MODELAGEM DOS CASOS DE COVID-19 NO BRASIL PELO MÉTODO DE QUADRADOS MÍNIMOS

Eric Fukuyama e Letícia Oliveira

15 de Dezembro de 2021

UFPR

Introdução

A modelagem de casos de COVID-19 no Brasil é muito importante hodiernamente, uma vez que é uma doença que acarretou em diversas mazelas no mundo. Além disso, tal problema é importante pois é possível fazer previsões de como a quantidade de infectados variaria. Sendo assim, importante para o Estado e a população entender a problemática.

Método de Quadrados Mínimos

- Consiste em minimizar a soma dos quadrados das distâncias entre os pontos (x_i, y_i) e $(x_i, f(x_i))$;
- Logo, queremos minimizar D tal que

$$D = \sqrt{\sum_{i=1}^{n} (f(x_i) - y_i))^2}.$$

Método de Quadrados Mínimos

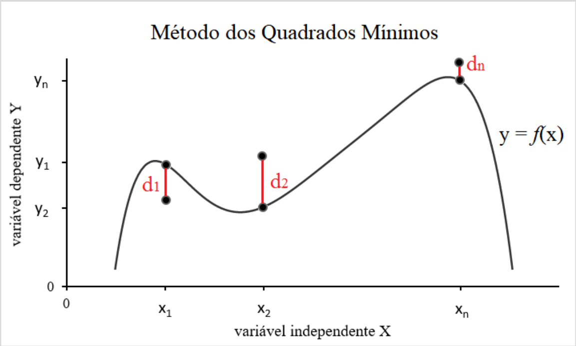

Figure 1: Método dos Quadrados Mínimos. Fonte: Autores.

Problema de Otimização

- Por simplicidade, é preferível otimizar a função E que a D, tal que E é

$$E = \sum_{i=1}^{n} (f(x_i) - y_i))^2,$$

- Nesse trabalho foram usados
    - x_i: número da semana;
    - y_i: casos reais de contaminados por COVID-19 na semana x_i;
    - $f(x_i)$: expressa o valor esperado para o número de contaminados na semana x_i;
    - n: quantidade de pontos na tabela dos dados do problema.

Problema de Otimização

Baseado em [2], as funções escolhidas para o estudo foram:

- Modelo Linear
  - $f(x) = a_0 + a_1 x$;
  - $E(a_0, a_1) = \sum_{i=1}^{n} (a_0 + a_1 x_i - y_i))^2$;
- Modelo Quadrático
  - $f(x) = a_0 + a_1 x + a_2 x^2$;
  - $E(a_0, a_1, a_2) = \sum_{i=1}^{n} (a_0 + a_1 x_i + a_2 x_i^2 - y_i))^2$;
- Modelo Cúbico
  - $f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3$;
  - $E(a_0, a_1, a_2, a_3) = \sum_{i=1}^{n} (a_0 + a_1 x_i + a_2 x_i^2 + a_3 x_i^3 - y_i))^2$;

Assim, os coficientes a_j, tal que $j = 0, 1, 2$ ou 3 são as incógnitas do nosso problema.

Modelagem Matemática

Para a Modelagem foi utilizado o seguinte teorema presente em [3]

Teorema: Seja $f : \mathbb{R}^n \to \mathbb{R}$ diferenciável no ponto $x^* \in \mathbb{R}^n$. Se x^* é um minimizador local de f, então

$$\nabla f(x^*) = 0.$$

Logo, se busca a solução dos 3 problemas que seguem

- Para o Modelo Linear

$$\begin{bmatrix} 2\sum_{i=1}^{n}(a_0 + a_1 x_i - y_i) \\ 2\sum_{i=1}^{n} x_i(a_0 + a_1 x_i - y_i) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

o que pode ser reescrito como a seguinte multiplicação de matrizes

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix}.$$

- Para o Modelo Quadrático

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 \\ \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 & \sum_{i=1}^{n} x_i^4 \end{bmatrix} . \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} x_i^2 y_i \end{bmatrix} .$$

- Para o Modelo Cúbico

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 & \sum_{i=1}^{n} x_i^4 \\ \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i^3 & \sum_{i=1}^{n} x_i^4 & \sum_{i=1}^{n} x_i^5 \\ \sum_{i=1}^{n} x_i^3 & \sum_{i=1}^{n} x_i^4 & \sum_{i=1}^{n} x_i^5 & \sum_{i=1}^{n} x_i^6 \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} x_i^2 y_i \\ \sum_{i=1}^{n} x_i^3 y_i \end{bmatrix}.$$

Método do Gradiente

O Método do Gradiente foi implementado seguindo o algoritmo baseado em [3]:

Algoritmo 1.
Dados $f(z_i)$, $\nabla(z_i)$ e z^0
$k = 0$, $t = 1$, $\gamma = 0.5$ e $\eta = 0.1$
REPITA enquanto $\nabla f(z^k) > 1e - 4$ e $k < 10000$
    Defina $d^k = -\nabla f(z^k)$
    REPITA enquanto $f(z^k + td) > f(z^k) + \eta t \nabla f(z^k)^T d$
        $t_k = \gamma t_{k-1}$
    Faça $z^{k+1} = z^k + t_k d^k$
    k = k + 1
Retorne z

Método de Newton

O Método de Newton foi implementado seguindo o algoritmo baseado em [3]:

Algoritmo 2.
Dados $f(z_i)$, $\nabla(z_i)$ e z^0
$k = 0$, $t = 1$, $\gamma = 0.5$ e $\eta = 0.1$
REPITA enquanto $\nabla f(z^k) > 1e - 4$ e $k < 10000$
    Defina $d^k = -(\nabla^2 f(z^k))^{-1} \nabla f(z^k)$
    REPITA enquanto $f(z^k + td) > f(z^k) + \eta t \nabla f(z^k)^T d$
        $t_k = \gamma t_{k-1}$
    Faça $z^{k+1} = z^k + t_k d^k$
    k = k + 1
Retorne z

Resultados

- Os dados utilizados em nossos experimentos numéricos foram obtidos em [1].
- Foram analisados dados dos casos de COVID-19 no período de 08 de Agosto de 2020 até 06 de Agosto de 2021.

