

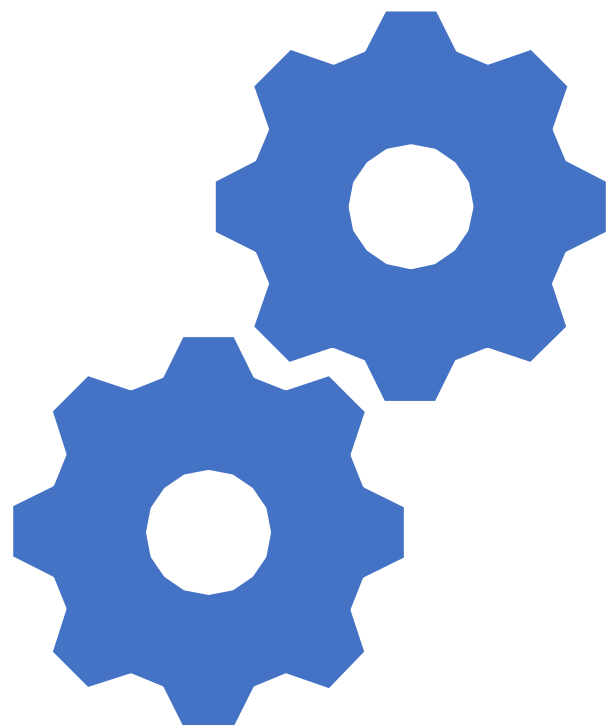


PECE Programa de
Educação Continuada

Escola Politécnica da USP

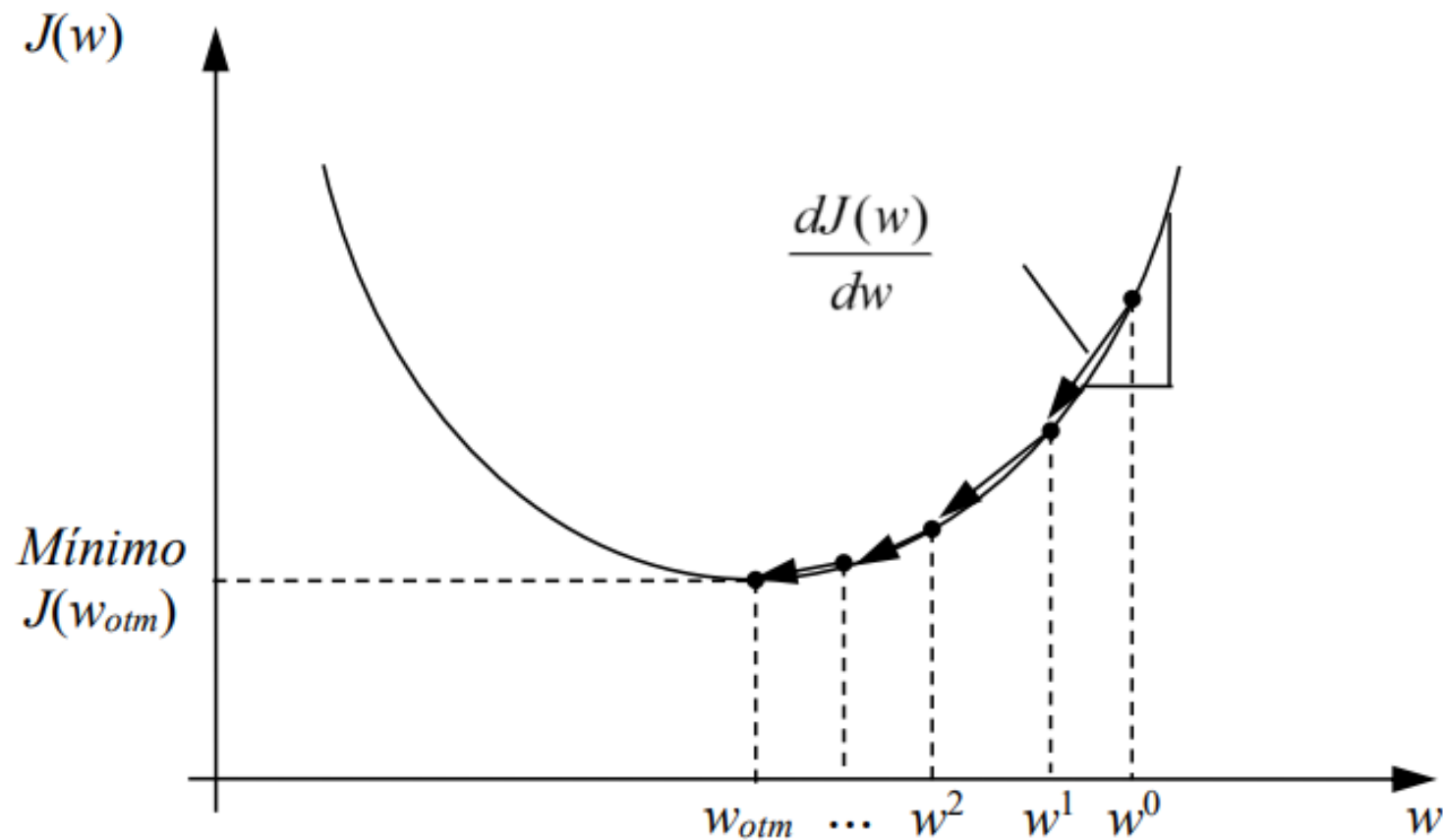
Introdução a Redes Neurais

Marlon Sproesser Mathias

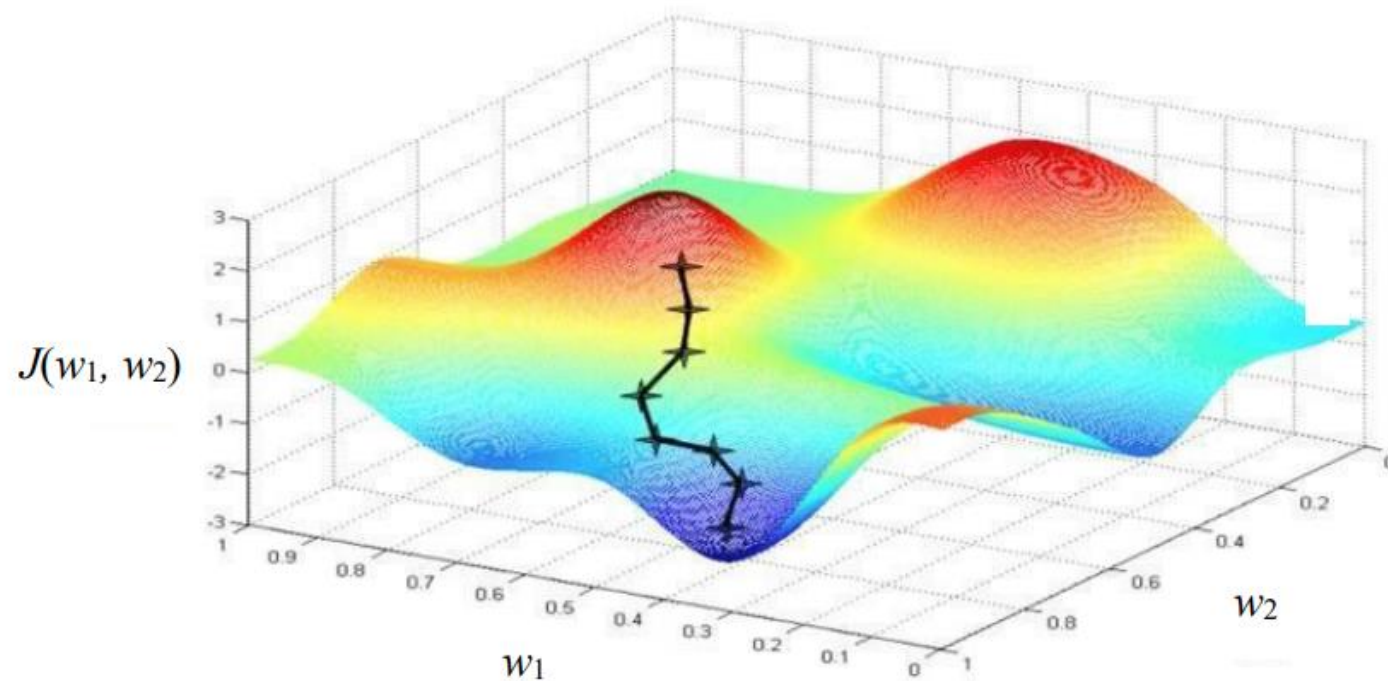


Métodos de otimização

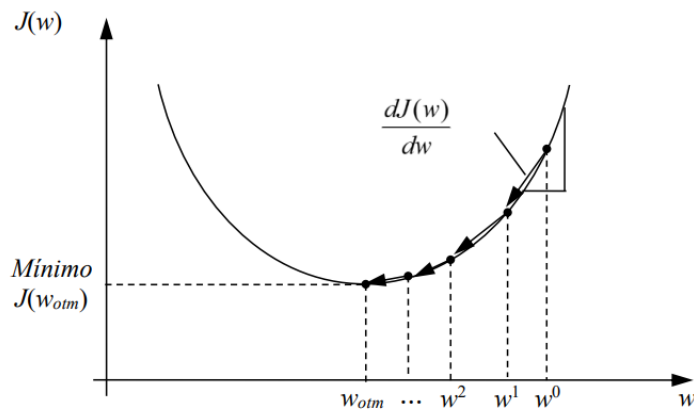
Otimização por gradiente descendente



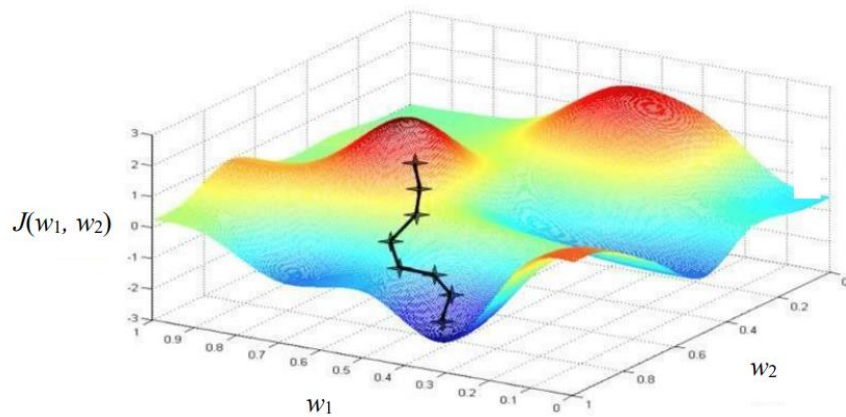
