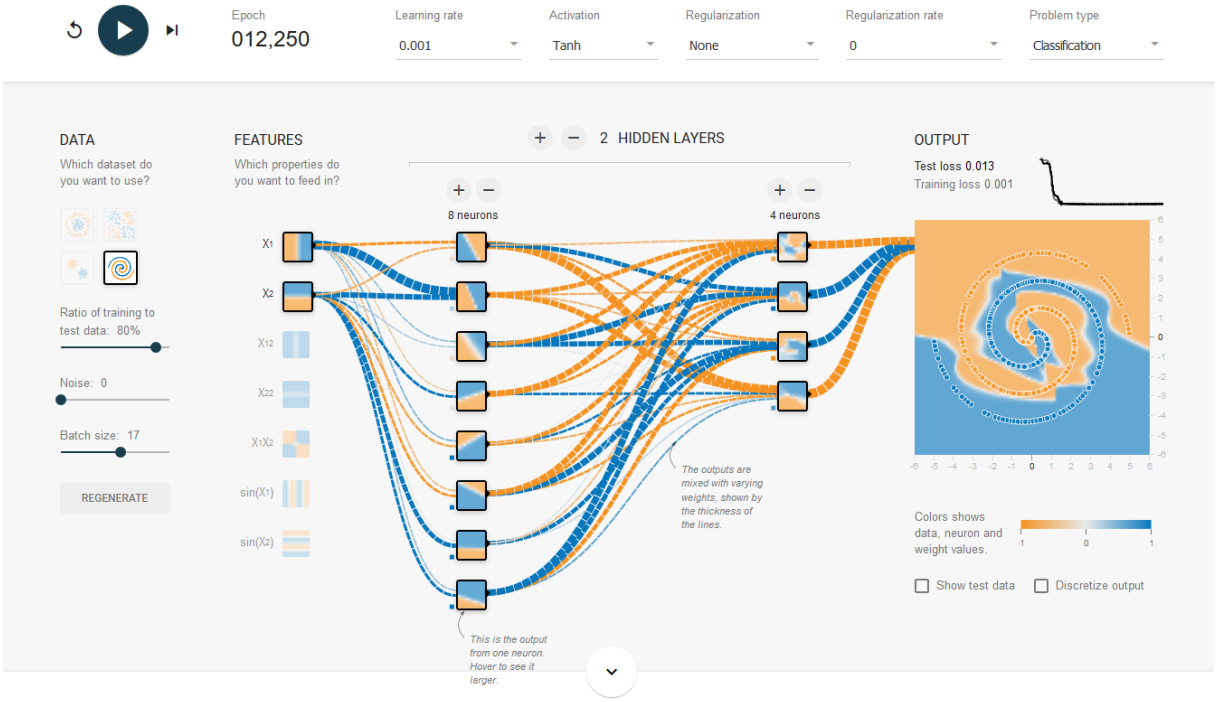
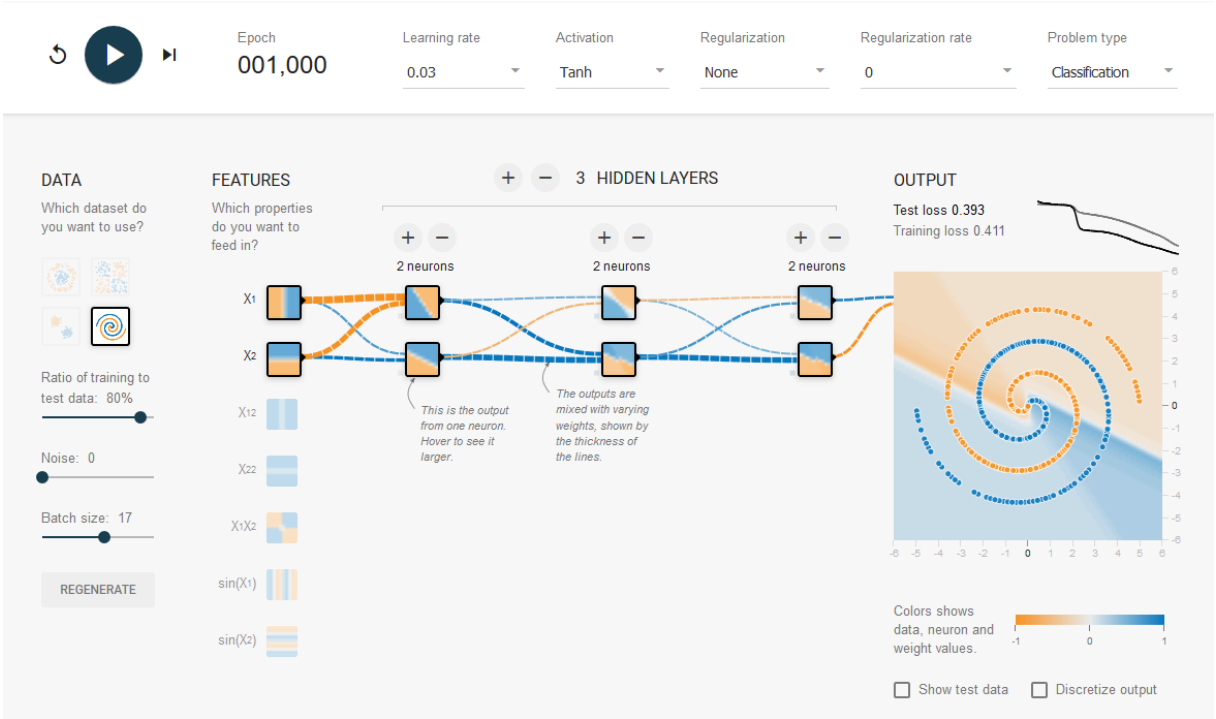
Exercise 1

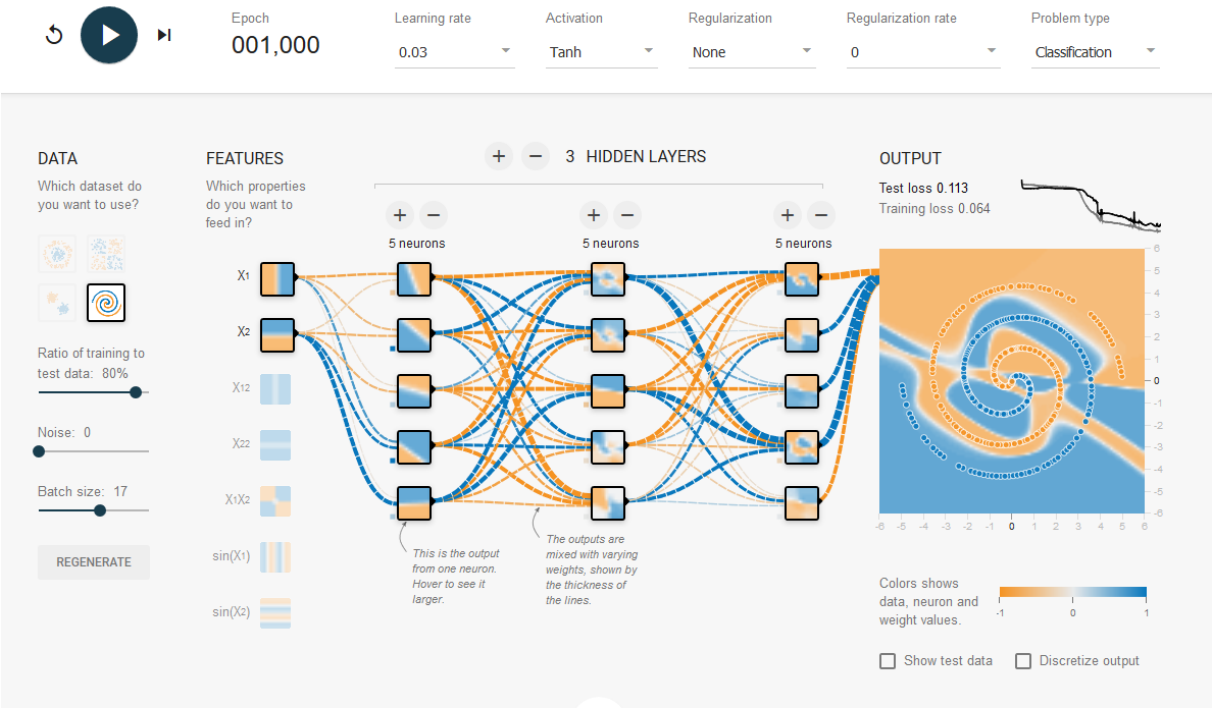
