



1. Com uma camada oculta de 3 neurônios já foi possível resolver o problema
  - O melhor resultado foi uma perda no conjunto de teste de 0,009.
  - A função de ativação ReLu foi a que obteve melhor resultado.
  - Também foi possível resolver o problema com Tangente Hiperbólica e Sigmoides como funções de ativação, mas elas precisaram de muito mais épocas para isso.
  - Não foi possível resolver o problema com uma função de ativação linear.
2. Ao adicionar mais neurônios (e.g. 4 ou 6), mesmo com uma única camada oculta e a ReLu como função de ativação, os resultados foram ainda melhores, uma vez que uma perda de 0,006 foi obtida com pouco mais de 100 épocas iteradas.
3. Usar um learning rate de 3 ou 10 trouxe resultados bem instáveis e perda variável (crescente e decrescente), ou seja, não foi possível resolver um problema. Por outro lado, com um learning rate de 1, foi possível rapidamente chegar a uma perda de quase 0 em ambos os conjuntos de dados (treino e teste).
4. Usar um learning rate muito pequeno (e.g. 0,0001) também pode resolver o problema, mas são necessárias milhares de épocas para isso.
5. Adicionar mais de uma camada oculta também resolve o problema, mas adiciona uma complexidade desnecessária ao modelo, uma vez que uma única camada oculta, com 3 ou mais neurônios, já resolve o problema.
6. Usar regularização também ajuda a resolver o problema, mas não é necessária
  - Com uma taxa de regularização mais alta ( $\geq 1$ ), não é possível resolver o problema, independente do número de camadas ocultas e neurônios.