

Ajuste Não-Linear De Uma Função Exponencial A Um Conjunto De Pontos

JOÃO V. MEYER, GABRIEL S. VERNIZI, CAIO B. DEWNIG

31 de Julho de 2024

1 Introdução

1.1 Descrição do Problema

Quando temos um conjunto de pontos com ruído entre eles, que se comportam de maneira similar a uma função exponencial, podemos recorrer a ajustes não-lineares para encontrar a melhor função que se aproxima a esse conjunto de pontos.

De maneira mais técnica, isso significa achar números a e b no conjunto dos reais de modo que a função $f(x) = ae^{bx}$ melhor se aproxima a um conjunto X e Y de pontos dado. A forma que esses números a e b são encontrados leva em conta um ajuste não-linear, isso é, que não pode ser "traduzido" para um sistema linear e que utiliza outras técnicas daquelas usadas para resolver sistemas lineares.

1.2 Diferença entre Ajustes Lineares e Ajustes Não-Lineares

Um ajuste linear de uma função consiste, de maneira generalizada, em resolver o seguinte problema: dado um conjunto de m funções $\{f_1(x), f_2(x), \dots, f_m(x)\}$ e n pares ordenados $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, encontrar os coeficientes $\beta = \{a_1, a_2, \dots, a_m\}$ de modo que a função dada por

$$f(x) = \sum_{i=1}^m a_i f_i(x) = a_1 f_1(x) + a_2 f_2(x) + \dots + a_m f_m(x)$$

se aproxime dos pontos dados, isto é, minimize a soma dos resíduos ao quadrado

$$L = \sum_{i=1}^n r_i^2$$

onde r é o vetor das diferenças entre o valor esperado e o valor aproximado

$$r_i = f(x_i) - y_i$$

A linearidade do problema aparece ao calcularmos as derivadas parciais da função objetivo com respeito aos coeficientes

$$\frac{\partial L}{\partial a_i} = \frac{\partial L}{\partial r} \frac{\partial r}{\partial a_i} = 2 \sum_{j=1}^n (f(x_j) - y_j) f_i(x_j)$$

que como pode ser observado são lineares em relação aos coeficientes, tornando possível encontrar o mínimo da função objetivo resolvendo um sistema linear.

Quando consideramos um ajuste não linear, generaliza-se a função que utilizamos para aproximar os pontos, de tal forma que os parâmetros se tornam argumentos da função utilizada para a aproximação. Neste caso, o resíduo é dado por:

$$r_i = f(x_i, \beta) - y_i$$

e o gradiente é dado por:

$$\frac{\partial L}{\partial a_i} = \frac{\partial L}{\partial r} \frac{\partial r}{\partial a_i} = 2 \sum_{j=1}^n (f(x_j) - y_j) \frac{\partial f}{\partial a_i}.$$

Como é possível ver, a linearidade do problema é comprometida justamente porque não há como garantir que as derivadas parciais $\frac{\partial f}{\partial a_i}$ são lineares.

2 Metodologia

Nosso objetivo é encontrar os parâmetros a e b de tal forma que a função $f(x, a, b) = ae^{bx}$ melhor se aproxime dos pontos $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$. Para isso, devemos ter um critério objetivo do que é nossa função estar próxima dos pontos. No Método dos Quadrados Mínimos, este critério é dado pela função

$$L(a, b) = \sum_{i=1}^n (ae^{bx_i} - y_i)^2$$

que pode ser interpretada como a soma das distâncias ao quadrado da aproximação e da função real. É simples perceber que esta função representa o erro da nossa aproximação, pois quão mais distante a aproximação estiver do valor esperado, maior vai ser o valor de L .

Para encontrar o mínimo desta função, queremos achar seu ponto estacionário (onde seu gradiente é 0). Como não é possível encontrar uma fórmula direta para o 0 do gradiente, devemos recorrer aos métodos iterativos, onde uma aproximação inicial é aprimorada iterativamente até que tenhamos uma aproximação boa o suficiente.

