funções $\varphi(x) = \alpha_1 g_1(x) + \alpha_2 g_2(x) + \ldots \alpha_n g_n(x)$ conhecidas que descreve por exemplo o problema observado.

O método dos quadrados mínimos vai encontrar os melhores α_i tais que minimizam a distância ao quadrado entre $f(x)$ e $\varphi(x)$.
A distância (ao quadrado) entre as duas curvas continuas $f(x)$ e $\varphi(x)$ em $[a, b]$ é medida usando

$$d = \int_a^b (f(x) - \varphi(x))^2 dx$$

Esta distância estende a distancia discreta em m pontos

$$\sum_{k=1}^{m} (f(x_k) - \varphi(x_k))^2$$

usada nos problemas de quadrados mínimos para o caso continuo onde $m \to \infty$, ou seja quando temos infinitas observações continuas.

Agora em vez de um número finito de observações (x_k, y_k) com $x_k \in [a, b]$ teremos infinitos dados $(x, f(x))$ e a distancia deles do modelo sugerido $\varphi(x) = \alpha_1 g_1(x) + \alpha_2 g_2(x) + \ldots \alpha_n g_n(x)$ será medida usando o integral que como sabemos generaliza o conceito de soma no caso continuo.

$$\sum_{k=1}^{m} (f(x_k) - \varphi(x_k))^2 \rightarrow_{\text{para } k \text{ infinito}} \rightarrow \int_a^b (f(x) - \varphi(x))^2 dx$$

Métodos dos quadrados mínimos no caso contínuo

O método de quadrados mínimos no caso continuo consiste no minimizar ao variar dos α_j a função distância ao quadrado

$$F(\alpha_1, \ldots, \alpha_n) = \int_a^b (f(x) - \varphi(x))^2 dx$$

onde $\varphi(x) = \sum_{j=1}^{n} \alpha_j g_j(x)$. Procederemos como no caso discreto: minimizar F ao variar dos α_j equivale a encontrar os α_j tais que

$$\frac{\partial F}{\partial \alpha_i}(\alpha_1, \ldots, \alpha_n) = 0 \quad \text{para cada } i = 1, \ldots, n \tag{1}$$

Cada uma das n equações em cima será uma equação linear nos n coeficientes incógnitos α_j. Portanto (1) é equivalente a resolver um sistema linear $A\alpha = b$, onde $\alpha = (\alpha_1, \ldots, \alpha_n)$

Sistema linear associado ao método dos quadrados mínimos no caso continuo

Observamos que para cada $i = 1, \ldots, n$,

$$\frac{\partial F}{\partial \alpha_i} = \frac{\partial}{\partial \alpha_i} \int_a^b (f(x) - \sum_{j=1}^{n} \alpha_j g_j(x))^2 dx =$$

$$2 \int_a^b (f(x) - \sum_{j=1}^{n} \alpha_j g_j(x)) \cdot (-\frac{\partial}{\partial \alpha_i} \sum_{j=1}^{n} \alpha_j g_j(x)) dx$$

Note que $\frac{\partial}{\partial \alpha_i} \sum_{j=1}^{n} \alpha_j g_j(x) = g_i(x)$. Portanto multiplicando por -2

$$\frac{\partial F}{\partial \alpha_i} = 0 \iff \int_a^b (f(x) - \sum_{j=1}^{n} \alpha_j g_j(x)) g_i(x) dx = 0$$

A ultima igualdade pode ser escrita como a seguinte equação linear nos coeficientes α_j:

$$\sum_{j=1}^{n} (\int_a^b \; g_i(x) g_j(x) dx) \; \alpha_j = \int_a^b \; f(x) g_i(x) dx \quad \text{para cada } i = 1, \ldots, n$$

Expressão do sistema linear usando o produto escalar de funções continuas

Indicado com $(\phi, \psi) := \int_a^b \phi(x)\psi(x)dx$ o produto escalar das duas funções continuas genéricas ϕ e ψ teremos que as equação lineares para cada $j = 1, \ldots, n$

