

SME0892 - Cálculo Numérico para Estatística

Trabalho 2 - Mínimos Quadrados Não Linear

Carolina Spera Braga - Número USP: 7161740

26 de julho, 2022

Introdução

Neste trabalho faremos uma análise do método dos Mínimos Quadrados aplicado a dois conjuntos de dados que sabemos representar uma função não linear. Quando linearizamos uma função não linear e aplicamos o método, estamos encontrando uma solução para uma combinação linear da função objeto de estudo, o que não indica uma solução ótima.

Consideremos que o conjunto de dados representa a distribuição Normal com média a e variância σ^2 dada por:

$$f(t; a, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-a)^2}{2\sigma^2}}$$

Os valores dos parâmetros a e σ que melhor aproximam o conjunto de dados (t_i, y_i) configura um problema de minimização.

Neste estudo aplicaremos tanto o método dos Mínimos Quadrados ao problema linearizado quanto o Método de Gauss-Newton para problemas de mínimos quadrados não lineares.

O método de Gauss-Newton pode ser visto como uma modificação do Método de Newton para encontrar o mínimo de uma função. Diferentemente do Método de Newton, ele apenas pode ser usado para minimizar uma soma dos valores quadrados da função, mas tem a vantagem de que as derivadas segundas, que podem ser difíceis de calcular, não são necessárias.

Método de Gauss-Newton

Para o método de Gauss-Newton consideramos a função resíduo

$$r_i(x) = y_i - f(t_i; a, \sigma), i = 1, 2, \dots, n$$

e o Jacobiano de $r_i(x)$

$$J_r(x) = \begin{bmatrix} \frac{(t-a)}{\sigma^3\sqrt{2\pi}} e^{-\frac{(t-a)^2}{2\sigma^2}} \\ -\frac{e^{-\frac{(t-a)^2}{2\sigma^2}}}{\sigma^3\sqrt{2\pi}} \left(\frac{1}{2} + \frac{1}{\sigma^2}\right) \end{bmatrix}$$

Com essas informações, em linguagem *Python*, utilizamos a função *numpy.linalg.lstsq()* para resolver o problema dos mínimos quadrados, em que

$$x = \begin{bmatrix} a \\ \sigma \end{bmatrix}$$

$$x^{(k+1)} = x^{(k)} - (J_r^T J_r)^{-1} J_r^T r(x^{(k)})$$

Através do processo iterativo obtemos os parâmetros a e σ e a norma do resíduo dividida por 2 é obtida por

$$\phi(x) = \frac{1}{2} \|r(x)\|^2$$

O algoritmo foi rodado com uma tolerância de 10^{-6} , calculada através do erro relativo, e limite máximo de iterações igual a 1000. Para o conjunto de dados 1, com chute inicial $x = (5, 1)$, o método convergiu na iteração 104, e para o conjunto de dados 2, com chute inicial $x = (2, 2)$, o método convergiu na iteração 7.

Iteração	a	σ	$r(x_i)$
1	5.602124	0.867157	2.815362
2	5.145803	0.800203	2.529144
3	5.438051	0.752014	2.300667
4	5.277063	0.715829	2.153148
5	5.355941	0.688495	2.061053
6	5.323173	0.667394	2.000743
7	5.335514	0.650668	1.957293
8	5.331747	0.637069	1.924378
9	5.332972	0.625779	1.898780
10	5.332774	0.616251	1.878588
\vdots	\vdots	\vdots	\vdots
95	5.3332	0.536536	1.788759
96	5.3332	0.536527	1.78866
97	5.3332	0.536519	1.788660
98	5.3332	0.536511	1.788661
99	5.3332	0.536504	1.7886617
100	5.3332	0.536497	1.7886624
101	5.3332	0.53649	1.788663
102	5.3332	0.536484	1.7886635
103	5.3332	0.536479	1.788664
104	5.3332	0.536474	1.7886644

Tabela 1: Convergência do método de Gauss-Newton para o conjunto de dados 1 com chute inicial (5, 1).

Iteração	a	σ	$r(x_i)$
1	3.717434	4.122696	2.500891
2	4.501416	-5.7611	2.725857
3	6.73867	35.419567	3.5224198
4	-634.716194	-16635.881042	3.730788
5	63740013024.2353	83758344200907.64	3.731248
6	-2.208797e+30	-1.732653e+19	3.731248
7	-2.208797e+30	-1.732653e+19	3.731248

Tabela 2: Convergência do método de Gauss-Newton para o conjunto de dados 2 com chute inicial (2, 2).