It was possible to fit the dataset very good with the following configuration (changed the learning rate for the last 2000 iterations to 0.001, before it was at 0.03)



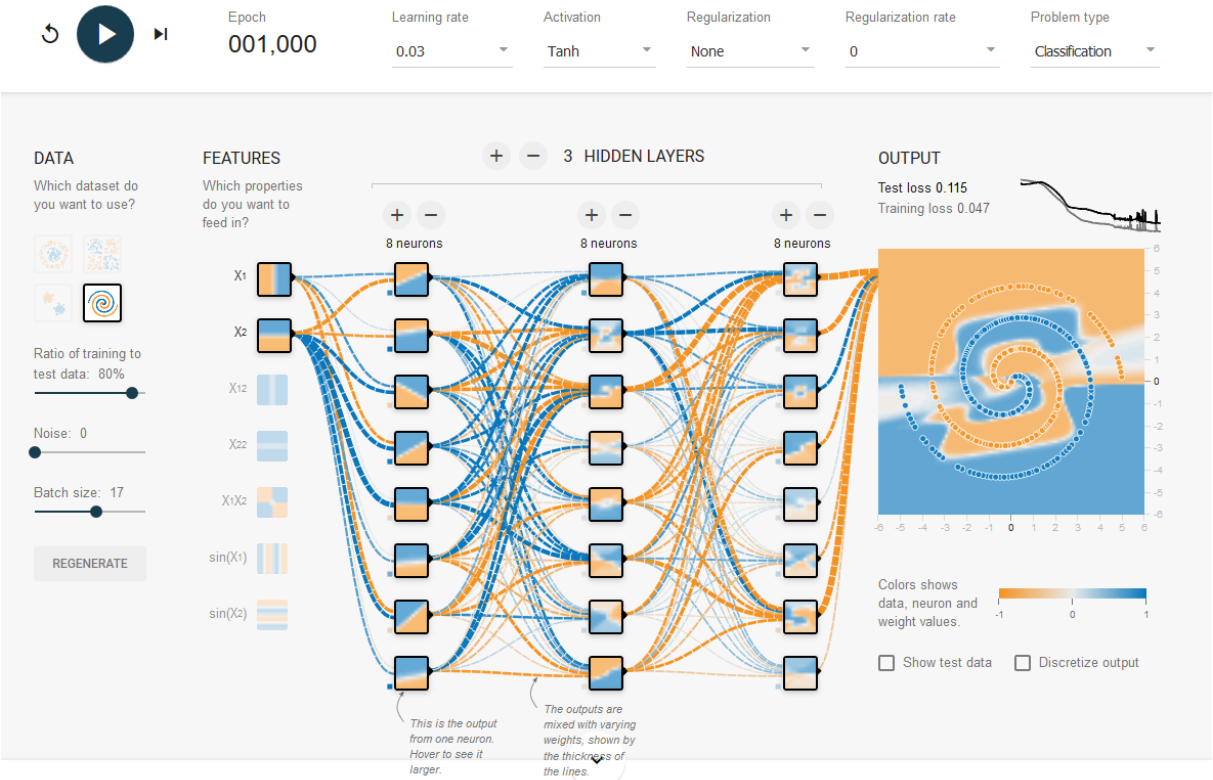
With a number of 3 layers and 2 neurons per layer, the dataset did not fit very well as seen in the following picture.



