

Machine Learning com Python

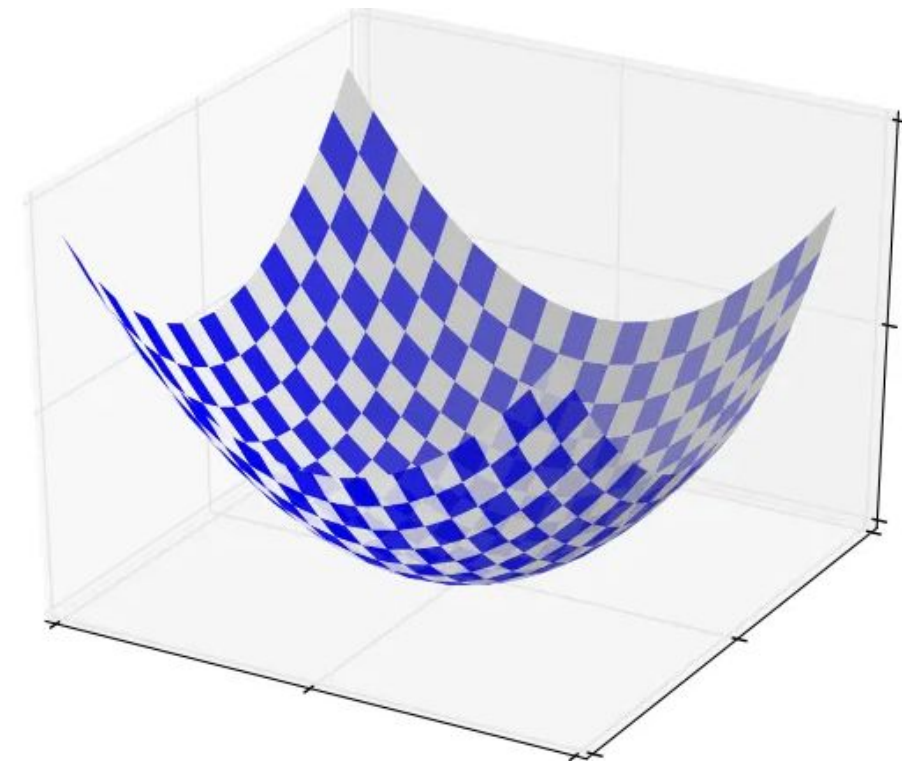