Dado o chute inicial do processo iterativo, podemos observar a convergência dos valores dos parâmetros aos valores ideais, representados nas Tabelas 1 e 2, e podemos verificar a convergência gráfica da função Gaussiana ao conjunto de dados.

Nas Tabelas 1 e 2 temos os resultados obtidos para os conjuntos de dados 1 e 2, respectivamente, e nas figuras 1 e 2 vemos a evolução do ajuste da função Gaussiana ao conjunto de dados para os dados 1 e 2, respectivamente.

Vemos uma boa aproximação para o conjunto de dados 1, com a função Gaussiana bem ajustada ao conjunto de dados, e não conseguimos uma aproximação para o conjunto de dados 2, para o qual não tivemos ajuste da função Gaussiana, o que indica que para o segundo caso devemos realizar o processo com outro chute inicial.

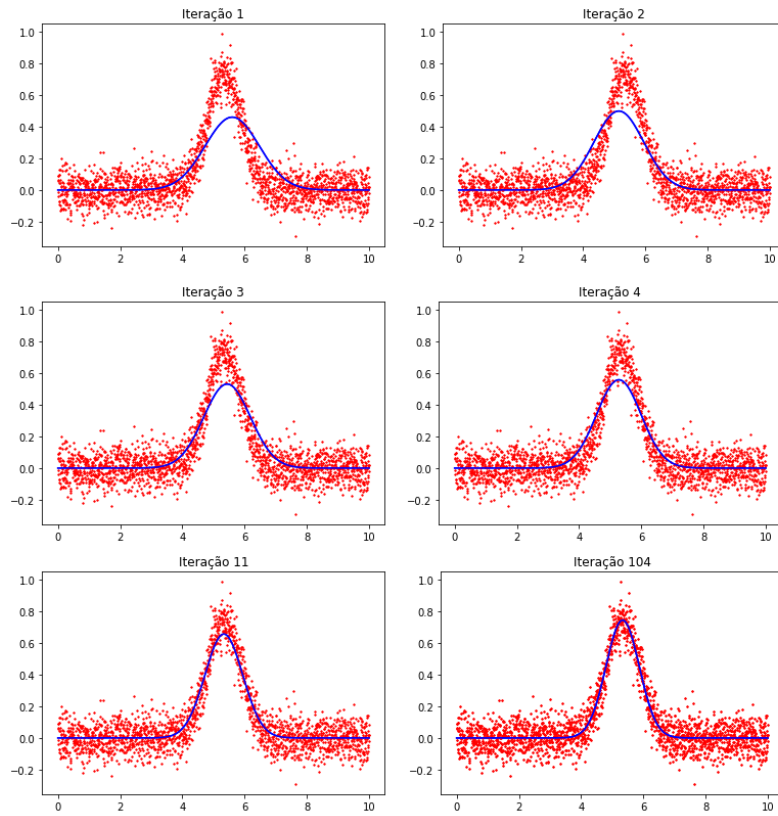


Figura 1: Ajuste da função Gaussiana ao conjunto de dados 1 pelo método de Gauss-Newton com chute inicial (5, 1).

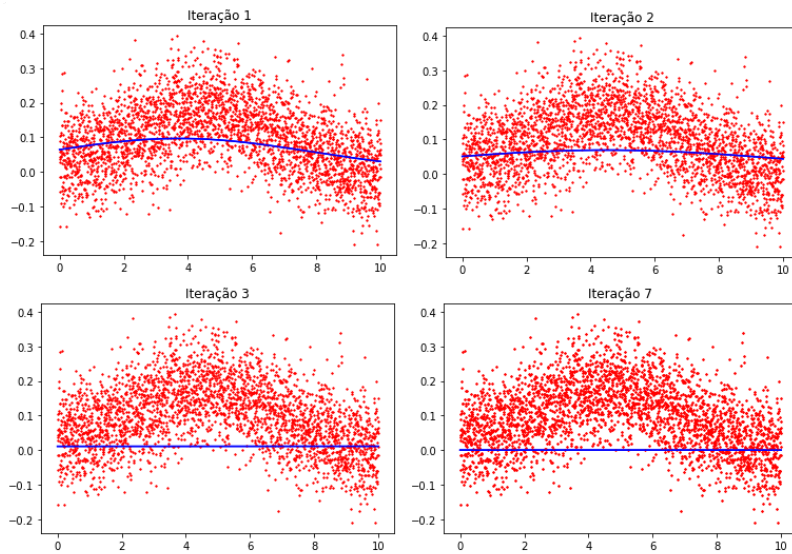


Figura 2: Ajuste da função Gaussiana ao conjunto de dados 2 pelo método de Gauss-Newton com chute inicial (2, 2).

