

Atividade Playground

Nome do Aluno: Gustavo Nicolau Jacintho

E-mail: gnj@cesar.school

1ª tentativa

Épocas executadas: 621

Camadas ocultas: 6 (4n, 6n, 9n, 9n, 4n, 4n)

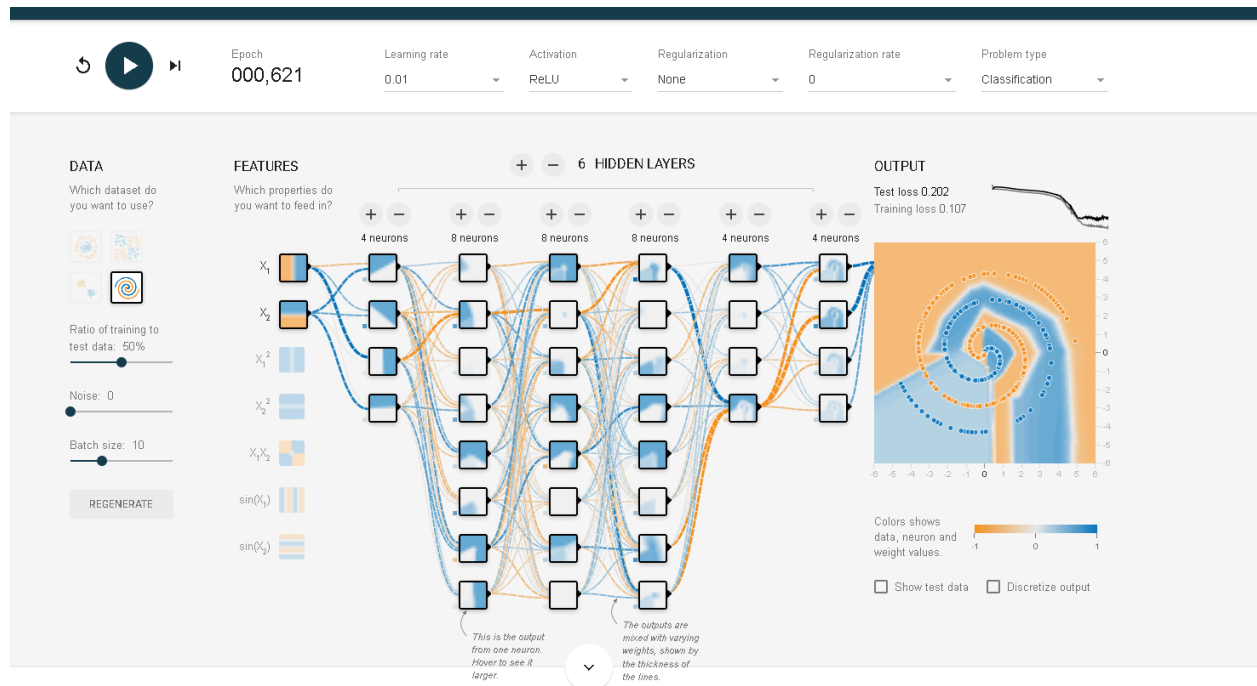
Learning Rate: 0.01

Activation: ReLU

Regularization: None

Regularization rate: 0

Evidência:



Comentários: Comecei com uma rede com muitos recursos, já que se trata de um problema aparentemente de alta complexidade. Notei que ocorreu overfitting e alguns neurônios foram poucos utilizados. Parei a execução na época 621, e o treinamento ainda não tinha sido resolvido. Verifiquei se nenhuma camada configurada se tornou gargalo na rede.

2ª tentativa

Épocas executadas: 1180

Camadas ocultas: 4 (8n, 6n, 6n, 4n)