2.1 O Problema do Valor Inicial

A utilização de métodos iterativos requer que tenhamos uma aproximação inicial, e a qualidade desta aproximação interfere diretamente na velocidade em que o método converge. Isso pode ser exemplificado a partir da seguinte analogia: imagine que você está em um ponto A e quer chegar ao ponto B. A velocidade em que você chegará ao seu objetivo depende de dois fatores: a direção que você anda e a distância inicial dos pontos. Assumindo que você tenha que ir em uma linha reta de um ponto ao outro, a velocidade com que você chegará ao ponto B depende apenas da sua distância inicial até ele. Para um método iterativo, o mesmo vale: quão mais próximo estiver sua aproximação inicial, menos passos (iterações) serão necessárias para a convergência.

A natureza do nosso problema nos permite obter uma boa aproximação inicial. Devemos perceber que por mais que $y \approx ae^{bx}$ não seja um problema com derivadas lineares, ao aplicar a função $\ln(x)$ aos dois lados da última relação, obtemos

$$\ln(y) \approx \ln(a) + bx,$$

que é um problema de quadrados mínimos linear, onde $f_1(x) = 1$ e $f_2(x) = x$. É importante notar que, como y_i não é exatamente uma exponencial de x_i , esta aproximação inicial não será perfeita, e de fato precisará de um aprimoramento.

2.2 Gradiente Descendente

O método do gradiente descendente utiliza do fato que o gradiente aponta para a direção de maior crescimento de uma função, e consequentemente, o sentido oposto do gradiente aponta para a direção de maior decrescimento da função. Portanto, se quisermos minimizar $L(a, b)$, dada uma solução inicial (a_0, b_0) , podemos "caminhar" na direção oposta ao gradiente com nossos parâmetros:

$$(a, b)_{n+1} = (a, b)_n - \lambda \nabla L(a, b)_n$$

onde $\nabla L(a, b)$ é o gradiente de L avaliado em (a, b) , e λ é um hiperparâmetro (conhecido como taxa de aprendizado), geralmente pequeno ($|\lambda| \ll 1$), que é necessário pois por mais que o oposto do gradiente nos leve ao maior decrescimento, nós não temos conhecimento de por quanto tempo aquela direção se manterá sendo uma direção de decrescimento, e portanto, por segurança, damos pequenos passos nesta direção.

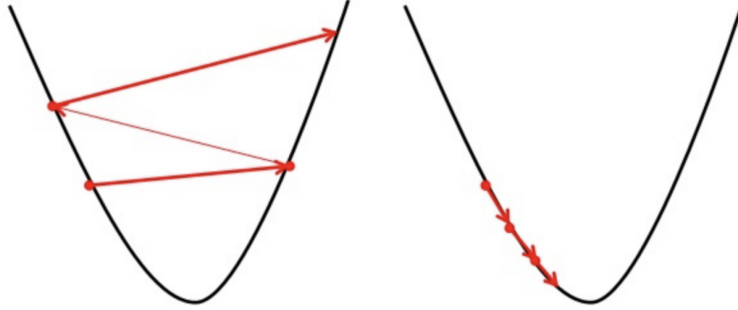


Figure 1: Diferença de uma taxa de aprendizado grande e uma pequena

2.3 Método de Gauss-Newton

O método de Newton é comumente utilizado para encontrar um zero de uma função $y = f(x)$, fazendo uma aproximação linear da função num ponto e encontrando seu zero de forma fácil, e então repetindo o processo:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

A mesma interpretação pode ser feita para várias dimensões, onde o método de Newton geralmente é utilizado para resolver sistemas não lineares: Aproximamos o sistema não linear com um sistema linear, resolvemos o sistema linear, e repetimos o processo.