Método dos Mínimos Quadrados ao problema linearizado

Para este caso precisamos linearizar a função Gaussiana, e quando fazemos isso encontramos o seguinte resultado:

$$\ln(y) = -(\ln(\sqrt{2\pi}) + \ln(\sigma) + \frac{a^2}{2\sigma^2} + \frac{at}{\sigma^2} - \frac{t^2}{2\sigma^2})$$

Aqui vemos que o método dos mínimos quadrados se aplica à função $\ln(y)$, o que gera uma inconsistência com nossos dados, pois estes contém números negativos para y .

Uma forma de resolver o problema é elevar ambos os lados da igualdade da função Gaussiana ao quadrado, e linearizar o problema da forma a seguir:

$$(f(t; a, \sigma))^2 = \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-a)^2}{2\sigma^2}} \right)^2$$

$$\ln(y^2) = -\ln(2\pi\sigma^2) + \frac{a^2}{4\sigma^2} - \frac{a^3t}{\sigma^2} + \frac{3a^2t^2}{2\sigma^2} - \frac{at^3}{\sigma^2} + \frac{t^4}{4\sigma^2}$$

Agora temos que o método será aplicado à função $\ln(y^2)$ que é consistente com os nossos dados, temos como base $B = \{1, t, t^2, t^3, t^4\}$, e como coeficientes $\alpha_0, \alpha_1, \alpha_2, \alpha_3$ e α_4 :

$$\alpha_0 = \frac{a^2}{4\sigma^2} - \ln(2\pi\sigma^2)$$

$$\alpha_1 = -\frac{a^3}{\sigma^2}$$

$$\alpha_2 = \frac{3a^2}{2\sigma^2}$$

$$\alpha_3 = -\frac{a}{\sigma^2}$$

$$\alpha_4 = \frac{1}{4\sigma^2}$$

Com essas informações, construímos a matriz de Vandermonde X a partir de $B = \{1, t, t^2, t^3, t^4\}$,

$$X = \begin{bmatrix} 1 & t_0 & t_0^2 & t_0^3 & t_0^4 \\ 1 & t_1 & t_1^2 & t_1^3 & t_1^4 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_n & t_n^2 & t_n^3 & t_n^4 \end{bmatrix}$$

e utilizando a função `numpy.linalg.lstsq()`, com a matriz de Vandermonde e $\ln(y^2)$ como parâmetros, encontramos os coeficientes $\alpha_0, \alpha_1, \alpha_2, \alpha_3$ e α_4 .

Para encontrar os parâmetros a e σ precisamos apenas dos valores de α_3 e α_4 . Com α_4 encontramos σ e com α_3 encontramos a .

Os resultados obtidos encontram-se na Tabela 3.

Conjunto de dados 1		Conjunto de dados 2	
a_1	σ_1	a_2	σ_2
5.177775	3.539479	4.761149	5.893628

Tabela 3: Mínimos Quadrados aplicado ao problema linearizado.

Sensibilidade do método

Para analisar a sensibilidade do método de Gauss-Newton, iremos variar o chute inicial para ambos os conjuntos de dados. Uma pequena variação no chute inicial pode fazer o método divergir, de forma que a escolha dos chutes iniciais pode parecer não diferir muito, mas os resultados obtidos se mostram muito distintos.

Podemos ver nas Tabelas 4 e 5 que com os chutes iniciais (5,1) e (6,2) obtivemos boas estimativas para os parâmetros para ambos os conjuntos de dados, e com o restante não conseguimos ajustar a função Gaussiana ao conjunto de pontos, o que pode ser visualizado nos gráficos das Figuras 3 e 4.

Chute inicial	a	σ	$r(x_i)$	Número de iterações
(1,1)	4.835637e+66	1.213314e+43	5.429664	8
(1,2)	-4.046303e+47	-2.125281e+30	5.429664	6
(2,6)	2.979275e+151	6.84512e+75	5.429664	4
(5,1)	5.333200	0.536474	1.788664	104
(6,2)	5.333200	0.536330	1.788678	114
(10,3)	1.17884e+31	-7.032115e+19	5.429664	5

Tabela 4: Método de Gauss-Newton com variações do chute inicial para o conjunto de dados 1.