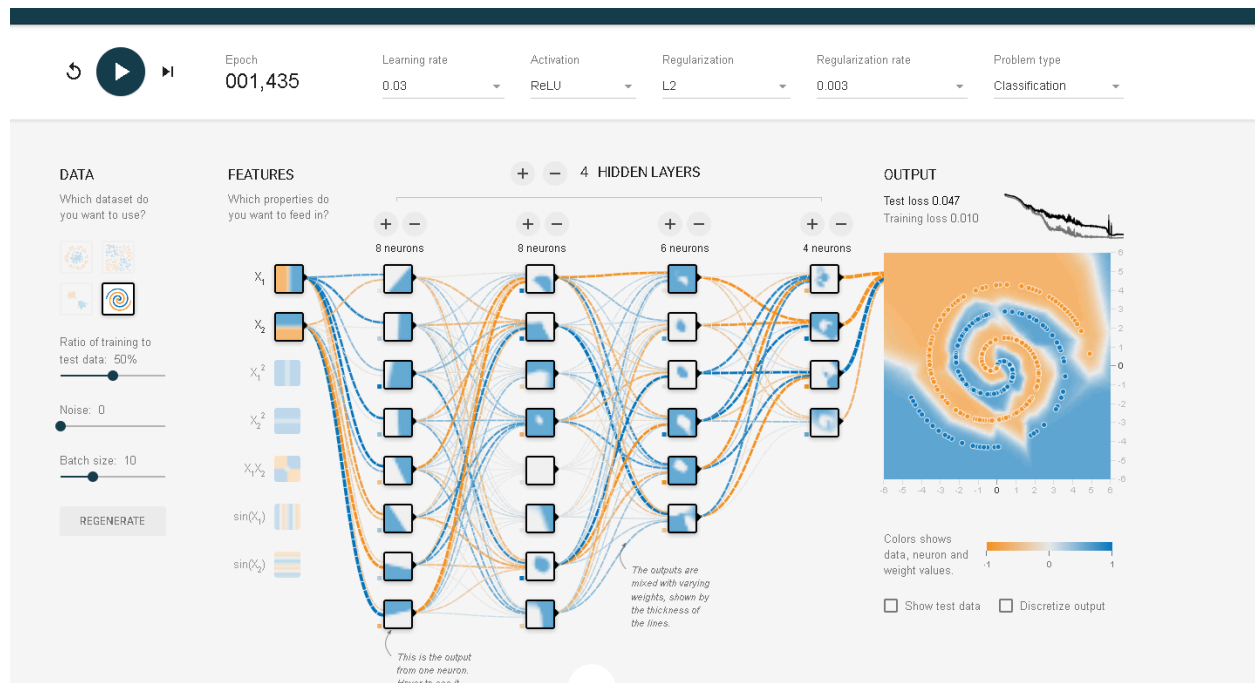
Learning Rate: 0.01

Activation: ReLU

Regularization: L2

Regularization rate: 0.003

Evidência:



Comentários: Ao notar que vários neurônios não foram executados na primeira tentativa, tentei reduzir o número de camadas ocultas e aumentar o número de neurônios na entrada. Também adicionei o tipo de regularização para L2, com um rate de 0.003 para tentar evitar o Overfitting detectado na primeira tentativa. O treinamento foi resolvido após 1400 épocas.

3ª tentativa

Épocas executadas: 2006 épocas

Camadas ocultas: 4 (6n, 5n, 4n)

