Relatório da disciplina de Redes Neurais

Vitor dos Santos Amorim*, Wesley Pereira Pimentel†
Universidade Federal do Espírito Santo
Vitória, Espírito Santo, Brasil
Matrículas: 2022230782, 2022231404

***Abstract*—Relatório apresentado na disciplina de Redes Neurais, ministrada pelo professor dr. Thomas Walter Rauber, como forma parcial de avaliação. O presente trabalho tem como finalidade utilizar o método de descida de gradientes para resolução de problemas de minimização.**

***Index Terms*—Inteligência Artificial, Otimização, Redes Neurais**

I. MÉTODO

Utilizou-se o método de descida de gradientes para resolver os problemas de otimização propostos neste trabalho, e adotou-se como bibliografia de referencia [1]. O cálculo da derivada ocorreu de duas formas, a primeira pelo método de diferenças finitas, conforme representado na Equação 1, e de forma analítica, utilizando-se da biblioteca *sympy*. Posteriormente, os valores obtidos foram comparados conforme apresentado na Tabela 1.

O cálculo das derivadas parciais de uma função de duas variáveis, x_1 e x_2, determinando o passo, $h = 0.01$, utilizado neste trabalho pode ser representada por:

$$
\begin{aligned}
f_x(x_1, x_2) &= \frac{f(x_1 + h, x_2) - f(x_1, x_2)}{h} \\
f_y(x_1, x_2) &= \frac{f(x_1, x_2 + h) - f(x_1, x_2)}{h}
\end{aligned}
\tag{1}
$$

Adotou-se como hiperparâmetros para o problema de minimização, o número máximo de iterações, $k_{\max}$, o tamanho da derivada, $g_{\min}$, e a taxa de aprendizagem alpha, α. O valor de cada um desses parâmetros foi definido de acordo com o problema abordado.

O programa foi implementado em linguagem de programação *Python*, com o compilador *Python* 3.8, num ambiente cuja máquina possuía um processador da Intel Core i5-8250U CPU @ 1.60GHz, Memória RAM de 8GB, no sistema operacional Ubuntu 22.04.1 LTS.

II. RESULTADOS

A. Descida de gradientes: unidimensional

A análise realizada levou em consideração a Equação 2. E foi determinado que o valor inicial para a variável, x, seria $x_0 = 1$, e a taxa de aprendizagem, alpha, $\alpha = 0.1$. Fez-se o cálculo da derivada para calcular a descida de gradientes utilizando-se o método de diferenças finitas(MDF), adotando-se o passo, $h = 0.01$. Posteriormente, foi calculada a derivada analiticamente.

$$f(x) = x(x^2 - x - 1)e^{-x} \tag{2}$$

A Equação 2 também pode ser escrita como:

$$f(x) = (x^3 - x^2 - x)e^{-x} \tag{3}$$

Derivada	**MDF**	**Analítica**	$\bar{x}$	$\lvert\sigma\rvert$
x_0	1.0000	1.0000	1.0000	0.0000
x_1	0.9627	0.9632	0.9630	0.0005

Table I: Comparação entre o próximo valor de x, utilizando-se aproximação da derivada e calculando-a de forma analítica.

Nesta próxima etapa, foi utilizado o método de descida de gradientes, cuja derivadas foram obtidas analiticamente. O valor inicial definido foi de $x_0 = 3$, e taxa de aprendizagem, alpha, $\alpha = 0.1$. As condições de parada implementadas foram o tamanho da derivada, $gmin = 0.1$ e o número máximo de iterações atingidos, $k_{max} = 100$. Obteve-se como menor valor de $f(0.789) = -0.4185$

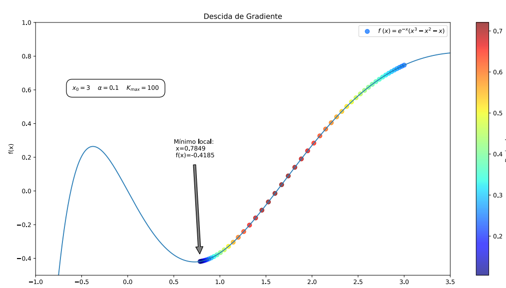

Figure 1: Atualização do valor de x, atualizando-o com base no gradiente.