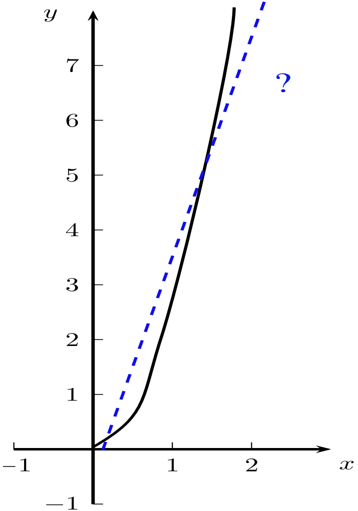
$$\sum_{j=1}^{n} \int_a^b g_i(x)g_j(x)dx\alpha_j = \int_a^b f(x)g_i(x)dx$$

podem escritas todas juntas como $A\alpha = b$ com $A = (a_{ij})$ onde $a_{ij} = (g_i, g_j)$ e $b_i = (f, g_i)$. Portanto o sistema linear para resolver tem dimensão n e matriz A simétrica de coeficientes e vetor b dados por

$$A = \begin{pmatrix} (g_1, g_1) & (g_1, g_2) & \cdots & (g_1, g_n) \\ (g_2, g_1) & (g_2, g_2) & \cdots & (g_2, g_n) \\ \vdots & \vdots & \ldots & \vdots \\ (g_n, g_1) & (g_n, g_2) & \cdots & (g_n, g_n) \end{pmatrix}, b = \begin{pmatrix} (f, g_1) \\ (f, g_2) \\ \vdots \\ (f, g_n) \end{pmatrix}$$

Exemplo

Queremos achar a melhor reta $\varphi(x) = ax + b$ que aproxima a função $f(x) = 2x^2$ no intervalo $[0, 2]$



Para determinar a melhor reta $\varphi(x) = a + bx$ usamos o método dos quadrados mínimos continuo com $n = 2$, $g_1(x) = 1$, $g_2(x) = x$ assim teremos $\varphi(x) = \alpha_1 g_1(x) + \alpha_2 g_2(x)$ com $\alpha_1 = a$ e $\alpha_2 = b$. O sistema linear $A\alpha = b$ para resolver tem dimensão 2 com

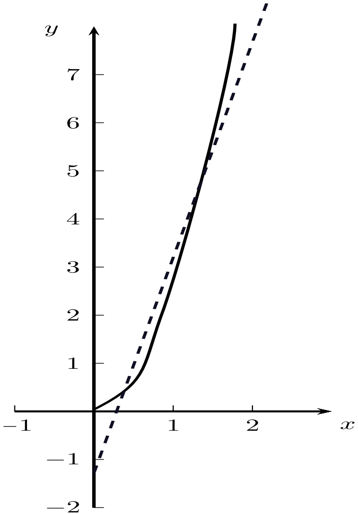
$$A = \begin{pmatrix} (g_1, g_1) & (g_1, g_2) \\ (g_2, g_1) & (g_2, g_2) \end{pmatrix} = \begin{pmatrix} \int_0^2 g_1^2(x)dx & \int_0^2 g_1(x)g_2(x)dx \\ \int_0^2 g_2(x)g_1(x)dx & \int_0^2 g_2(x)^2 dx \end{pmatrix}$$

Observamos que $a_{11} = \int_0^2 g_1^2(x) = \int_0^2 1dx = 2$,
$a_{21} = a_{12} = \int_0^2 g_1(x)g_2(x)dx = \int_0^2 xdx = \frac{x^2}{2}|_0^2 = 2$,
$a_{22} = \int_0^2 x^2 dx = \frac{x^3}{3}|_0^2 = \frac{8}{3}$.
$b_1 = \int_0^2 f(x)g_1(x)dx = \int_0^2 2x^2 dx = \frac{16}{3}$,
$b_2 = \int_0^2 f(x)g_2(x)dx = \int_0^2 2x^3 dx = 2\frac{x^4}{4}|_0^2 = 2\frac{16}{4} = 8$.

O sistema para resolver é então

$$\begin{pmatrix} 2 & 2 \\ 2 & \frac{8}{3} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} \frac{16}{3} \\ 8 \end{pmatrix}$$

A solução será $\alpha_1 = -\frac{4}{3}$ e $\alpha_2 = 4$, portanto a melhor reta será $\varphi(x) = -\frac{4}{3} + 4x$



$$y = -\frac{4}{3} + 4x$$

Problema dos quadrados mínimos no caso não linear

Não sempre o ajuste de uma função $f(x)$ em $[a, b]$ ou de um numero finito de pontos (x_k, u_k) é requerido para satisfazer um modelo linear. Por exemplo pode ser que um fenômeno satisfaz uma relação não linear exponencial $\varphi(x) = \alpha_1 + e^{\alpha_2 x}$. Notamos que esta $\varphi(x)$ é não linear respeito α_2 porque este coeficiente incógnito é argumento da função não linear exponencial. Suponhamos de ter observado o fenômeno m vezes e de ter medido os valores (x_k, y_k) com $k = 1, \dots, m$. Então:
achar os melhores α_1, α_2 que minimizam a soma das distancias(erros) ao quadrado $\sum_{k=1}^{m}(y_k - \varphi(x_k))^2$ é chamado problema dos quadrados mínimos no caso discreto não linear. Em vez se procuramos os melhores α_1 , α_2

que minimizam $\int_a^b (f(x) - \varphi(x))^2 dx$ com φ uma função não linear nos α_i será dito problema dos quadrados mínimos no caso continuo não linear.