Otimização por gradiente descendente



<https://github.com/RNogales94/ISI-Tensorflow>

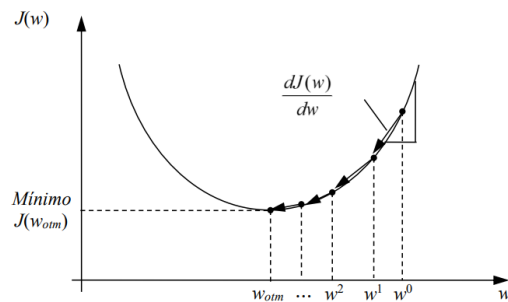


Otimização por gradiente descendente

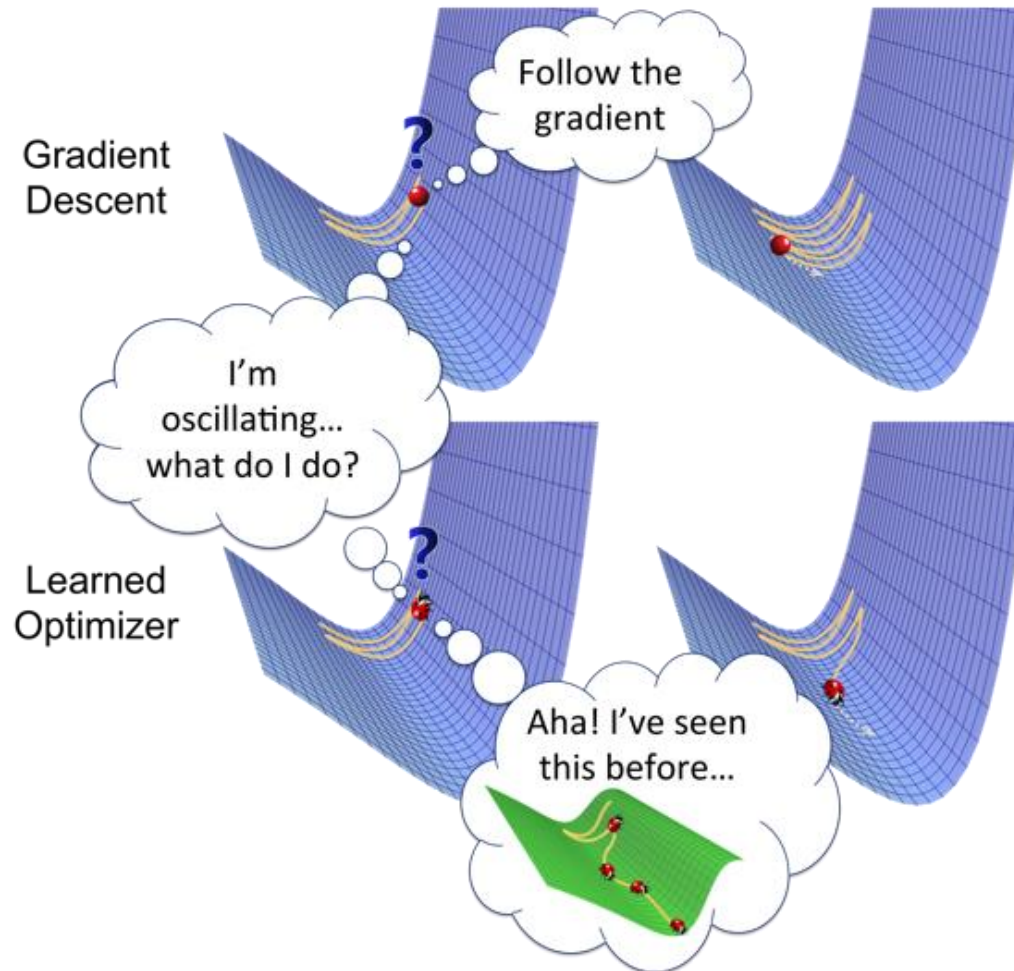
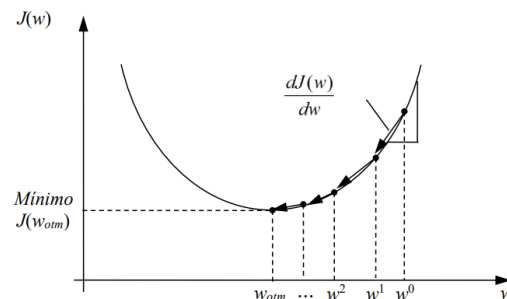
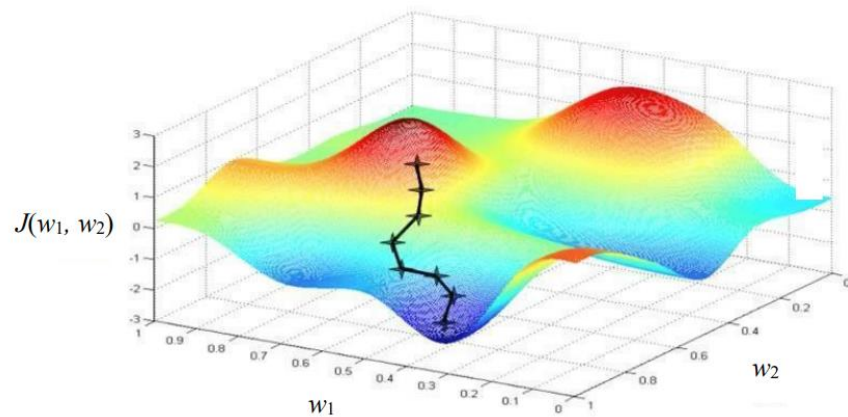


Possíveis dificuldades:

- Mínimos locais
- Diferentes magnitudes dos gradientes



Otimização por gradiente descendente



Gradiente descendente

- Simples
- Fácil de implementar
- Pode parar em mínimos locais
- Percorre todo o dataset antes de atualizar os pesos

$$\theta = \theta - \alpha \cdot \nabla J(\theta)$$

Gradiente descendente estocástico

- Simples
- Fácil de implementar
- Pode parar em mínimos locais
- Atualiza a cada avaliação

$$\theta = \theta - \alpha \cdot \nabla J(\theta; x(i); y(i))$$

onde $\{x(i), y(i)\}$ são os exemplos de treinamento

Gradiente descendente por mini-batch

- Simples
- Fácil de implementar
- Pode parar em mínimos locais
- Atualiza após cada batch

$$\theta = \theta - \alpha \cdot \nabla J(\theta; B(i))$$

onde $\{B(i)\}$ são os lotes

Momento

- Reduz oscilações
- Evita mínimos locais
- Adiciona o hiper-parâmetro de momento