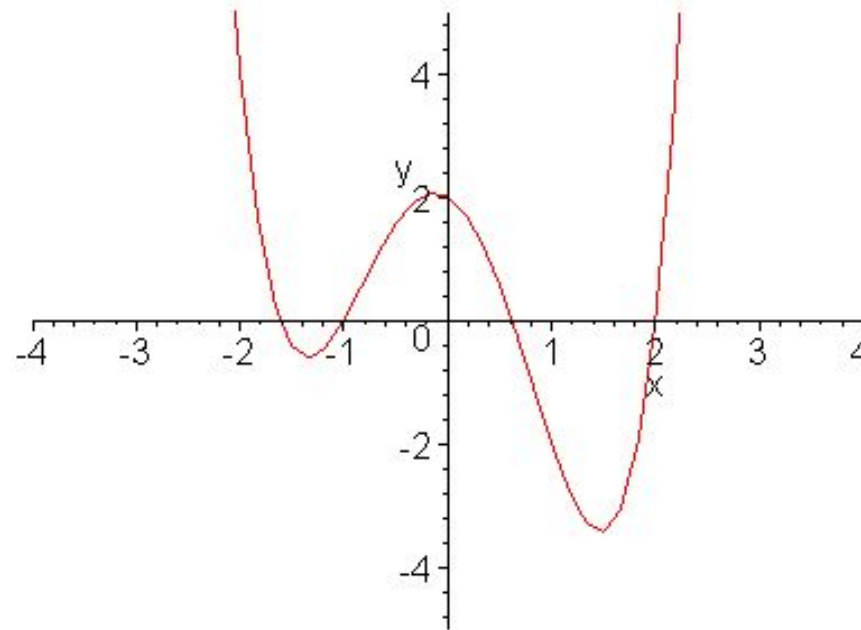
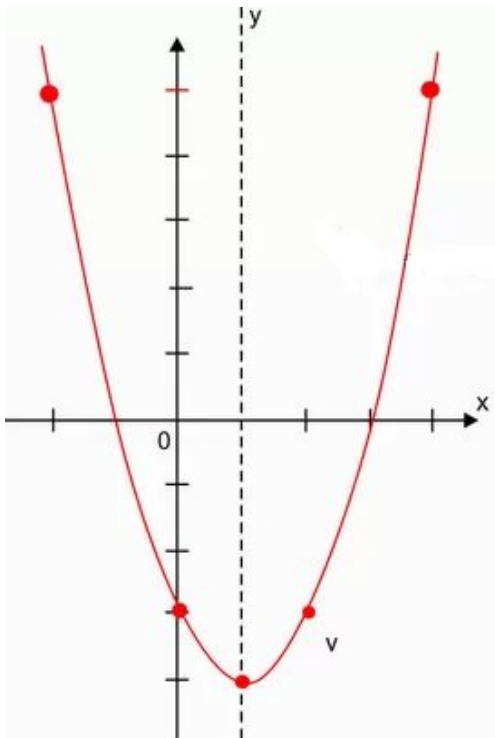
Prof. Luciano Galdino