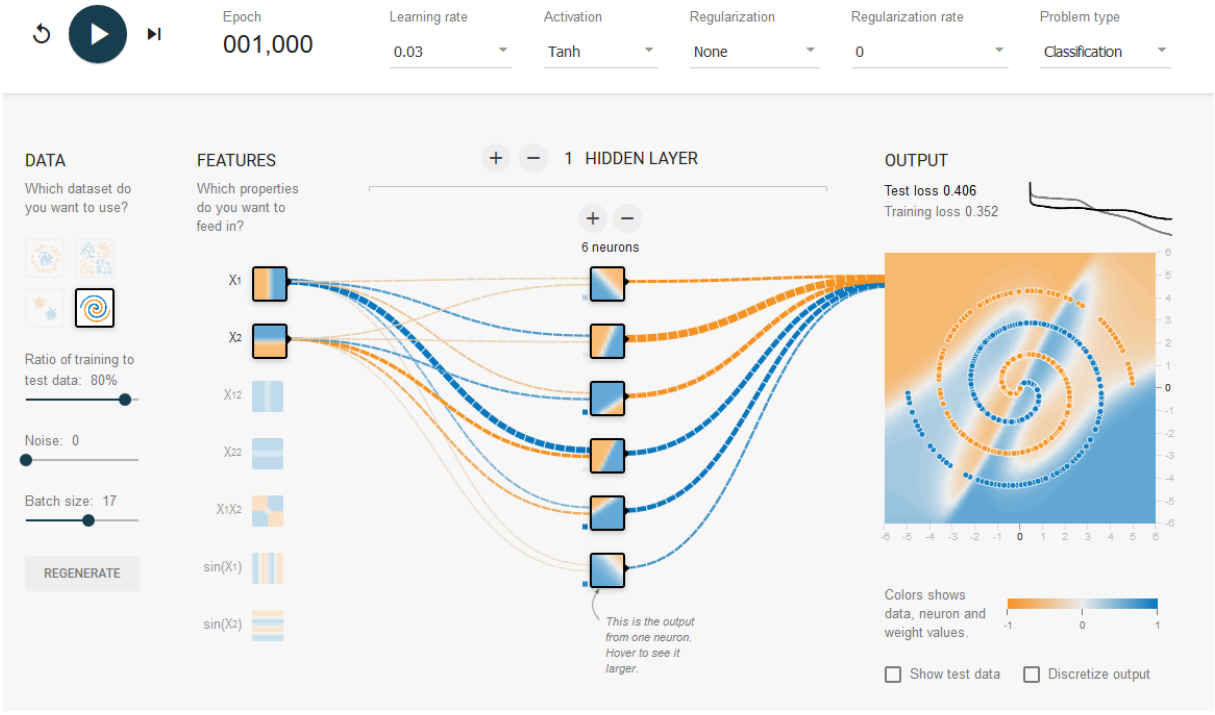
Increasing the number of neurons per layer to 5 improves the neural network extremely



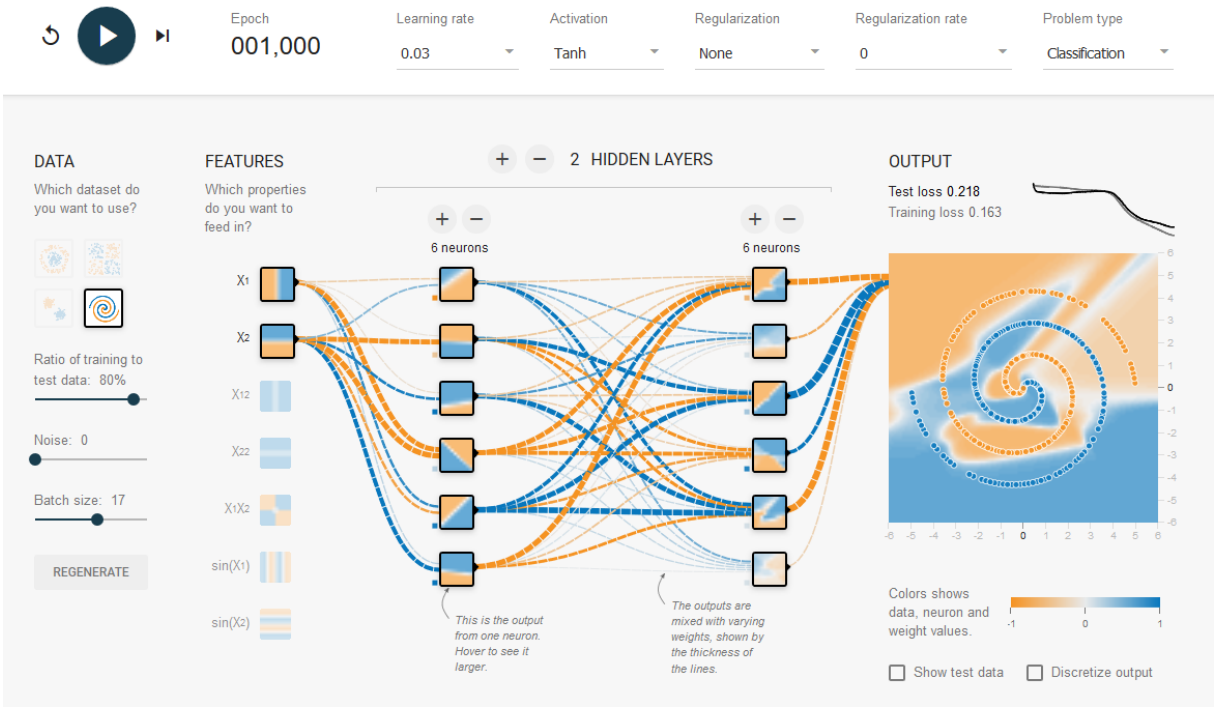
The effect when increasing the number of