B. Descida de gradientes: bidimensional

O objetivo abordado neste problema é utilizar o método de diferenças finitas, bem como o método analítico, para o cálculo da derivada de uma função de duas variáveis. A função analisada é representada pela Equação 4.

$$f(x_1, x_2) = (4 - 2.1x_1^2 + \frac{x_1^3}{3})x_1^3 + x_1x_2 + (-4 + 4x_2^2)x_2^2 \tag{4}$$

$$f(x_1, x_2) = 4x_1^3 - 2.1x_1^5 + \frac{x_1^6}{3} + x_1x_2 - 4x_2^2 + 4x_2^4 \tag{5}$$

O gradiente das funções calculados analiticamente, por meio do algoritmo estão representados pela Equação 6.

$$\nabla f(x_1, x_2) = \begin{bmatrix} \dfrac{\partial f(x_1, x_2)}{\partial x_1} \\ \\ \dfrac{\partial f(x_1, x_2)}{\partial x_2} \end{bmatrix} \tag{6}$$

$$\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1^5 - 10.5x_1^4 + 12x_1^2 + x_2 \\ \\ 16x_2^3 - 8x_2 + x_1 \end{bmatrix}$$

Executou-se o algoritmo elaborado para estimar mínimos da Equação 5, e foi implementado o método das diferenças finitas para duas variáveis, bem como o calculo do gradiente de forma analítica. A Figura 2 representa a curva de nível da Equação 4 no domínio de $-1 \leq x_1 \leq 1$, e $-1 \leq x_2 \leq 1$.

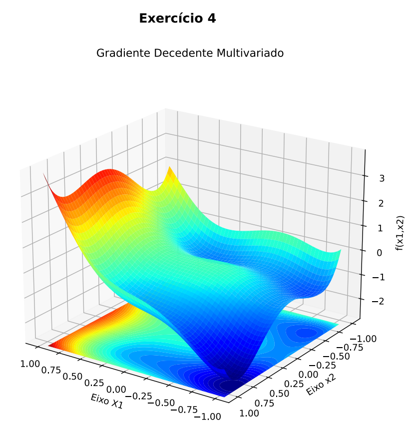

Figure 2: Superfície tridimensional

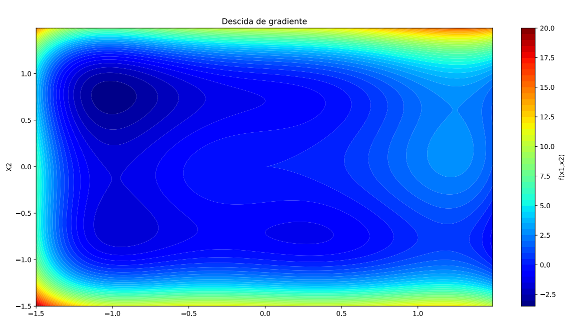

Figure 3: Superfície representado no domínio $-1.5 \leq x_1 \leq 1.5$, e $-1.5 \leq x_2 \leq 1.5$.

Utilizando a Equação 5 e o método de descida de gradiente estudado em sala de aula foi possível encontrar os seguintes valores demonstrados na Tabela 2.

	Menor valor $f(x_1, x_2)$		
Ponto Inicial	**MDF**	**Analítico**	σ
[1, 1]	-3.3034	-3.3038	0.0004
[-0.5, -0.5]	-1.9184	-1.9190	0.0005
[0,0]	0	0	0
[0.3, -0.2]	-1.1172	-1.1177	0.0004
[0.7, 1]	-3.3030	-3.3038	0.0007
[1, -0.5]	-1.1177	-1.1177	0.0000

Além da análise numérica, fez-se também uma visual dos caminhos percorridos na superfície de nível, Figura 4, utilizando outro conjunto de pontos iniciais para encontrar outros mínimos locais. Foi observado que os novos valores encontrados para função, $f(x_1, x_2)$, resultaram em mínimos equivalentes aos da Tabela 2, pois alguns pontos encontravam-se na mesma zona de convergência.

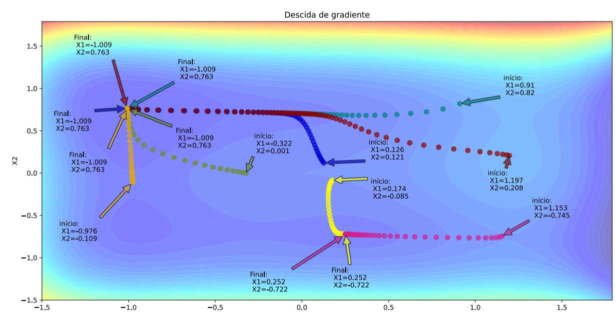

Figure 4: Superfície representado no domínio $-1.5 \leq x_1 \leq 1.5$, e $-1.5 \leq x_2 \leq 1.5$.

III. Discussão

Conforme apresentado na Tabela 2, o menor valor obtido pela função foi $f(1, 1) = -3.3038$. que representa um ponto de mínimo local, não sendo possível afirmar que se trata de mínimo global, uma vez que foi analisado apenas um subdomínio.

É exequível inferir que o menor valor encontrado para $f(x_1, x_2)$, calculando a derivada analiticamente, e por meio do método de diferenças finitas difere muito pouco para o caso estudado, todavia é recomendado utilizar a derivada em sua forma analítica a fim de evitar a propagação de erro.

References

[1] T. W. Rauber, "Aprendizagem de máquina," Set. 2022.