Resultados

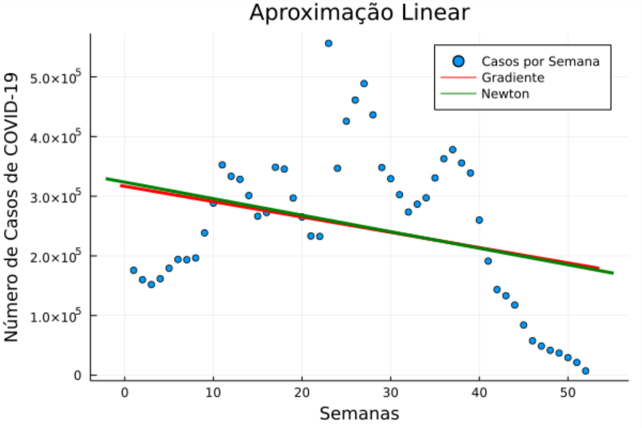

Figure 2: Aproximação Linear. Fonte: Autores.

Resultados

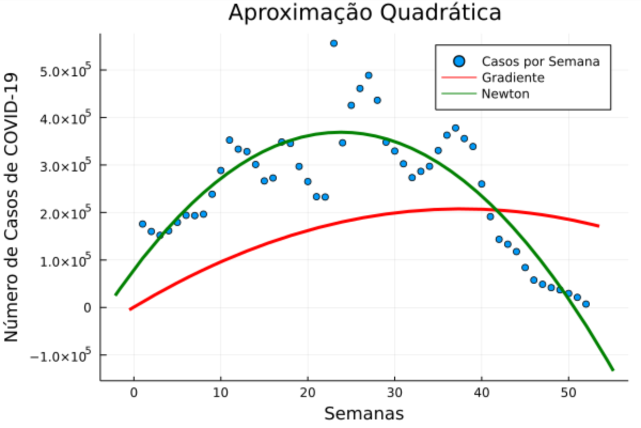

Figure 3: Aproximção Quadrática. Fonte: Autores.

Resultados

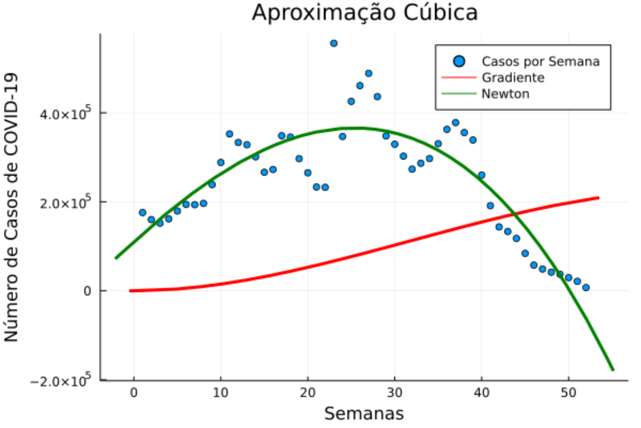

Figure 4: Aproximção Cúbica. Fonte: Autores.

Resultados

- Devido a ordem de grandeza dos dados analizados, ao avaliar qual método encontrou a melhor aproximação, vamos utilizar o Erro Relativo [4], cujo valor é determinado por:

$$E_{rel} = \frac{D}{\|y\|} = \frac{\sqrt{\displaystyle\sum_{i=1}^{n}(f(x_i) - y_i))^2}}{\sqrt{\displaystyle\sum_{i=1}^{n} y_i^2}}.$$

Resultados

Aproximação	Erro Gradiente	Erro Newton
Linear	0.428972	0.428802
Quadrática	0.576880	0.222370
Cúbica	0.817927	0.219287

Table 1: Erro Absoluto das Aproximações. Fonte: Autores

Conclusão

- Com base nos gráficos apresentados nas imagens 2, 3 e 4 e também nos valores apresentados na tabela 1, verificamos que o Método de Newton foi mais eficiente em minimizar o erro para cada um dos modelos de aproximação.

Conclusões

- O Método de Newton resolveu o problema com 1 iteração para cada modelo, enquanto o do Gradiente utilizou 10000 iterações para chegar no resultado apresentado aqui.
- A aproximação cúbica obtida pelo Método de Newton foi a que apresentou o menor erro, sendo assim, a função que melhor modelou os dados analisados.

References

[1] Fundação Oswaldo Cruz. Monitora covid-19: Casos e Óbitos. https://bigdata-covid19.icict.fiocruz.br/, 2021. Acessado: 2021-12-10.

[2] Stefani V. Marques. Método de quadrados mínimos em época de pandemia de covid-19. https://docs.ufpr.br/~ewkaras/ic/TCC_Stefani.pdf, 2020. Acessado: 2021-12-10.

[3] Ademir A. Ribeiro and Elizabeth W. Karas. *Otimização Contínua : Aspectos Teóricos e Computacionais*. Makron Books, Rio de Janeiro, 1996.

[4] Marcia A. Gomes Ruggiero and Vera Lucia da Rocha Lopes. *Cálculo Numérico : Aspectos Teóricos e Computacionais*. Makron Books, Rio de Janeiro, 1996.

Obrigado!