Aprendizagem com descida do gradiente

Aprendizagem com descida do gradiente

Descida do gradiente: **algoritmo** que tem por objetivo encontrar o ponto de mínimo de uma função.

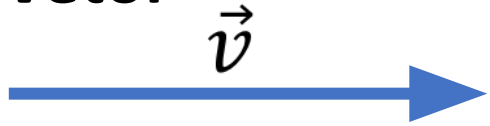
Uma função pode ter vários pontos de mínimo (mínimos locais e globais) e o objetivo é encontrar o mínimo global.



Gradiente

Vetor cujo módulo é a **derivada** direcional máxima (sentido da maior variação).
Aponta para onde a grandeza resultante da função tem seu **maior** crescimento.

Vetor



$|\vec{v}| = \text{módulo}$ (Ex.: 12)

Direção: horizontal.

Sentido: para direita.

Derivada: taxa de variação instantânea entre grandezas (variáveis).

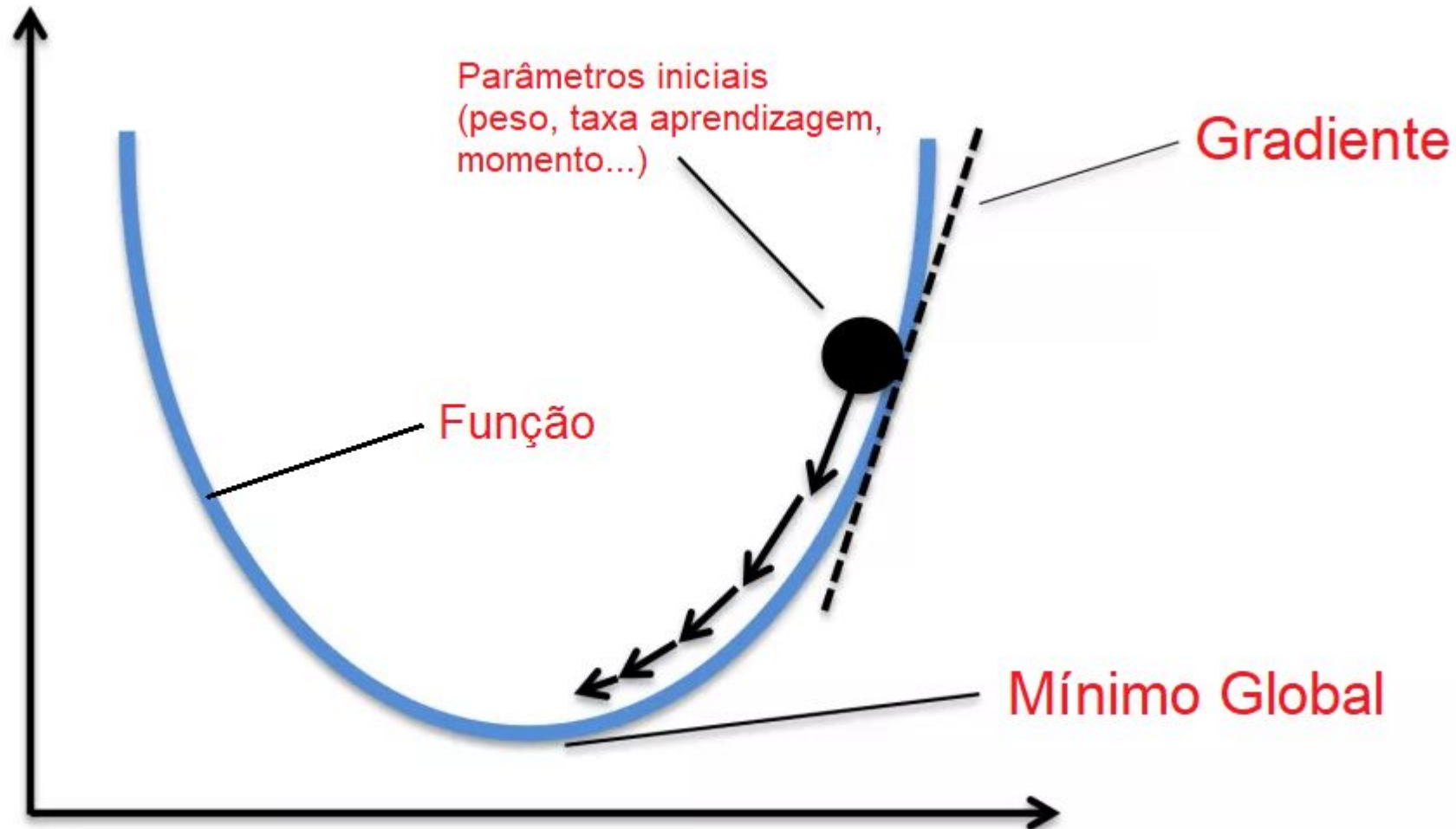
Gradiente

$$\bar{G} = \frac{\partial \Phi}{\partial x} \bar{a}_x + \frac{\partial \Phi}{\partial y} \bar{a}_y + \frac{\partial \Phi}{\partial z} \bar{a}_z$$