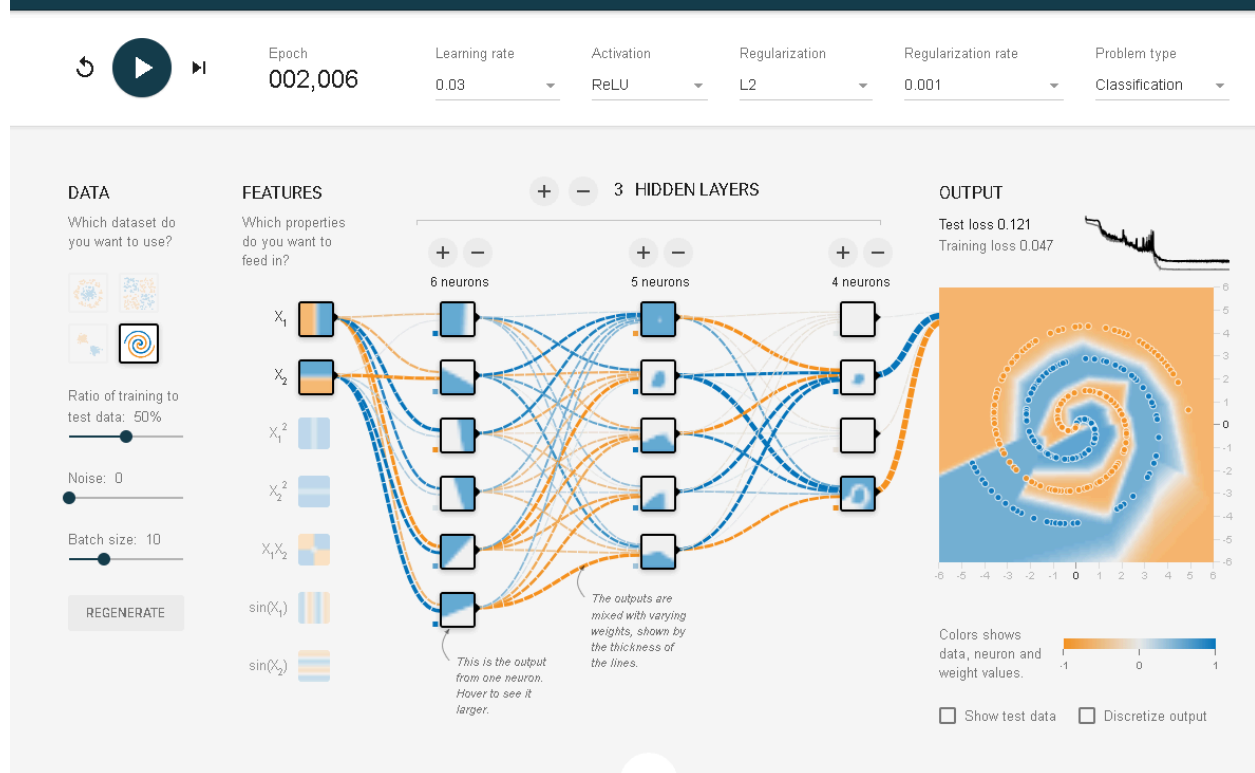
Learning Rate: 0.03

Activation: ReLU

Regularization: L2

Regularization rate: 0.001

Evidência:



Comentários: Novamente tentei otimizar a rede, mas com menos camadas e neurônios, aumentei o learning rate para 0.03. Mesmo após mais de 2000 épocas, o treinamento não convergiu em um resultado, e parece que ficou preso em com overfitting.

Para diminuir o overfitting, aumentei o valor do Regularization rate para 0.003 na próxima tentativa.

4ª tentativa

Épocas executadas: 984 épocas

Camadas ocultas: 4 (6n, 5n, 4n)

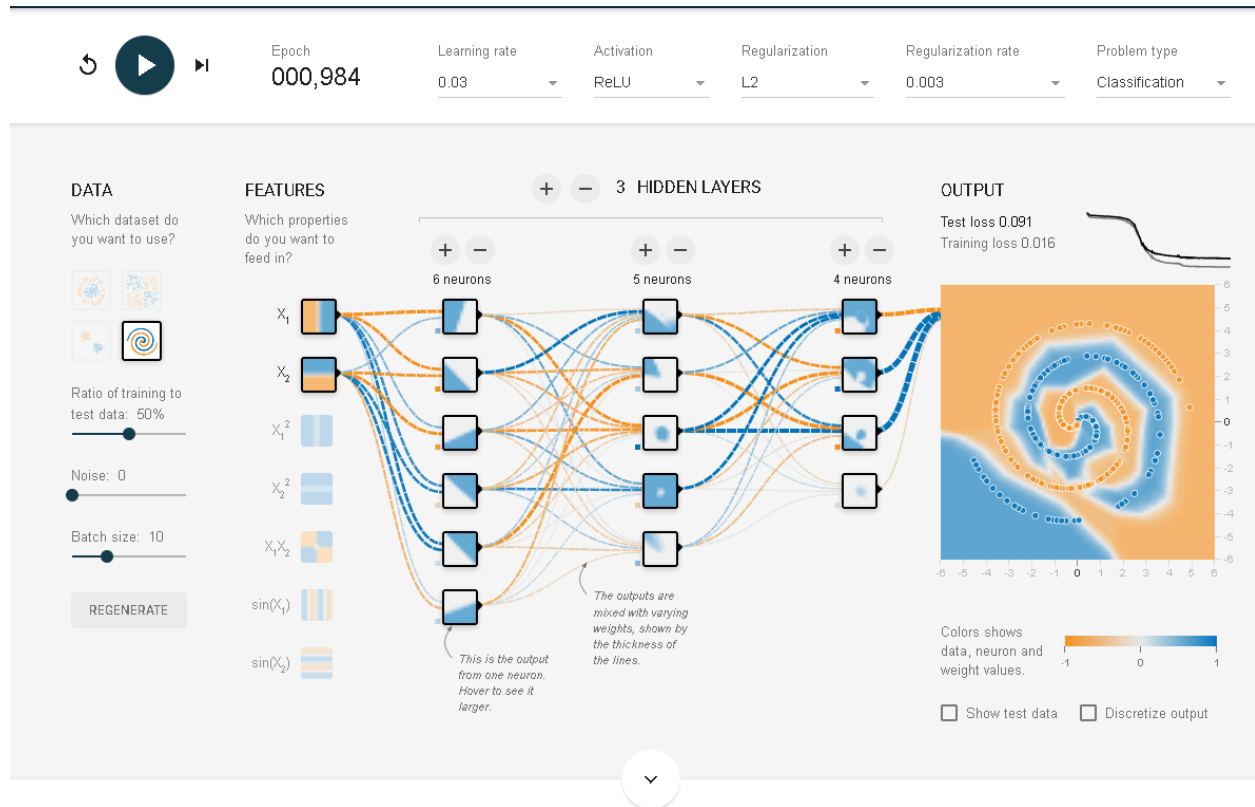
Learning Rate: 0.03

Activation: ReLU

Regularization: L2

Regularization rate: 0.003

Evidência:



Comentários: Aumentando o valor do Regularization rate para 0.003 o problema do overfitting foi resolvido e o treinamento convergiu para um resultado após 980 épocas. Esta rede tornou-se muito mais eficaz que a rede da segunda tentativa. A otimização funcionou.

