



**Data Science
Academy**

www.datascienceacademy.com.br

Machine Learning

Momentum e Learning Rate Decay



A forma mais comum para se treinar uma rede neural profunda é conhecida como Stochastic Gradient Descent. Esta abordagem não é utilizada apenas em redes neurais, e na verdade é bastante comum em casos em que existem grandes quantidades de dados nos quais métodos mais tradicionais podem demandar muito tempo para execução.

Tanto no GD quanto no SGD, você atualiza um conjunto de parâmetros de forma iterativa para minimizar uma função de erro. Enquanto com GD, você tem que percorrer TODAS as amostras em seu conjunto de treinamento para fazer uma única atualização para um parâmetro em uma iteração particular, com SGD, por outro lado, você usa SOMENTE uma amostra de seu conjunto de treinamento para fazer a atualização para um parâmetro em uma iteração específica.

Os pesos da rede neural podem ser atualizados conforme os dados são processados e os erros são calculados (abordagem conhecida como online learning) ou ao final do processo (abordagem conhecida como batch learning). Esta segunda abordagem é geralmente mais estável. Para otimizar o processamento da rede (muitas vezes devido a grande quantidade de dados) o tamanho do batch que possui os dados para atualização dos pesos é reduzido.

O principal parâmetro que controla a atualização dos pesos é conhecido como learning rate (taxa de aprendizado). Geralmente se utiliza-se taxas pequenas, como 0.1 ou 0.01 para este parâmetro. Dois parâmetros adicionais ainda podem ser usados na equação responsável pela atualização dos pesos:

Momentum

Incorpora as propriedades da atualização de pesos anterior e faz com que os pesos continuem sendo atualizados na mesma direção mesmo quando o erro diminui.

Learning Rate Decay

Learning rate decay, é usado para diminuir o valor da learning rate conforme os erros diminuem.