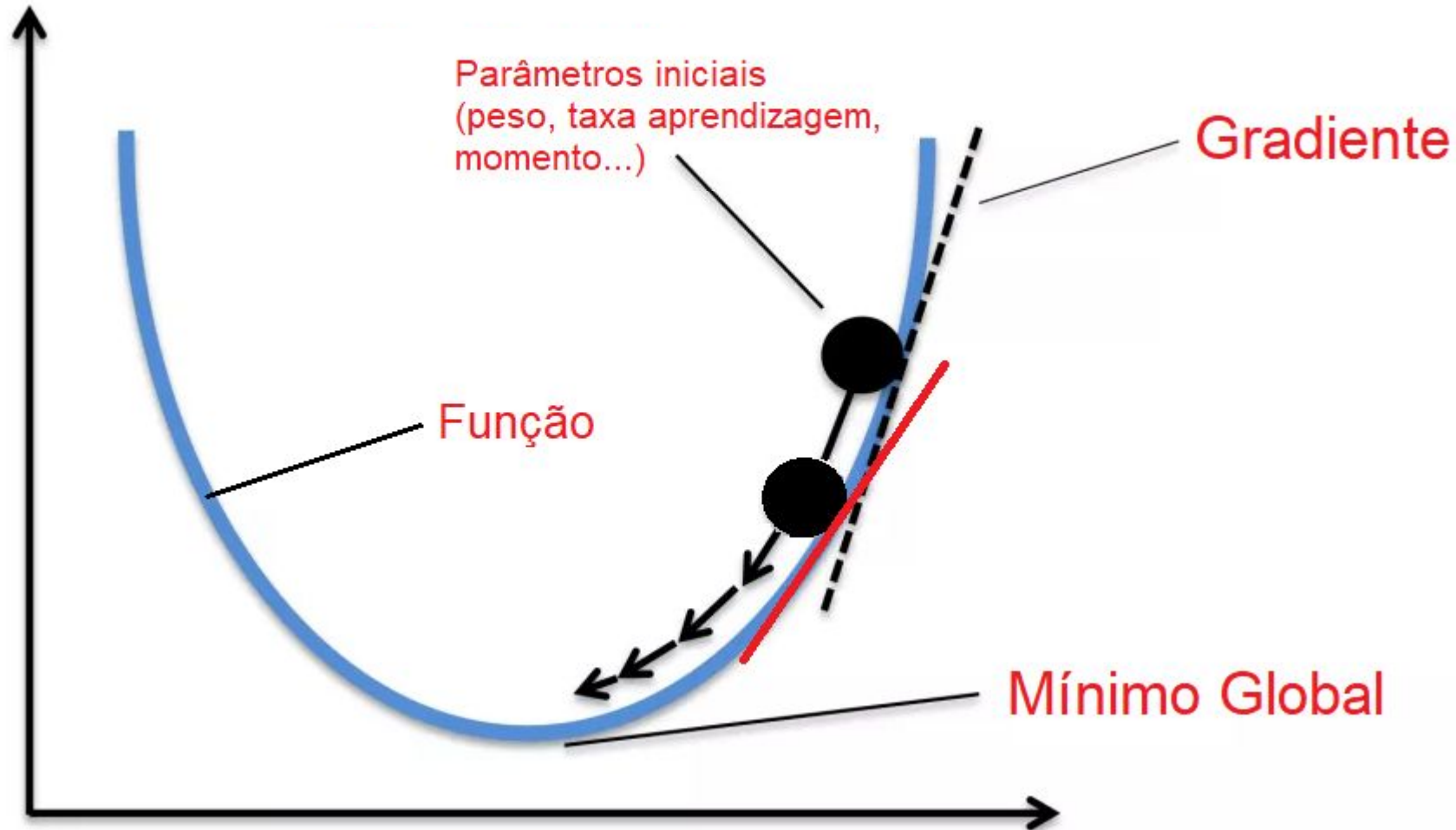


$$\text{derivada} = \frac{dy}{dx} = y'$$

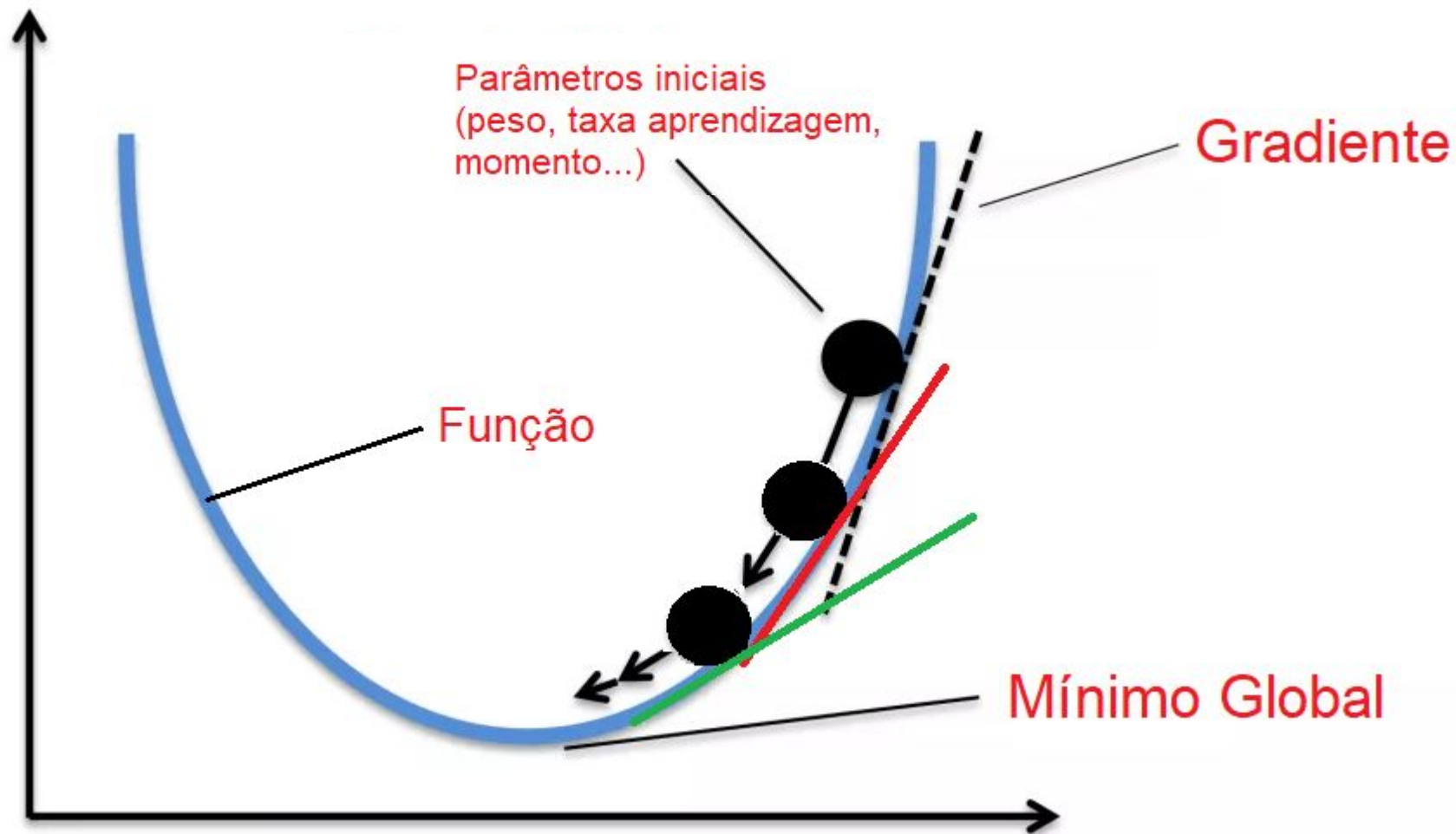
É utilizado para encontrar o mínimo de uma função de erro (função de perda – loss function ou função de custo - cost function) quando aplicado a algoritmos de aprendizagem de máquina