Dado o sistema não linear com m funções $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$

$$\begin{cases} f_1(\mathbf{x}) = 0 \\ f_2(\mathbf{x}) = 0 \\ \vdots \\ f_m(\mathbf{x}) = 0 \end{cases} \quad (1)$$

Definimos a matriz jacobiana $J \in \mathbb{R}^{m \times n}$ pelas suas entradas

$$J_{ij} = \frac{\partial f_i(\mathbf{x})}{\partial x_j}.$$

e então o método de Newton pode ser usado para aproximar a solução do sistema dado uma solução inicial \mathbf{x}_0 , através da seguinte relação de recursão:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \Delta$$

onde Δ é a solução para o sistema linear $J\Delta = -F(\mathbf{x}_k)$, e a função $F(\mathbf{x})$ é dada por $F(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_m(\mathbf{x}))^T$. Podemos observar que quando $m >$

n , o sistema $J\Delta = -F(\mathbf{x}_k)$ é sobredeterminado, e geralmente não apresenta solução. Nesses casos, podemos usar a solução que mais aproxima Δ da solução do sistema (segundo o método dos quadrados mínimos), dado por

$$\Delta = -(J^T J)^{-1} J^T F(\mathbf{x}_k)$$

Quando isso ocorre, o sistema de equações original é aproximado, e a solução encontrada é tal que minimize as diferenças ao quadrado, que é o que desejamos. Esta generalização do método de Newton para sistemas sobredeterminados é conhecida como Método de Gauss-Newton.

Se considerarmos as funções $f_i(a, b) = r_i = ae^{bx_i} - y_i$, resolver o sistema não linear significaria encontrar os parâmetros (a, b) que levam nossa função objetivo a 0. Como isso na prática geralmente não é possível, o Método de Gauss-Newton ao invés disso encontra o ponto que mais aproxima os resíduos a 0, minimizando a soma dos resíduos ao quadrado.

2.4 Método de Levenberg-Marquardt

Considerando as mesmas condições do Método de Gauss-Newton, podemos acrescentar uma variável de "amortecimento" à expressão $(J^T J)\Delta = -J^T F(\mathbf{x}_k)$ e transformá-la em

$$(J^T J + \lambda I)\Delta = -J^T F(\mathbf{x}_k).$$

A motivação de fazer tal acréscimo é poder fazer um método que diverge menos que o Método de Gauss-Newton (ou seja, aqui não precisamos dar um valor inicial tão bom aos parâmetros).

Se isolarmos para delta a última equação, chegamos que

$$\Delta = -(J^T J + \lambda I)^{-1} J^T F(\mathbf{x}_k).$$

Perceba que a medida que aumentamos o valor da variável de amortecimento λ na última equação, mais a matriz $(J^T J + \lambda I)^{-1}$ se aproxima de uma matriz diagonal, fazendo com que nos aproximemos do Método do Gradiente Descendente. Quando tomamos um valor pequeno para λ , temos que $(J^T J + \lambda I)^{-1} \approx (J^T J)^{-1}$ o que faz com que nos aproximemos do Método de Gauss-Newton.

Agora que já determinamos o papel dessa variável de amortecimento, podemos ajustá-la conforme o método se comporta. Uma maneira de ajustar tal parâmetro é segundo a seguinte regra: se o último ajuste piorou nossa aproximação, o método está divergindo e então aumenta-se o valor dessa variável e caso aconteça o contrário diminui-se o valor da variável de amortecimento.

2.5 A Irrelevância de usar Todos os Pontos

Em situações onde temos muitos pontos em um intervalo relativamente pequeno, nós não precisamos de todos os pontos para representar nossa função, pois um subconjunto reduzido pode ser suficiente para que identifiquemos suas estruturas, e podemos então aproximar nossos parâmetros de uma forma menos custosa, utilizando menos pontos.

A escolha desses pontos pode ser feita de maneira aleatória (usando a função "randperm" do Julia), desde que seja feita de maneira bem distribuída (isso é, sem priorizar algum extremo de intervalo).