$$\mathbf{V}(t) = \gamma \mathbf{V}(t-1) + \alpha \cdot \nabla J(\theta)$$

$$\theta = \theta - \mathbf{V}(t)$$

Momento acelerado de Nesterov

- Evita “passar reto” pelo mínimo
- Olha adiante para atualização da velocidade

$$V(t) = \gamma V(t-1) + \alpha \cdot \nabla J(\theta - \gamma V(t-1))$$

$$\theta = \theta - V(t)$$

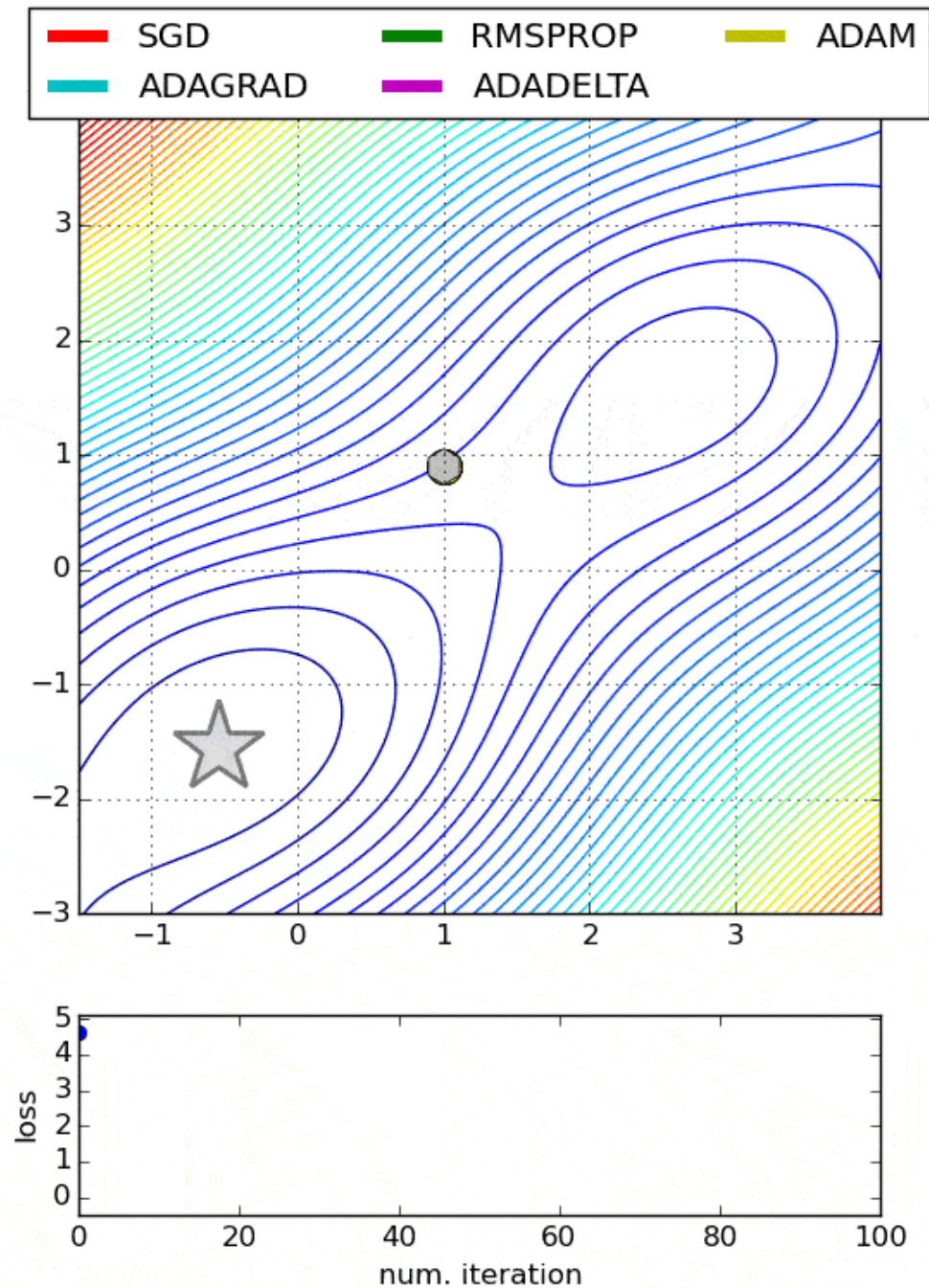


Adagrad, AdaDelta, Adam

- Taxa de aprendizado adaptável
- Cada parâmetro tem sua taxa
- Mais caros e complexos

Comparação

<https://towardsdatascience.com/optimizers-for-training-neural-network-59450d71caf6>



Gradientes que desaparecem



- “Vanishing gradients”
- Com sigmoides e tanh, ocorre em valores de módulo alto
- Com ReLU, ocorre com valores muito negativos

Gradientes que desaparecem

O que fazer?

- Camada de renormalização
- Conexões residuais
- Leaky ReLU