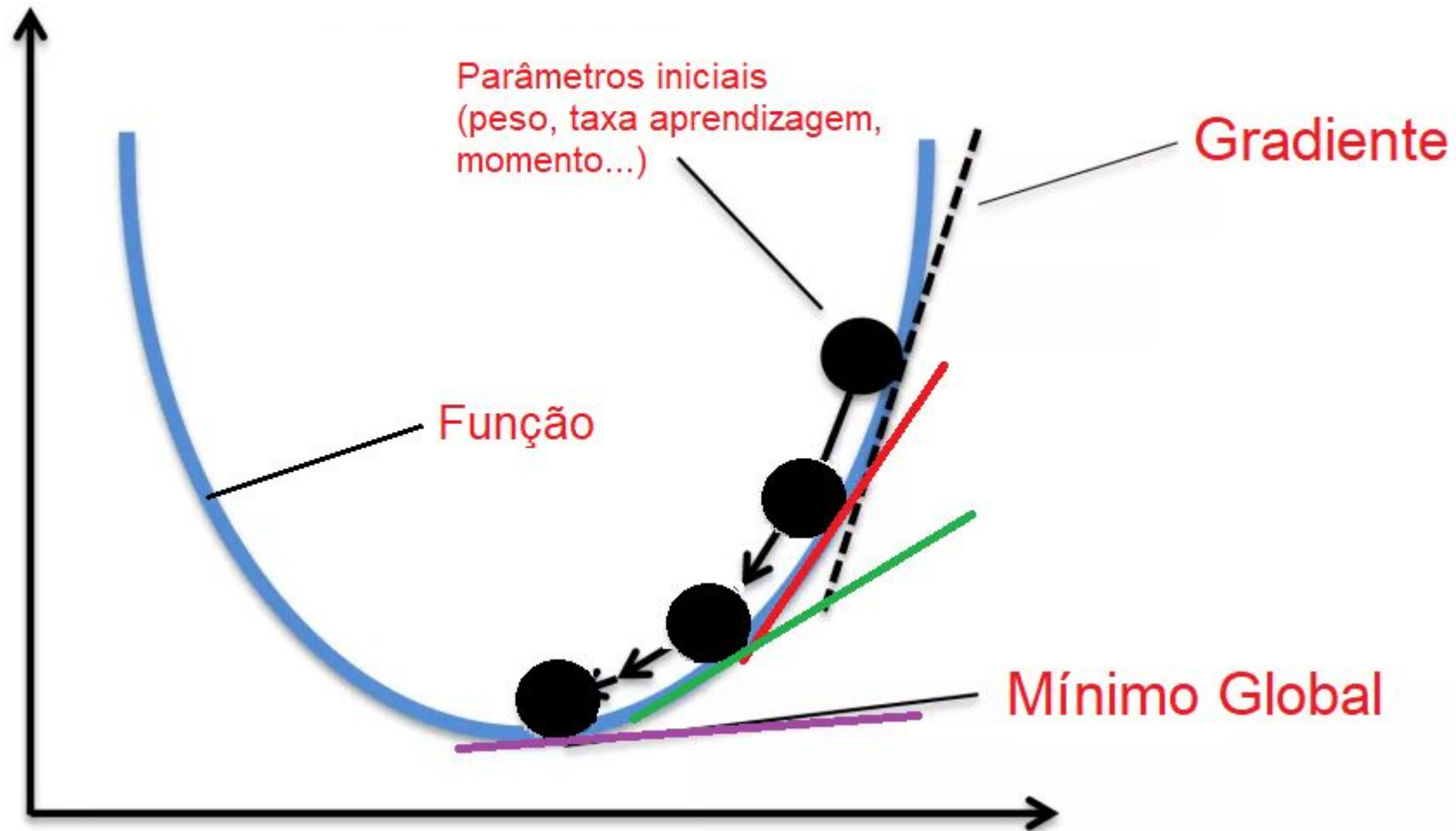
É utilizado para encontrar o mínimo de uma função de erro (função de perda – loss function ou função de custo - cost function) quando aplicado a algoritmos de aprendizagem de máquina



