



Impacto dos diferentes parâmetros:

Camadas: É um fator crítico para a solução pois enquanto resolvia o exercício foi fácil perceber que redes mais rasas não tem a capacidade de resolver o dataset spiral que demanda mais poder. Redes com 3 camadas podem resolver, mas demoram mais tempo. Já redes com 4 camadas resolvem o dataset spiral com velocidade.

Neuronios em cada camada: O dataset spiral tem bastante features para serem aprendidas, portanto um número razoável de camadas entre 5 e 10 é necessário. 4 neurônios parecem ser poucas, e partir de 5 o resultado melhorou. Mais de 10 neurônios parecem ser excessivos e podem levar a overfitting.

Ativação:

Função Sigmoide falha pois os pesos se aproxima de zero rapidamente.

Função ReLU: Funciona muito bem. É simples de calcular e muito mais rápida do que a função sigmoide.

Learning Rate: O intervalo ótimo para a learning rate ótima é entre 0.03 e 0.1, pois permite uma convergência rápida.

Regularização: L2 é um fator de regularização bom o suficiente para esse problema porque não forçou os pesos a atualizarem muito rapidamente nem devagar demais. Evitando overfitting ou underfitting.