Chute inicial	a	σ	$r(x_i)$	Número de iterações
(1,1)	3.922927e+77	-6.899072e+49	3.731248	11
(1,2)	2.551626e+32	-2.980064e+20	3.731248	6
(2,6)	4.1953148e+30	-2.303339e+19	3.731248	5
(5,1)	4.511657	2.313012	2.192798	28
(6,2)	4.511657	2.313019	2.192798	31
(10,3)	-5.975812e+28	-1.089221e+18	3.731248	5

Tabela 5: Método de Gauss-Newton com variações do chute inicial para o conjunto de dados 2.

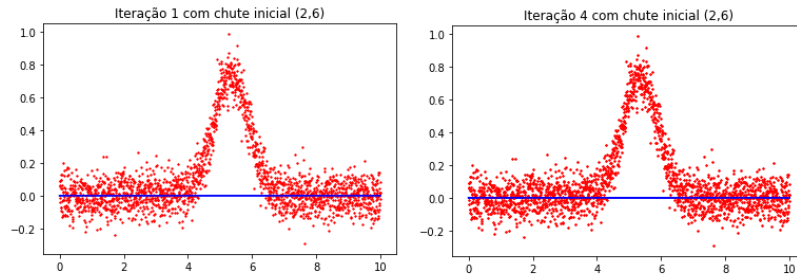


Figura 3: Ajuste da função Gaussiana ao conjunto de dados 1 pelo método de Gauss-Newton com chute inicial (2, 6).

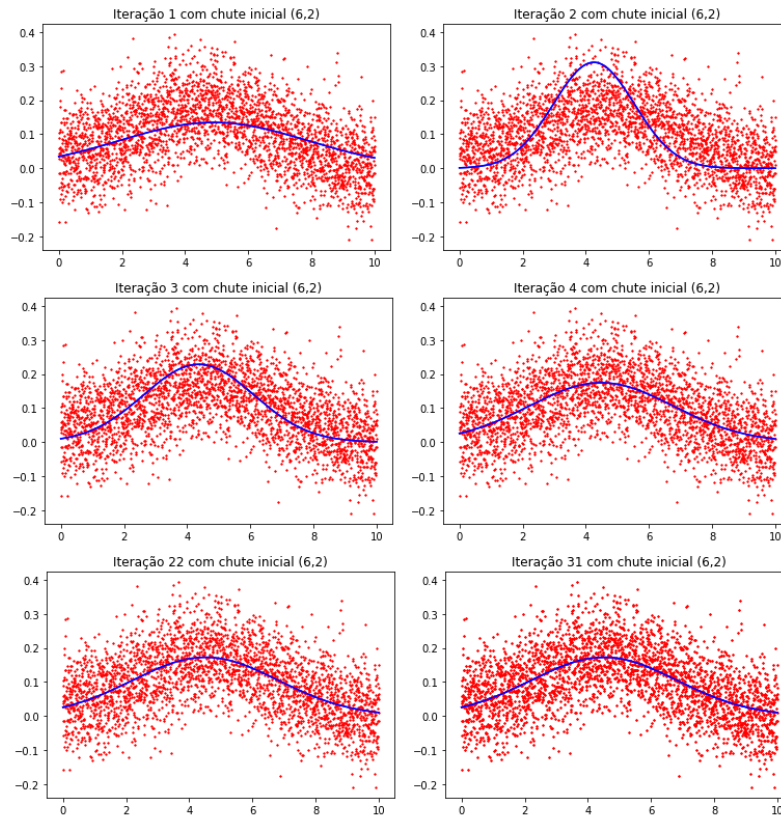


Figura 4: Ajuste da função Gaussiana ao conjunto de dados 2 pelo método de Gauss-Newton com chute inicial $(6, 2)$.

A Figura 3 representa o método de Gauss-Newton aplicado com chute inicial $(2, 6)$ ao conjunto de dados 1. Vemos que a Gaussiana não consegue se ajustar aos dados, uma situação oposta à que observamos na Figura 1, quando o procedimento foi realizado com chute inicial $(5, 1)$.

Na Figura 4 temos o método de Gauss-Newton aplicado com chute inicial $(6, 2)$ ao conjunto de dados 2. E como vimos na análise da Figura 2 que com chute inicial $(2, 2)$ a Gaussiana não se ajustava aos dados, e portanto uma mudança no chute inicial se fazia necessária, aqui resolvemos o problema, com a função Gaussiana bem ajustada ao conjunto de dados.

Comparação entre os métodos

De acordo com o que vimos sobre a sensibilidade do método de Gauss-Newton, é fácil observar que um bom chute inicial para ambos os conjuntos de dados pode ser $(5, 1)$, iremos utilizar esse chute inicial para comparar os resultados obtidos por este método com os resultados obtidos pelo método dos Mínimos Quadrados aplicado ao problema linearizado.

