

CENTRO UNIVERSITÁRIO FEI  
SILVIO ROMERO DE ARAÚJO JÚNIOR

Métodos Numéricos: Descida de Gradiente

São Bernardo do Campo  
2019

SILVIO ROMERO DE ARAÚJO JÚNIOR

Métodos Numéricos: Descida de Gradiente

Relatório de atividade para a disciplina  
Programação Científica - PEL216.

São Bernardo do Campo

2019

## RESUMO

O método de descida de gradiente pode ser definido em termos computacionais, como um algoritmo para encontrar o valor mínimo de uma função, partindo da premissa que o gradiente de tal função possa ser calculado. O gradiente revela em qual direção a função diminui mais rapidamente. Desta forma, é possível buscar de forma iterativa o mínimo local da função a partir de um ponto inicial. Dada sua ampla utilização em algoritmos de aprendizado de máquina e problemas de otimização, assim implementar e avaliar uma solução computacional é de fundamental importância.

Palavras-chave: gradiente descendente, otimização, mínimos locais.

## ABSTRACT

The gradient descent method can be computationally defined as an algorithm for finding the minimum value of a function, assuming that the gradient of such a function can be calculated. The gradient reveals which direction the function decreases most rapidly. This way, it is possible to look for the iterative form the local minimum of the function after a starting point. Its wide use in machine learning algorithms and optimization problems, how to implement and evaluate a computational solution is important.

Keywords: gradient descent, optimization, local minimum.

## LISTA DE TABELAS

Tabela 1 - Dados para a Função $x^2$ .....	14
Tabela 2 - Dados para a Função $x^3 - 2x^2 + 2$ .....	15

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>11</b>
1.1	OBJETIVO.....	11
1.2	MOTIVAÇÃO.....	11
1.3	METODOLOGIA .....	11
<b>2</b>	<b>REVISÃO DA LITERATURA.....</b>	<b>12</b>
2.1	GRADIENTE DESCENDENTE.....	12
<b>3</b>	<b>IMPLEMENTAÇÃO PROPOSTA.....</b>	<b>13</b>
3.1	ALGORITMO UTILIZADO .....	13
<b>4</b>	<b>EXPERIMENTOS E RESULTADOS.....</b>	<b>14</b>
<b>5</b>	<b>TRABALHOS CORRELATOS.....</b>	<b>15</b>
<b>6</b>	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>16</b>
	<b>REFERÊNCIAS.....</b>	<b>17</b>

# 1 INTRODUÇÃO

Todo engenheiro de aprendizado de máquina está sempre procurando melhorar o desempenho de seus modelos. É aqui que entra a otimização, um dos campos mais importantes no aprendizado de máquina [RUDER, 2017]. A otimização permite selecionar os melhores parâmetros, associados ao algoritmo de aprendizado de máquina ou ao método que estamos usando, para nosso problema. Existem vários tipos de algoritmos para tal tarefa. Talvez o mais popular seja o algoritmo do gradiente descendente [HALLÉN, 2017].

## 1.1 OBJETIVO

Este trabalho tem como objetivo apresentar uma proposta de implementação do método de descida do gradiente para que se encontre o valor mínimo de uma função, utilizando a linguagem de programação C++.

## 1.2 MOTIVAÇÃO

Dada as diversas aplicações no campo de aprendizado de máquina e problemas de otimização, um estudo sistematizado do método de descida de gradiente, faz-se necessário. Além de suas aplicações nos temas supracitados, uma implementação computacional baseada em linguagem de programação comercial é algo que agrega conhecimento à disciplina de Programação Científica.

## 1.3 METODOLOGIA

Foi efetuada uma pesquisa bibliográfica com base em artigos científicos abordando os conceitos aplicados ao relatório. Após esta etapa, aplicou-se a linguagem de programação C++ para implementação computacional e análise de duas funções mencionadas devidamente no capítulo “Experimentos e Resultados” utilizando o método de descida de gradiente.

## 2 REVISÃO DA LITERATURA

A otimização matemática tem um impacto fundamental resolvendo muitos problemas relativos ao aprendizado de máquina. Este fato é aparente após uma verificação em qualquer grande conferência sobre o assunto, onde um número significativo de artigos usa técnicas de otimização [RUDER, 2017]. A descida de gradiente é um dos algoritmos mais populares para realizar tal tarefa e a maneira mais comum de otimizar as redes neurais por exemplo. Ao mesmo tempo, cada biblioteca Deep Learning de última geração contém implementações de vários algoritmos com o objetivo de otimizar a descida de gradientes [RUDER, 2017].

### 2.1 GRADIENTE DESCENDENTE

Conforme HALLÉN, 2017, o método de descida de gradiente é um algoritmo de otimização que ajuda os modelos de aprendizado de máquina a convergirem a um valor mínimo por meio de iteração. Essencialmente, a descida de gradiente é usada para minimizar uma função localizando o valor que fornece a saída mais baixa dessa função. Muitas vezes, essa função geralmente é uma função de perda. As funções de perda medem o quão ruim o nosso modelo funciona em comparação com as ocorrências reais. Portanto, só faz sentido reduzir essa perda. Uma maneira de fazer isso é via a descida de gradiente. Ele encontra os valores dos parâmetros (coeficientes) de uma função ( $f$ ) que minimiza uma função de custo. A descida de gradiente é melhor usada quando os parâmetros não podem ser calculados analiticamente (por exemplo, usando álgebra linear) e devem ser procurados por um algoritmo de otimização.



