

# Gradiente estocástico e aprendizado de máquinas

Carlos Henrique Venturi Ronchi \*  
Bacharelado em Matemática - UFPR  
*carloshvronchi@gmail.com*

Prof. Abel Soares Siqueira  
Departamento de Matemática - UFPR  
*abel.s.siqueira@gmail.com*

Setembro 2016

**Palavras-chave:** Otimização, Gradiente Estocástico, Aprendizado de Máquinas.

## Resumo:

A otimização se encontra presente no dia a dia, quando se quer encontrar as melhores condições para lucrar com uma empresa, ou quando se quer minimizar os gastos de uma linha de produção. Para solucionar esses problemas muitos métodos foram criados, gerando uma teoria muito abrangente. Um dos casos é a minimização sem restrição, que consiste em minimizar uma função  $f$  sem nenhuma condição imposta no seu domínio, denotada por:

$$\min_{x \in \mathbb{R}^n} f(x), \quad (1)$$

onde  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  é chamada de função objetivo [1]. Esta função  $f$  é proveniente da modelagem do problema, como os custos da linha de produção ou o custo de funcionamento de uma empresa. Esse problema de minimização pode ser difícil de resolver, para isso foram criados alguns métodos, como o Método do Gradiente, Método dos Gradientes Conjugados e o Método de Newton.

O Método do Gradiente é um método de busca que utiliza o vetor oposto ao gradiente da função a ser minimizada como a direção de busca. Tal método se mostra muito eficiente porque não exige nenhuma informação de segunda ordem e é computacionalmente barato, além de obter resultados satisfatórios.

O Método do Gradiente pode ser utilizado no aprendizado de máquinas. Esta é uma área da inteligência artificial que desenvolve algoritmos para que o computador possa aprender, como um algoritmo que prevê se uma pessoa possui um tumor ou não dado um conjunto de informações.

Um dos problemas dessa área é como prever, qual será a função e também qual é o erro estimado. Seja  $(x^i, y^i)$  conjunto de exemplos, em que  $x^i$  são os atributos e  $y^i$  são a classe. O objetivo é, dado um conjunto de exemplos, desenvolver um algoritmo

---

\*Bolsista PAJT - CAPES

de modo que é possível prever nos atributos com apenas os dados já existentes do conjunto de exemplos. Por exemplo, se  $x_i, y_i \in \mathbb{R}$  e supormos um modelo linear

$$h(x) = \alpha_1 + \alpha_2 x,$$

ou  $x^i \in \mathbb{R}^2, y^i \in \mathbb{R}$ , e usamos um modelo quadrático

$$h(x) = \alpha_1 + \alpha_2 x_1 + \alpha_3 x_2 + \alpha_4 x_1 x_2 + \alpha_5 x_1^2 + \alpha_6 x_2^2.$$

Em ambos os casos, a dependência de  $\alpha$  é linear, mas isso não é necessário. Podemos ter um modelo mais complexo, como

$$h(x) = \alpha_1 e^{\alpha_2 x_2} + \alpha_3 x_1 + \alpha_4.$$

Em todo caso, queremos encontrar os parâmetros  $\alpha_1, \dots, \alpha_n$  tal que  $h(x^i)$  e  $y^i$  sejam os mais próximos possível. Uma maneira de quantificar isso é definir a função

$$f(\alpha) = \sum_{i=1}^m [h(x^i) - y^i]^2,$$

e procurar o minimizador de  $f$ . Para resolver este problema, podemos aplicar o Método do Gradiente.

Quando se trabalha com um banco de dados de larga escala, o Método do Gradiente já não se mostra tão eficiente, pois a cada passo do algoritmo, o método analisa todos os dados  $(x^i, y^i)$  inseridos. O gradiente de  $f$  é

$$\nabla f = \left( \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right)$$

Por essa razão, utiliza-se o Método do Gradiente estocástico, ou gradiente incremental, método que utiliza apenas um subconjunto dos dados disponíveis. Além de ser mais barato, ele também é eficiente para não ficar preso em mínimos locais, devido às direções de busca. Ele pode ficar oscilando próximo de um mínimo, mas considerando passos convenientes, os valores obtidos são suficientes.

O objetivo do trabalho é fazer uma comparação de ambos no aprendizado de máquinas, notando que o método do gradiente estocástico é mais eficiente para problemas de larga escala [2], em que há muito mais dados.

## Referências:

- [1] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, 2 edition, 2006.
- [2] Shai Shalev-Shwartz and Shai Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 1 edition, 2014.