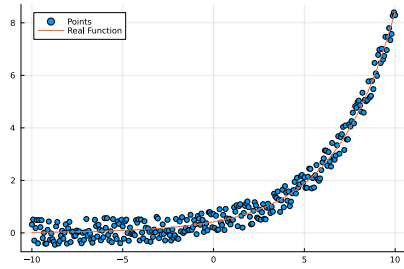


Figure 2: Muitos pontos

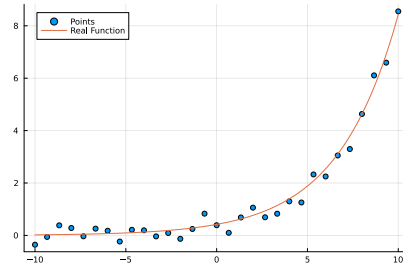


Figure 3: Poucos pontos

3 Resultados e Discussão

Aplicando o Método de Gauss-Newton, do Gradiente Descendente e de Levenberg-Marquardt à função $f(x) = a.e^{bx}$, com alguns valores de a e b preestabelecidos chegamos as seguintes aproximações:

Nota: para cada um dos exemplos foram considerados 100 pontos aleatórios gerados pela função "rand" da Julia, com um ruído de 3. Aqui o ruído produz pontos a uma certa variância (de $[-1.5, 1.5]$ entre a função real e os pontos destacados). As funções foram plotadas no intervalo $[-10, 10]$ juntamente com os pontos com ruído, para uma melhor análise dos resultados).

Nos exemplos abaixo, podemos ver que o método de Gauss-Newton e de Levenberg-Marquardt convergiram e aproximaram muito bem a função, porém o método do Gradiente descendente não foi capaz de aproximar com tanta precisão. Ao analisar os erros realizados pelas funções aproximadas, vemos que o método de Gauss-Newton convergiu mais rápido que o de Levenberg, que por sua vez convergiu mais rápido que o Gradiente. Além disso, podemos ver claramente que a vantagem do Gradiente é justamente sua estabilidade quando comparada aos outros métodos.