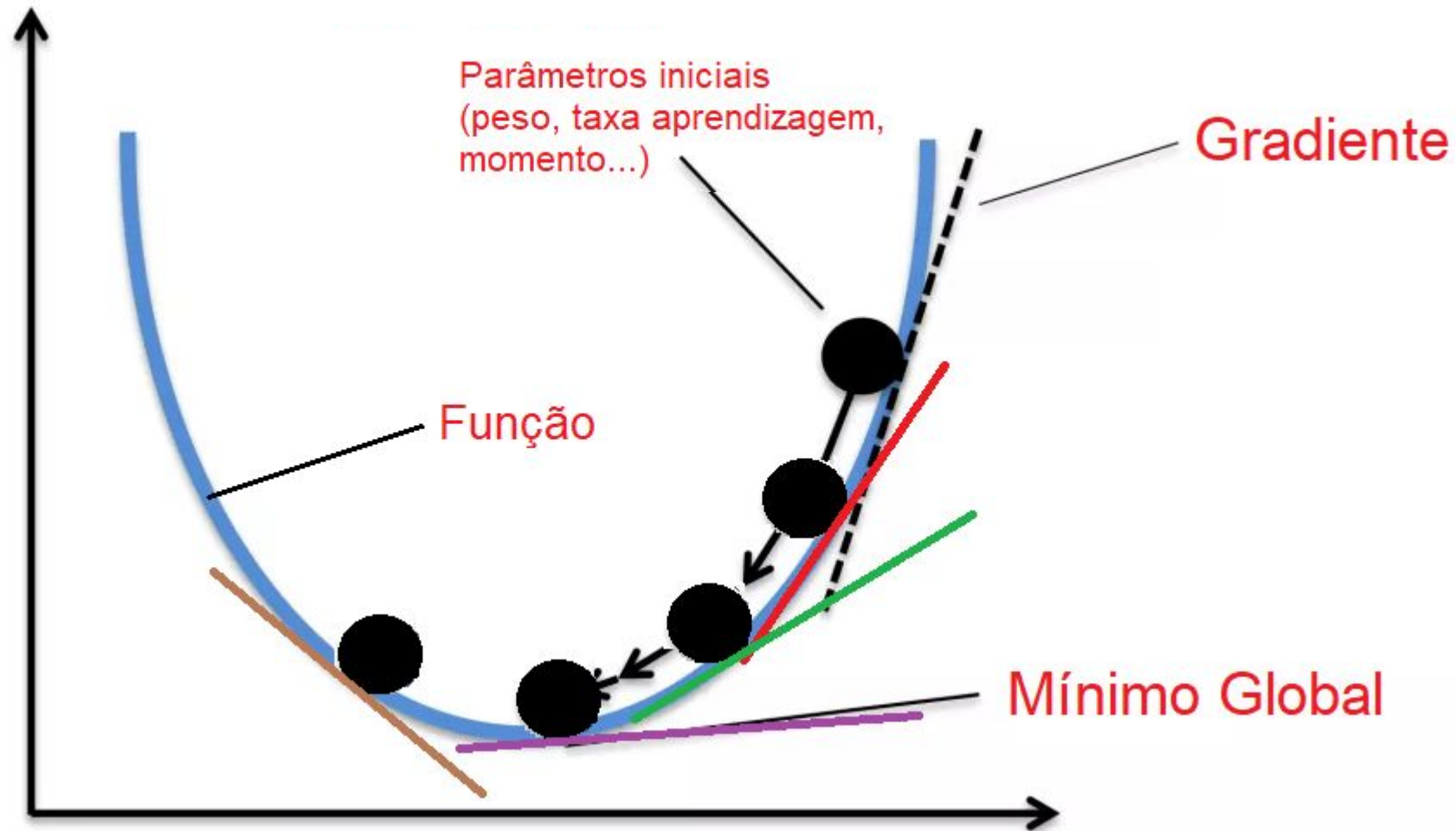
É utilizado para encontrar o mínimo de uma função de erro (função de perda – loss function ou função de custo - cost function) quando aplicado a algoritmos de aprendizagem de máquina



É utilizado para encontrar o mínimo de uma função de erro (função de perda – loss function ou função de custo - cost function) quando aplicado a algoritmos de aprendizagem de máquina



É utilizado para encontrar o mínimo de uma função de erro (função de perda – loss function ou função de custo - cost function) quando aplicado a algoritmos de aprendizagem de máquina



Evolução do algoritmo de descida do gradiente

SGD (Stochastic Gradient Descent): Descida de Gradiente Estocástica. Aumenta o número de atualizações nas interações (todos os dados atualizam pesos). Evita erro no mínimo local, mas tem excesso de atualizações.

SGD Mini-batch: Descida de Gradiente estocástica com mini lotes. Esse é o algoritmo principal da descida do gradiente. Diminui o número de atualizações e aumenta a velocidade de processamento.

Momento (Momentum): Técnica para aumentar a velocidade do algoritmo de descida do gradiente, reduzir instabilidades e evitar mínimos locais. Seu valor varia de 0 (não utilização) a 1. O valor recomendado para o termo momentum é 0.3.