neurons to 8 per layer makes the result slightly better but has not such a dramatic effect.



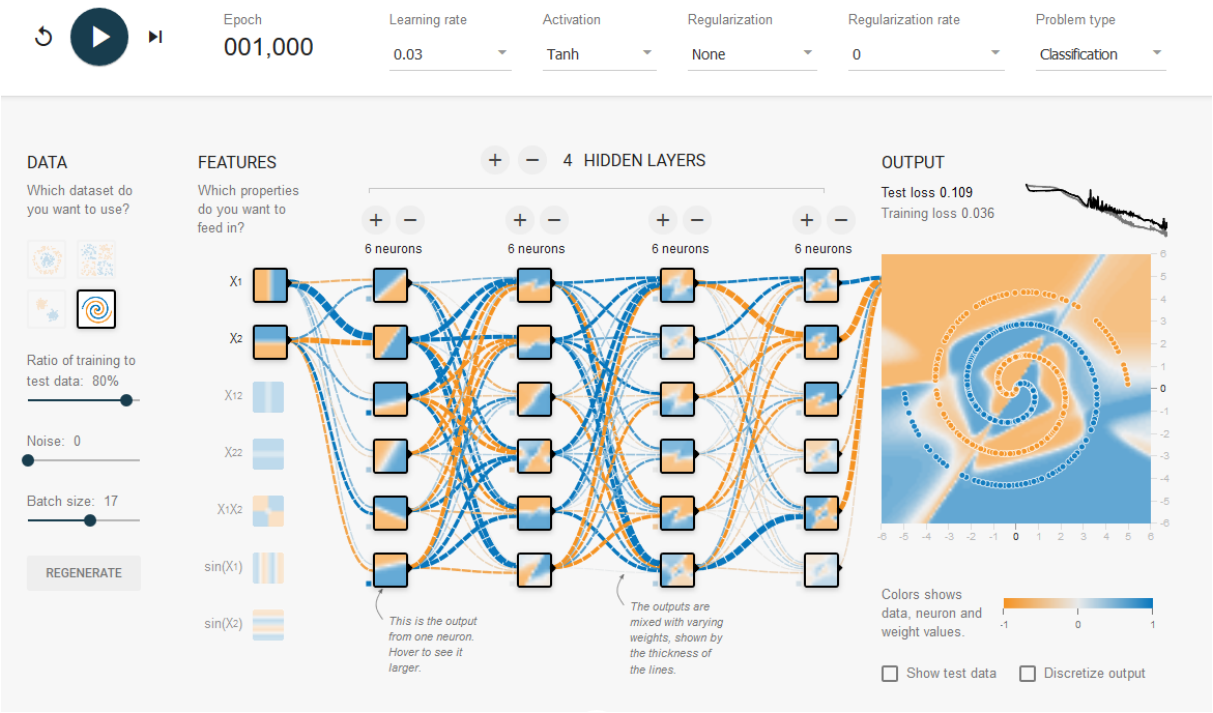
Now this results will be compared to the results of different numbers of layers.
