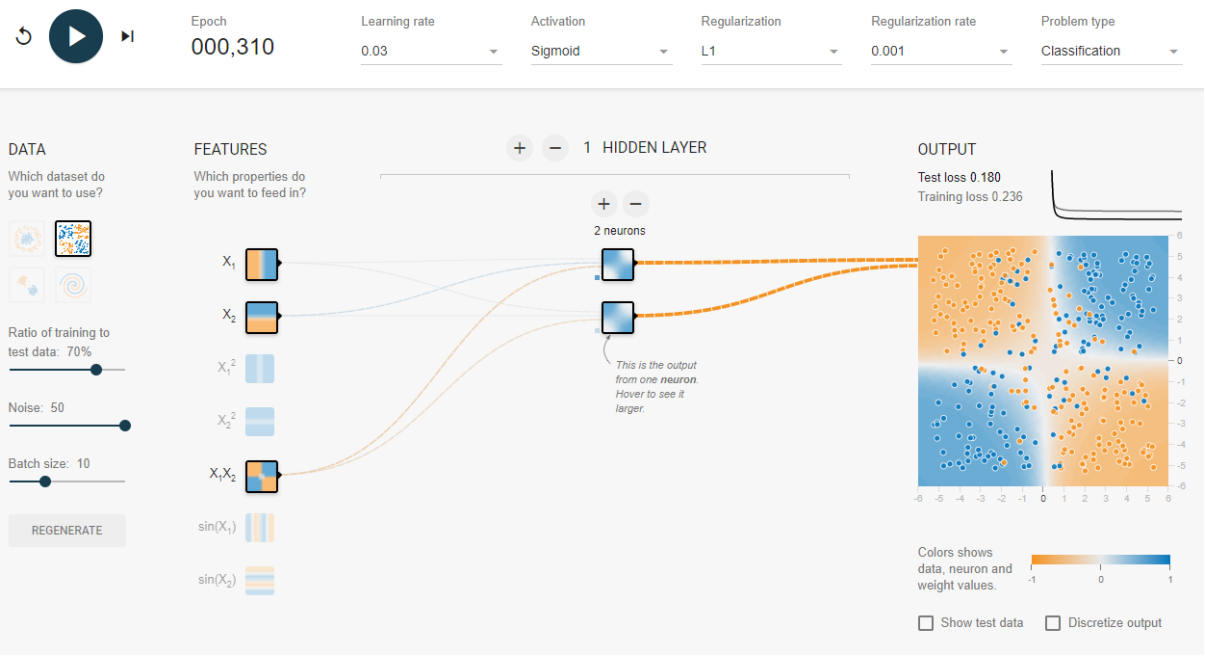

The Network is optimized to recognize the 2 categories of the XOR dataset by implementing X1, X2, and X1X2 features. This allows the network to easily distinguish the 4 relevant corners of the dataset. Since the dataset is so simple, a single hidden layer of 2 neurons is all that is needed. In this case, both neurons learned the same pattern, however an inverse pattern would be a just as accurate prediction. We use the Sigmoid activation function to relatively easily produce a bounded probability of the points. We use slight regularization to counteract the noise in the data.