

**O otimizador Adam (Adaptive Moment Estimation) é um algoritmo de otimização popular usado em redes neurais para ajustar os pesos durante o treinamento. Ele combina os benefícios do otimizador Gradiente Descendente Estocástico (SGD) com adaptações baseadas em momentos.**

*O algoritmo Adam mantém uma taxa de aprendizado adaptativa para cada peso, ajustando-a de acordo com as estimativas de momento do gradiente de primeira ordem (média móvel do gradiente) e do gradiente de segunda ordem (média móvel do quadrado do gradiente). Isso permite que o algoritmo se adapte automaticamente à taxa de aprendizado ideal para cada peso e, assim, acelere o treinamento.*

**Aqui estão os principais componentes e etapas do algoritmo Adam:**

- 1. \*\*Momentum\*\*:** O algoritmo Adam utiliza o conceito de momentum para acelerar o treinamento e ajudar a superar os mínimos locais. Ele mantém uma média móvel exponencial do gradiente de primeira ordem (média móvel do momento) para direcionar o algoritmo de otimização na direção certa.
- 2. \*\*Adaptação do learning rate\*\*:** O Adam ajusta a taxa de aprendizado de forma adaptativa para cada peso com base nas estimativas de momentos do gradiente. Isso ajuda a garantir que a taxa de aprendizado seja apropriada para cada peso individualmente.
- 3. \*\*Bias correction\*\*:** No início do treinamento, as estimativas de momentos são inicializadas como zero, o que pode levar a uma taxa de aprendizado inicial baixa. Para compensar isso, o Adam realiza uma etapa de correção de viés, chamada de correção de viés do momento, para ajustar as estimativas de momento iniciais.
- 4. \*\*Atualização dos pesos\*\*:** O Adam atualiza os pesos da rede neural usando os gradientes ponderados pela taxa de aprendizado adaptativa. A taxa de aprendizado adaptativa é calculada como uma fração do momento do gradiente de primeira ordem e da raiz quadrada do momento do gradiente de segunda ordem.

*Em resumo, o otimizador Adam combina as vantagens do SGD com adaptações baseadas em momentos e uma taxa de aprendizado adaptativa. Essas características ajudam a acelerar o treinamento e melhorar a convergência das redes neurais.*