



Utilizei somente as entradas  $x_1$  e  $x_2$  e fiz ajuste fino dos hiperparâmetros até obter boa generalização. A arquitetura final — **três camadas escondidas (8, 6 e 3 neurônios), ativação ReLU, learning rate 0.03 e regularização L1** — apresentou desempenho consistente, atingindo **Training Loss de 0.020 e Test Loss de 0.065**. Esses valores, bem abaixo do limite solicitado de **0.1**, demonstram que a rede conseguiu aprender o padrão complexo da espiral de forma eficaz e com boa capacidade de generalização.

Durante o processo, **testei diversas combinações de camadas, neurônios, taxas de aprendizado e tipos de regularização**, observando como algumas arquiteturas não convergiam adequadamente ou sofriam com o overfitting até chegar à configuração mais estável. A partir dessas experimentações, ficou claro que o aprendizado de padrões complexos como o dataset Spiral depende diretamente da combinação entre profundidade da rede, quantidade adequada de neurônios, ativação não linear, taxa de aprendizado e regularização. A arquitetura 8/6/3 com learning rate 0.03 apresentou melhor estabilidade e generalização, atingindo novamente Test Loss = 0.065. Este exercício mostrou, na prática, como cada hiperparâmetro influencia o comportamento da rede e como ajustes cuidadosos tornam o modelo capaz de resolver problemas altamente não lineares.