Problema dos quadrados mínimos no caso não linear, necessitam em geral de resolver sistemas não lineares

Se estamos no caso discreto e queremos minimizar $F(\alpha_1, \alpha_2) := dsum_{k=1}^{m}(y_k - \varphi(x_k))^2$ teremos de resolver um sistema não linear de $n = 2$ incógnitas α_1, α_2. Isso porque no caso $\varphi(x) = \alpha_1 + e^{\alpha_2 x}$ por exemplo

$$\frac{\partial F}{\partial \alpha_1} = 0 \iff 2\sum_{k=1}^{m}(y_k - \alpha_1 - e^{\alpha_2 x_k})(-1) = 0 \iff$$

$$\sum_{k=1}^{m}(y_k - \alpha_1 - e^{\alpha_2 x_k}) = 0;$$

$$\frac{\partial F}{\partial \alpha_2} = 0 \iff 2\sum_{k=1}^{m}(y_k - \alpha_1 - e^{\alpha_2 x_k})(-e^{\alpha_2 x_k} x_k) = 0 \iff$$

$$\sum_{k=1}^{m}(y_k - \alpha_1 - e^{\alpha_2 x_k})(e^{\alpha_2 x_k} x_k) = 0;$$

Então teremos de resolver no nosso exemplo o seguinte sistema não linear em α_1, α_2

$$\begin{cases} \displaystyle\sum_{k=1}^{m} (y_k - \alpha_1 - e^{\alpha_2 x_k}) = 0 \\ \displaystyle\sum_{k=1}^{m} (y_k - \alpha_1 - e^{\alpha_2 x_k})(e^{\alpha_2 x_k} x_k) = 0 \end{cases}$$

Que pode ser resolvido com o método de Newton para sistemas não lineares, ver Aula 16.

Caso geral discreto não linear

No caso geral discreto não linear com a função do modelo do fenômeno $\varphi(x) = \phi(x, \alpha_1, \ldots, \alpha_n)$ que depende dos n parâmetros coeficientes incógnitos teremos de resolver o seguinte sistema não linear de n equações

$$\begin{cases} \displaystyle\sum_{k=1}^{m} (y_k - \phi(x_k, \alpha_1, \ldots, \alpha_n)) \cdot \frac{\partial \phi}{\partial \alpha_1}(x_k, \alpha_1, \ldots, \alpha_n) = 0 \\ \displaystyle\sum_{k=1}^{m} (y_k - \phi(x_k, \alpha_1, \ldots, \alpha_n)) \cdot \frac{\partial \phi}{\partial \alpha_2}(x_k, \alpha_1, \ldots, \alpha_n) = 0 \\ \ldots \\ \displaystyle\sum_{k=1}^{m} (y_k - \phi(x_k, \alpha_1, \ldots, \alpha_n)) \cdot \frac{\partial \phi}{\partial \alpha_n}(x_k, \alpha_1, \ldots, \alpha_n) = 0 \end{cases}$$

Caso geral continuo não linear

Analogamente no caso continuo onde temos ajustar a função continua $f(x)$ num modelo não continuo $\varphi(x) = \phi(x, \alpha_1, \ldots, \alpha_n)$ que depende dos n parâmetros coeficientes incógnitos α_i teremos de resolver o seguinte sistema não linear de n equações

$$\begin{cases} \displaystyle\int_a^b (f(x) - \phi(x, \alpha_1, \ldots, \alpha_n)) \cdot \frac{\partial \phi}{\partial \alpha_1}(x, \alpha_1, \ldots, \alpha_n) dx = 0 \\ \displaystyle\int_a^b (f(x) - \phi(x, \alpha_1, \ldots, \alpha_n)) \cdot \frac{\partial \phi}{\partial \alpha_2}(x, \alpha_1, \ldots, \alpha_n) dx = 0 \\ \ldots \\ \displaystyle\int_a^b (f(x) - \phi(x, \alpha_1, \ldots, \alpha_n)) \cdot \frac{\partial \phi}{\partial \alpha_n}(x, \alpha_1, \ldots, \alpha_n) dx = 0 \end{cases}$$

Linearização de algum problema dos quadrados mínimos não lineares

Algum problema não linear dos quadrados mínimos não linear pode ser linearizado permitindo de resolver ele também através a resolução de um sistema linear, que resulta ter uma metodologia numérica normalmente mais eficiente (menos memória, menos iterações, menos custo computacional) de resolver numericamente sistemas não lineares.