Tentativas com Tanh

Épocas executadas: 984 épocas

Camadas ocultas: 4 (6n, 5n, 4n)

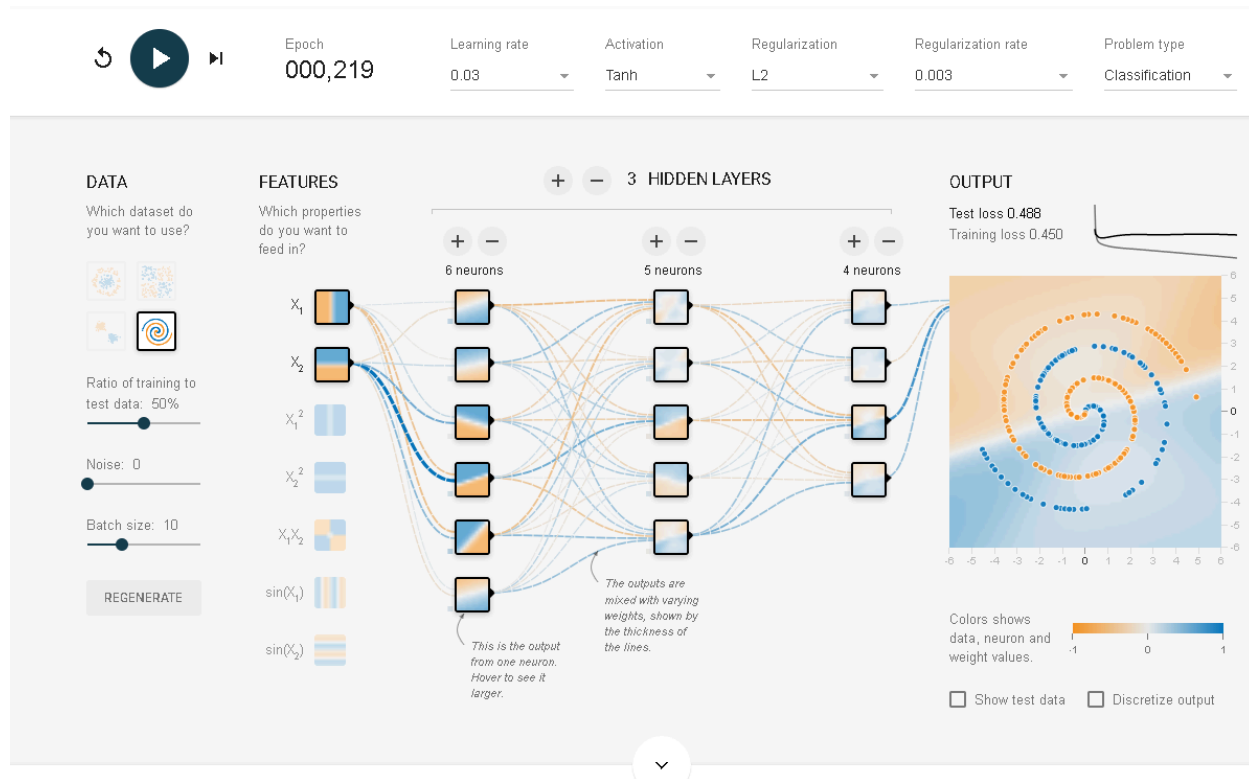
Learning Rate: 0.03

Activation: Tanh

Regularization: L2

Regularization rate: 0.003

Evidência:



Comentários: Tentei chegar em um resultado várias vezes com Tanh, mas não obtive sucesso. Na maioria dos casos, ela tende a ficar presa como a imagem acima. Apenas em uma execução, ela chegou muito próxima de resolver, porém não salvei o print e nem guardei os parâmetros que utilizei.

Enfim, os meus testes demonstraram que a ReLU é mais robusta e resiliente. E pelo que estudei, o modelo Tanh é mais enxuto que o ReLU, porém ele claramente parece depender mais de uma sorte na aleatoriedade inicial do que o modelo ReLU nos testes que realizei. Não que a sorte não ajude nos 2.

Tentativas com Linear e Sigmoid

Comentários: Não realizei tentativas com os 2 modelos, pois pelo que foi aprendido em sala de aula, o problema da espiral não é um problema linear, logo o modelo linear irá falhar e não irá conseguir convergir em um resultado. Já o sigmoid não funciona bem com redes profundas, e entendo que precisamos de várias camadas para resolver o problema da Espiral, já que uma rede com apenas uma camada tem uma capacidade limitada de aprender relações complexas.