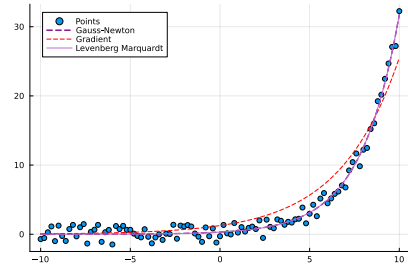


Figure 4: $a = 0.32132454$; $b = 0.4564658$

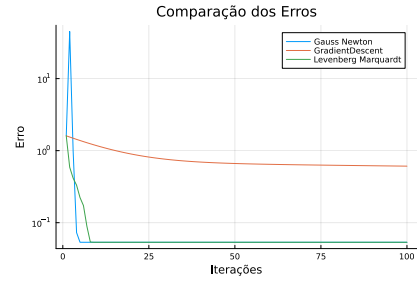


Figure 5: Erro das funções aproximadas

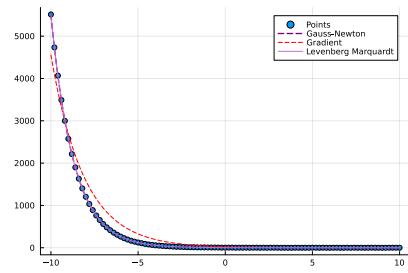


Figure 6: $a = \sqrt{3} + 1$; $b = -0.761$

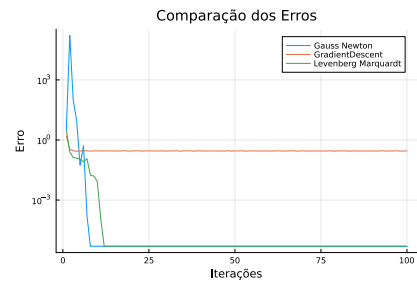


Figure 7: Erro das funções aproximadas

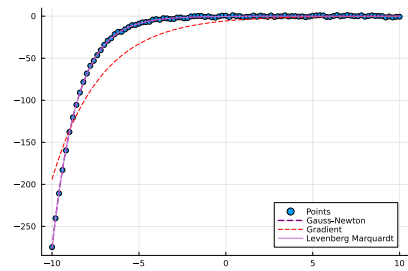


Figure 8: $a = -0.2738$; $b = -\sqrt{1.42} + 0.5$

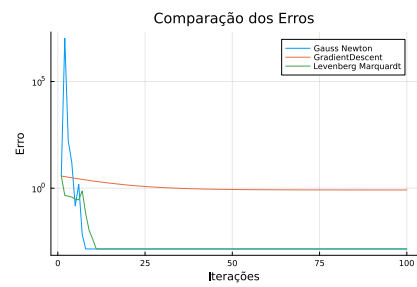
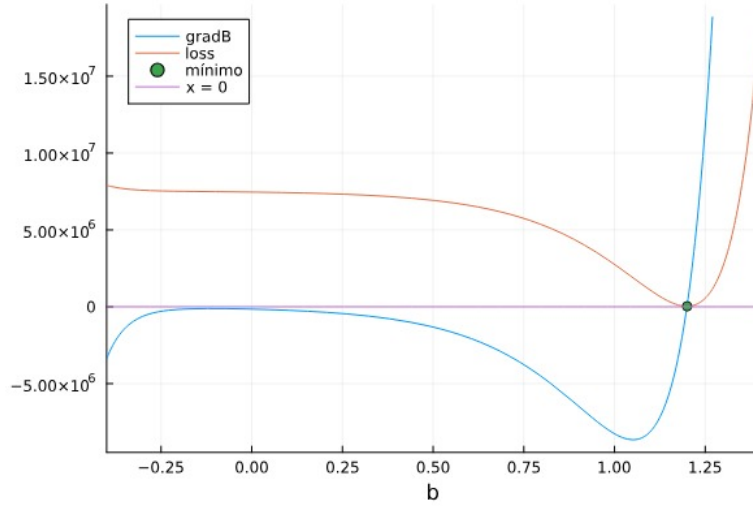


Figure 9: Erro das funções aproximadas

Vemos que em geral, em termos de velocidade, o método de Gauss-Newton e Levenberg-Marquardt superam o do Gradiente Descendente, porém a vantagem do Gradiente é que ele é o mais consistente entre todos, pois propicia pequenos incrementos ao longo do tempo, e portanto, é mais estável que os de Gauss-Newton e Levenberg-Marquardt.

Ao analisarmos o gradiente da nossa função objetivo, podemos observar alguns motivos para a baixa performance do método do gradiente:



Aproximadamente quando $b = -0.1$, podemos observar que o gradiente do parâmetro b é praticamente zero, tornando muito difícil para o método sair daquela região e ir para o mínimo. Escolher um λ maior poderia ajudar neste caso onde o gradiente é extremamente pequeno, porém iria tornar o método totalmente instável em outras regiões. Técnicas de taxa de aprendizado adaptativas poderiam tornar o método do gradiente competitivo com os outros, mas isto está fora do escopo deste trabalho

4 Conclusão

Como podemos ver, para ajustar uma função exponencial a um conjunto de dados com ruído, buscamos parâmetros a e b que minimizam a diferença entre a função $f(x) = ae^{bx}$ e os pontos dados. Isso é feito, nesse trabalho, através de métodos não-lineares, como o Método dos Quadrados Mínimos.

Em termos gerais, ajustes lineares envolvem a minimização de erros quadráticos com funções lineares em relação aos coeficientes, enquanto ajustes não-lineares

envolvem funções que não são lineares nos parâmetros, exigindo métodos iterativos para encontrar a solução.

O trabalho inteiro se voltou, de maneira generalizada, em encontrar parâmetros a e b para minimizar a função objetivo

$$L(a, b) = \sum_{i=1}^n (ae^{bx_i} - y_i)^2.$$

Utilizamos métodos iterativos como o Gradiente Descendente, o Método de Gauss-Newton e o Método de Levenberg-Marquardt para encontrar esses parâmetros (é importante ressaltar que cada um desses métodos exige um bom valor inicial para os parâmetros, pois assim necessita de menos iterações para sua convergência).