Por exemplo considere de ter m dados (x_k, y_k) observados de um problema descrito por $\varphi(x) = \alpha_1 e^{\alpha_2 x}$ onde se desconhecem os parâmetros α_1 e α_2.
Se continuássemos a trabalhar com a $\varphi(x)$ deste tipo teremos um problema não linear, que pode ser resolvido através a resolução numérica de um sistema não linear de duas equações obtido como na slide 16 anterior.

Mas observamos que

$$y_k \approx \varphi(x_k) = \alpha_1 e^{\alpha_2 x_k}$$

portanto teremos também que aplicando o logaritimo ln em base e

$$\ln(y_k) \approx \ln(\varphi(x_k)) = \ln(\alpha_1) + \alpha_2 x_k$$

Se denotamos $a_1 := \ln(\alpha_1)$ e $a_2 = \alpha_2$ teremos uma modelo linear que descreve o problema

$$\ln(y_k) \approx a_1 + a_2 x_k$$

- Agora com os m dados $(x_k, \ln(y_k))$ queremos achar os melhores a_1 e a_2 tais que a soma das distâncias ao quadrado de $z_k := \ln(y_k)$ respeito $a_1 + a_2 x_k$ seja a menor possível no sentido dos mínimos quadrados: $\min_{a_1,a_2} \sum_{k=1}^{m} (z_k - (a_1 + a_2 x_k))^2$

Note que este é um problema da reta dos mínimos quadrados (ver aula 22) portanto a_1,a_2 serão obtidos resolvendo o sistema linear

$$\begin{pmatrix} m & \sum_{k=1}^{m} x_k \\ \sum_{k=1}^{m} x_k & \sum_{k=1}^{m} x_k^2 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{m} z_k \\ \sum_{k=1}^{m} x_k z_k \end{pmatrix}$$

- Uma vez achados os melhores a_1 e a_2 do item anterior os melhores α_1 e α_2 que minimizam a soma dos quadrados $\sum_{k=1}^{m}(y_k - \varphi(x_k))^2$ serão necessariamente $\alpha_1 = e^{a_1}$ e $\alpha_2 = a_2$. Isso porque a função exponencial é continua e vale que

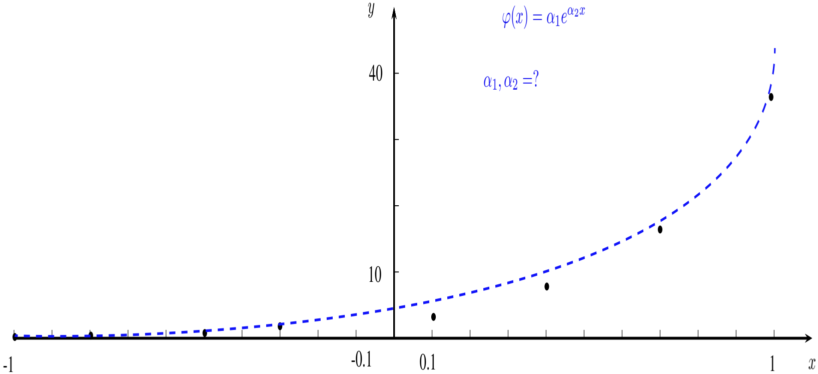
$$a_1 = \ln(\alpha_1) \iff \alpha_1 = e^{a_1}$$

e por definição $a_2 = \alpha_2$.

Exemplo não linear

x_k	-1.0	-0.8	-0.5	-0.2	0.1	0.4	0.7	1.0
y_k	0.246	0.406	0.860	1.82	3.852	8.155	17.264	36.547

Queremos achar os melhores α_1 α_2 que minimizam a soma dos quadrados das distâncias $\sum_{k=1}^{m}(y_k - \varphi(x_k))^2$ onde $\varphi(x) = \alpha_1 e^{\alpha_2 x}$ é a função não linear que descreve o fenômeno analisado.