Para o conjunto de dados 1, obtivemos pelo método de Gauss-Newton:

- $a = 5.333200$
- $\sigma = 0.536474$

E pelo método dos Mínimos Quadrados aplicado ao problema linearizado obtivemos:

- $a = 5.177775$
- $\sigma = 3.539479$

Os valores obtidos para o parâmetro a são bem similares, a diferença maior está no valor de σ .

Diferenças entre os valores obtidos são esperadas, pois como mencionado na introdução deste trabalho, quando linearizamos a Gaussiana estamos utilizando mínimos quadrados a uma combinação linear da função original, no caso deste trabalho estamos aplicando o método na função $\ln(y^2)$ e portanto obtemos os estimadores para os parâmetros desta função e não da função y .

Para o conjunto de dados 2, obtivemos pelo método de Gauss-Newton:

- $a = 4.511567$

- $\sigma = 2.313012$

E pelo método dos Mínimos Quadrados aplicado ao problema linearizado obtivemos:

- $a = 4.761149$
- $\sigma = 5.893628$

Neste caso podemos concluir o mesmo do conjunto de dados 1. Observamos uma variância maior para o conjunto de dados 2 em relação ao 1, o que é visível nos gráficos apresentados, pois os pontos estão mais dispersos em relação ao conjunto de dados 1.

Interpolação Quadrática

A interpolação quadrática foi implementada pela forma de Lagrange, que é calculada por

$$P_2(t) = y_0 l_0 + y_1 l_1 + y_2 l_2$$

Com l_0 , l_1 e l_2 dados por

$$l_0 = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)}$$

$$l_1 = \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)}$$

$$l_2 = \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)}$$

Foram utilizados os pontos extremos e o ponto médio de (t_i, y_i) do conjunto de dados 1, para obtenção dos coeficientes α_0 , α_1 e α_2 de

$$P_2 = \alpha_0 + \alpha_1 t + \alpha_2 t^2$$

O resultado obtido foi

$$P_2 = -0.29081064 + 0.03861545t + 0.00889886t^2$$

Ao utilizar o coeficiente do termo quadrático no método de Gauss-Newton, alteramos o passo de descida e temos o método de Gauss-Newton amortecido, que pode ser representado por

$$x^{(k+1)} = x^{(k)} - (J_r^T J_r)^{-1} J_r^T r(x^{(k)}) \alpha_k$$

Onde α_k representa o passo de descida. Nas análises feitas até agora utilizamos $\alpha_k = 1$ que representa o método de Gauss-Newton puro, agora vamos analisar o método de Gauss-Newton amortecido, com $\alpha_k = 0.00889886$.

Primeiramente notamos, ao rodar o código para o método de Gauss-Newton amortecido com chute inicial (1, 1), que ele para ao atingir o limite máximo de iterações proposto igual a 1000, e apresenta os resultados

- $a = 5.332093$
- $\sigma = 0.681338$
- $r(x) = 2.037381$

Resultados estes que estão de acordo com o método de Gauss-Newton puro para os chutes iniciais considerados ótimos, no caso do chute inicial utilizado (1, 1) com o método puro verificamos uma divergência, enquanto que com o método amortecido conseguimos estimativas boas para os parâmetros.

Ao rodar o método amortecido com chute inicial (5,1), temos que o limite máximo de iterações é atingido e obtemos os resultados

- $a = 5.332639$
- $\sigma = 0.635817$
- $r(x) = 1.921452$

Conforme mudamos o chute inicial percebemos que as estimativas para os parâmetros são sempre muito similares às mostradas acima, o que indica que o método de Gauss-Newton amortecido apresenta uma convergência certa não importando muito qual o chute inicial escolhido, enquanto que o método puro necessita de uma precisão maior ao se escolher o chute inicial, de forma que é fácil acontecer de o método divergir para pequenas variações do chute inicial.

Podemos concluir que o método de Gauss-Newton amortecido tem uma melhora considerável na sensibilidade em relação ao chute inicial em comparação com o método puro.

Com relação à convergência do processo iterativo, conforme aumentamos o limite máximo de iterações, percebemos que o método de Gauss-Newton amortecido tem uma ótima convergência, ajustando muito bem a função Gaussiana ao conjunto de dados. Em contrapartida, o método puro diverge muito facilmente.

O método puro tem a vantagem de ser muito rápido para encontrar a solução do problema, mas requer uma precisão na escolha do chute inicial para não divergir, já o método amortecido apresenta uma convergência lenta mas que não requer precisão na escolha do chute inicial.