O Método do Gradiente Descendente ajusta os parâmetros na direção oposta ao gradiente da função objetivo:

$$(a, b)_{n+1} = (a, b)_n - \lambda \nabla L(a, b).$$

Já o Método de Gauss-Newton usa a seguinte relação de recursão para resolver sistemas não-lineares:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - (J^T J)^{-1} J^T F(\mathbf{x}_k).$$

Nota: em sistemas sistemas sobredeterminados, o Método de Gauss-Newton minimiza a soma dos resíduos ao quadrado.

O Método de Levenberg-Marquardt adiciona uma variável de amortecimento (λ) ao Método de Newton:

$$(J^T J + \lambda I) \Delta = -J^T F(\mathbf{x}_k).$$

Essa variável evita que o Método de Gauss-Newton divirja com mais frequência. Quando usamos valores grandes para a variável de amortecimento, mais o método se aproxima do Método do Gradiente Descendente. Quando usamos valores menores, mais ele se aproxima do Método de Gauss-Newton.

Para efeito de simplificação, em grandes conjuntos de dados, um subconjunto representativo, selecionado de forma aleatória e distribuída, pode ser usado para reduzir o custo computacional.

Finalmente, quando os métodos foram aplicados a uma função exponencial com ruído, o Método de Gauss-Newton e Levenberg-Marquardt apresentaram melhor desempenho em termos de precisão e velocidade de convergência em comparação ao Gradiente Descendente, que é mais estável, mas mais lento. Uma possível explicação para a limitação do Método do Gradiente está no fato de que para valores específicos de b , o gradiente do parâmetro b se aproxima de 0, dificultando as futuras iterações do método de ir para o mínimo.

Dificuldades: devido ao crescimento extremamente rápido da função exponencial, tivemos que lidar com muitas instabilidades nos valores produzidos para a e b quando implementamos os diferentes métodos desse trabalho. Comumente era retornado NaN , $-inf$ e inf para os valores dos parâmetros durante as iterações.

5 Referências

- Universidade Federal do Rio Grande do Sul. *Ajuste Linear Geral*. Acesso em: 20 de Julho de 2024. URL: https://www.ufrgs.br/reamat/CalculoNumerico/livro-oct/adc-ajuste_linear_geral.html
- Wikipedia. *Gauss–Newton algorithm*. Acesso em: 23 de Julho de 2024. URL: https://en.wikipedia.org/wiki/Gauss%E2%80%93Newton_algorithm
- Wikipedia. *Non-linear least squares*. Acesso em: 20 de Julho de 2024. URL: https://en.wikipedia.org/wiki/Non-linear_least_squares
- Wikipedia. *Gradient descent*. Acesso em: 20 de Julho de 2024. URL: https://en.wikipedia.org/wiki/Gradient_descent
- Wikipedia. *Levenberg–Marquardt algorithm*. Acesso em: 22 de Julho de 2024. URL: https://en.wikipedia.org/wiki/Levenberg%E2%80%93Marquardt_algorithm
- Wikipedia. *Método de Newton–Raphson*. Acesso em: 21 de Julho de 2024. URL: https://pt.wikipedia.org/wiki/M%C3%A9todo_de_Newton%E2%80%93Raphson
- O. Tingleff K. Madsen H.B. Nielsen. *METHODS FOR NON-LINEAR LEAST SQUARES PROBLEMS*. Acesso em: 25 de Julho de 2024. URL: <https://www2.imm.dtu.dk/pubdb/edoc/imm3215.pdf>
- P. J. G. Teunissen. *Nonlinear least squares*. Acesso em: 27 de Julho de 2024. URL: https://espace.curtin.edu.au/bitstream/handle/20.500.11937/38758/185936_185936.pdf?sequence=2