Para poder linearizar o problema observamos que é suficiente aplicar o logaritmo e obteremos $\ln(\varphi(x)) = \ln(\alpha_1) + \alpha_2 x = a_1 + a_2 x$. Por isso obtemos uma função linear agora $\phi(x) = a_1 + a_2 x$ que descreve as $z_k := ln(y_k)$ porque $y_k \approx \varphi(x_k) \iff z_k = \ln(y_k) \approx \ln(\varphi(x_k)) = a_1 + a_2 x_k$.
Computamos os $z_k = \ln(y_k)$ usando os dados em cima

x_k	-1.0	-0.8	-0.5	-0.2	0.1	0.4	0.7	1.0
z_k	-1.402	-0.901	-0.151	0.599	1.349	2.099	2.849	3.599

O sistema linear para resolver será

$$\begin{pmatrix} m & \sum_{k=1}^{m} x_k \\ \sum_{k=1}^{m} x_k & \sum_{k=1}^{m} x_k^2 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{m} z_k \\ \sum_{k=1}^{m} x_k z_k \end{pmatrix}$$

ou seja com os dados do nosso problema

$$\begin{pmatrix} 8 & -0.3 \\ -0.3 & 3.59 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 8.041 \\ 8.6463 \end{pmatrix}$$

Obtemos então $a_1 = 1.099$ e $a_2 = 2.5$. Determinamos os parâmetros α_1, α_2 do problema original. Sendo $a_1 = \ln(\alpha_1)$ teremos $\alpha_1 = e^{a_1} = e^{1.099} = 3.001$ e $a_2 = \alpha_2 = 2.5$. Então a melhor curva exponencial que passa perto dos pontos (x_k, y_k) dados é $\varphi(x) = 3.001e^{2.5x}$.

Problemas não linearizáveis

Cuidado que não todas as $\varphi(x) = \Phi(x, \alpha_1, \ldots, \alpha_2)$ são linearizáveis aplicando a eles uma função simples se Φ for não linear. Por exemplo:

- $\varphi(x) = \alpha_1 + e^{\alpha_2 x}$ não é linearizável aplicando o logaritmo $\ln \varphi(x) = \ln(\alpha_1 + e^{\alpha_2}) \neq a + bx$.
- $\varphi(x) = \alpha_1 + \cos(\alpha_2 x)$ não é linearizável aplicando arccos
- $\varphi(x) = \alpha_1 \sin(\alpha_2 x)$ não é linearizável aplicando arcsin
- $\varphi(x) = \alpha_1 \log_{10}(\alpha_2 x)$ não é linearizável aplicando a função exponencial 10^x

Problemas PMQ linearizáveis

São poucos os casos de problemas dos mínimos quadrados que podem ser linearizados simplesmente

- $\varphi(x) = \frac{1}{\alpha_1+\alpha_2 x}$ pode ser linearizada usando a $g(y) = \frac{1}{y}$. Assim $g(\varphi(x)) = g(\frac{1}{\alpha_1+\alpha_2 x}) = \alpha_1 + \alpha_2 x$.
- $\varphi(x) = \alpha_1 \alpha_2^x$, aplicando $g(y) = \ln(y)$, obtemos $\ln(\varphi(x)) = \ln(\alpha_1) + x\ln(\alpha_2)$ esta nova função $\tilde{\varphi}(x) = \ln(\alpha_1) + x\ln(\alpha_2)$ é linear em $a_1 = \ln(\alpha_1)$ e $a_2 = \ln(\alpha_2)$ portanto $\tilde{\varphi}(x)$ é linear e podemos aplicar o problema dos mínimos quadrados nos pontos $(x_k, z_k) = (x_k, g(y_k)) = (x_k, \ln(y_k))$ achar a_1, a_2 usando o método da reta dos quadrados mínimos, e depois podemos por $\alpha_1 = e^{a_1}$, $\alpha_2 = e^{a_2}$.
- $\varphi(x) = \tan(\alpha_1 + \alpha_2 x)$ é linearizável usando $g(y) = \arctan(y)$
- $\varphi(x) = \alpha_1 e^{-\alpha_2 x}$ é linearizável usando $g(y) = \ln(y)$ $\tilde{\varphi}(x) = g(\varphi(x)) = a_1 + a_2 x$ com $a_1 = \ln(\alpha_1)$ e $a_2 = -\alpha_2$.