### 3 IMPLEMENTAÇÃO PROPOSTA

Na implementação do código proposto (enviado como anexo) no presente relatório utilizou-se os seguintes recursos de Hardware:

- MacBook Pro (Retina, 15-inch, Early 2013);
- Processador: 2,4 GHz Intel Core i7;
- Memória: 8 GB 1600 MHz DDR3.

Com relação aos softwares utilizados, citam-se:

- macOS Mojave Versão 10.14.5
- Xcode 10.2.1

#### 3.1 ALGORITMO UTILIZADO

Um Algoritmo de Descida de gradiente simples é o seguinte:

1. Obter uma função para minimizar  $F(x)$ .
2. Inicializar um valor  $x$  a partir do qual se deseja iniciar a descida ou otimização.
3. Especificar uma taxa de aprendizado que determinará quanto uma etapa será percorrida ou com que rapidez será a convergência para o valor mínimo.
4. Obter a derivada desse valor  $x$  (a descida).
5. Continuar a descer pela derivada desse valor multiplicado pela taxa de aprendizado.
6. Atualizar o valor de  $x$  com o novo valor decrescente.
7. Verificar sua condição de parada, isto é, a convergência.
8. Se a condição estiver satisfeita, parar. Se não, ir para o passo 4 com o novo valor de  $x$  e continuar repetindo o algoritmo até um número máximo de iterações.

## 4 EXPERIMENTOS E RESULTADOS

Depois de consultados os algoritmos por meio das referências bibliográficas, foi desenvolvido um código em linguagem C++ para execução dos experimentos. Conforme solicitado pela atividade, utilizou-se o método de descida de gradiente para encontrar o mínimo das seguintes funções  $x^2$  e  $x^3 - 2x^2 + 2$ , para  $x_0 = 2$  (ponto inicial de avaliação). Os resultados encontram-se nas tabelas 1 e 2 abaixo. Onde alfa é o critério de convergência ou valor mínimo desejado e beta é a taxa de aprendizagem.

$f(x) = x^2$ e $x_0=2$			
Alfa	Beta	Mínimo Local	Ponto
0,01	0,1	0,000023	0,004836
0,01	0,2	0,000019	0,004354
0,01	0,3	0,000011	0,003277
0,01	0,4	0,000010	0,003200
0,01	0,5	0,000000	0,000000
0,01	0,6	0,000010	0,003200
0,01	0,7	0,000011	-0,003277
0,01	0,8	0,000019	0,004354
0,01	0,9	0,000023	-0,004836
0,01	1,0	Não converge	Não converge

Tabela 1 - Dados para a Função  $x^2$

$f(x) = x^3 - 2x^2 + 2$ e $x_0=2$			
Alfa	Beta	Mínimo Local	Ponto
0,01	0,1	0,814823	1,335310
0,01	0,2	0,814821	1,331631
0,01	0,3	0,814823	1,331288
0,01	0,4	0,814822	1,331465
0,01	0,5	2,000000	0,000000
0,01	0,6	Não converge	Não converge
0,01	0,7	Não converge	Não converge
0,01	0,8	Não converge	Não converge
0,01	0,9	Não converge	Não converge
0,01	1,0	Não converge	Não converge

Tabela 2 - Dados para a Função  $x^3 - 2x^2 + 2$

A partir dos dados coletados, pode-se chegar às seguintes conclusões: quando a taxa de aprendizagem é muito pequena, os passos até o mínimo local, também são pequenos, isso significa que a convergência até o valor (alfa) estipulado é muito devagar. Poder-se-ia imaginar que para acelerar a convergência, bastaria aumentar o valor da taxa de aprendizagem (beta), porém, corre-se o risco de se perder o mínimo local e os valores não convergiriam, como na Tabela 2, por exemplo, para valores de beta maiores que 0,5. Cada vez que se aproxima do mínimo local, menos iterações são necessárias para a convergência.

## 5 TRABALHOS CORRELATOS

Alguns autores têm abordado o tema de métodos de otimização por meio da descida de gradiente (HALLÉN, 2017, RUDER, 2017). Em HALLÉN, 2017, o autor compara dois métodos: o Gradiente Descendente e o Gradiente Descendente Estocástico, comparando seus desempenhos e mostrando suas aplicações na análise de dados por meio de Regressão Logística. Já em RUDER, 2017, o autor apresenta uma visão geral dos algoritmos de otimização e aplicações do método no campo de aprendizado de máquina e redes neurais. Ambos os trabalhos formaram a base teórica do presente relatório.

## **6 CONSIDERAÇÕES FINAIS**

Este relatório apresentou os conceitos teóricos sobre o método de Descida de Gradiente. A importância deste método é essencial nos campos de Aprendizado de Máquina, Redes Neurais Artificiais e Regressão Logística. Por isso, é imperativo que se entenda o seu funcionamento. Este relatório apresentou uma forma mais simples do algoritmo de descida de gradiente, bem como sua implementação em C++. Visando uma compreensão intuitiva desse algoritmo e como aplicá-lo aos problemas do mundo real.

## REFERÊNCIAS

HALLÉN, R. **A Study of Gradient-Based Algorithms**. Gothenburg, Sweden. Lund University, 2017. Disponível em: <http://lup.lub.lu.se/student-papers/record/8904399>.

RUDER, S. **An Overview of Gradient Descent Optimization Algorithms**. Dublin, Ireland. arXiv preprint arXiv:1609.04747, 2017. Disponível em: <https://arxiv.org/abs/1609.04747v2>.