In the first place, the network with 1 layer and 6 neurons was not really able to fit the dataset



But only doubling the number of layers improves the results a lot



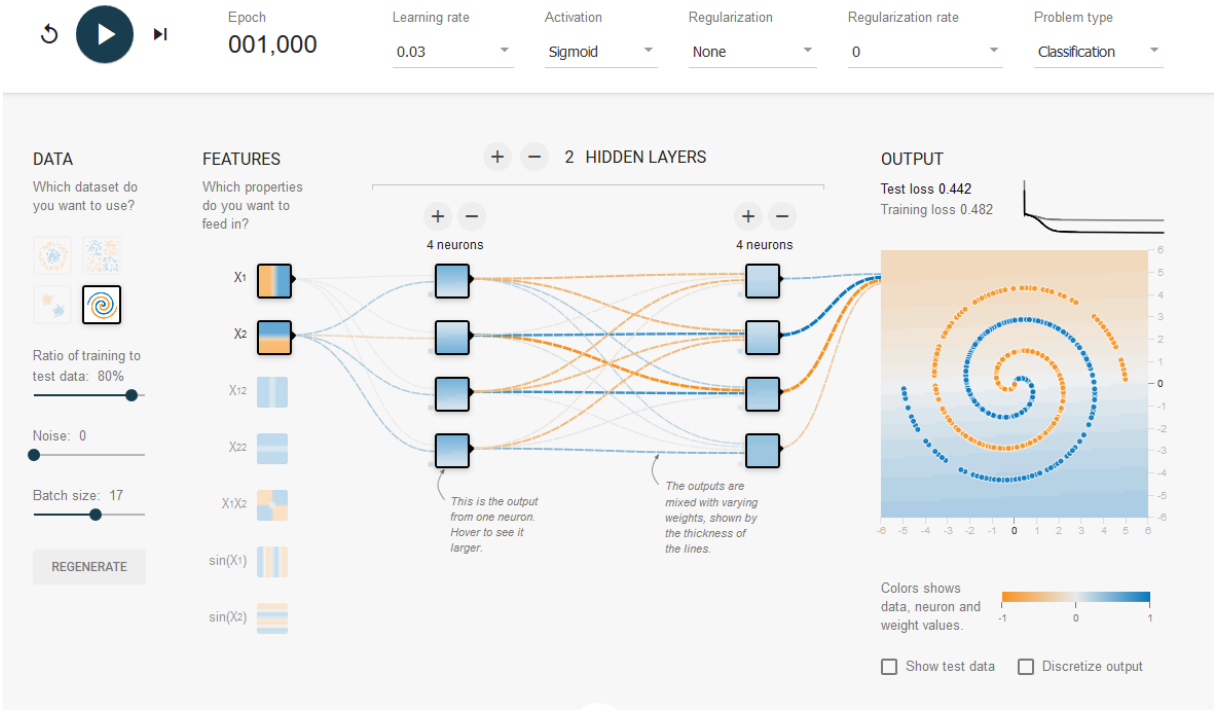
Doubling the number again also improves the Test result by the factor 2

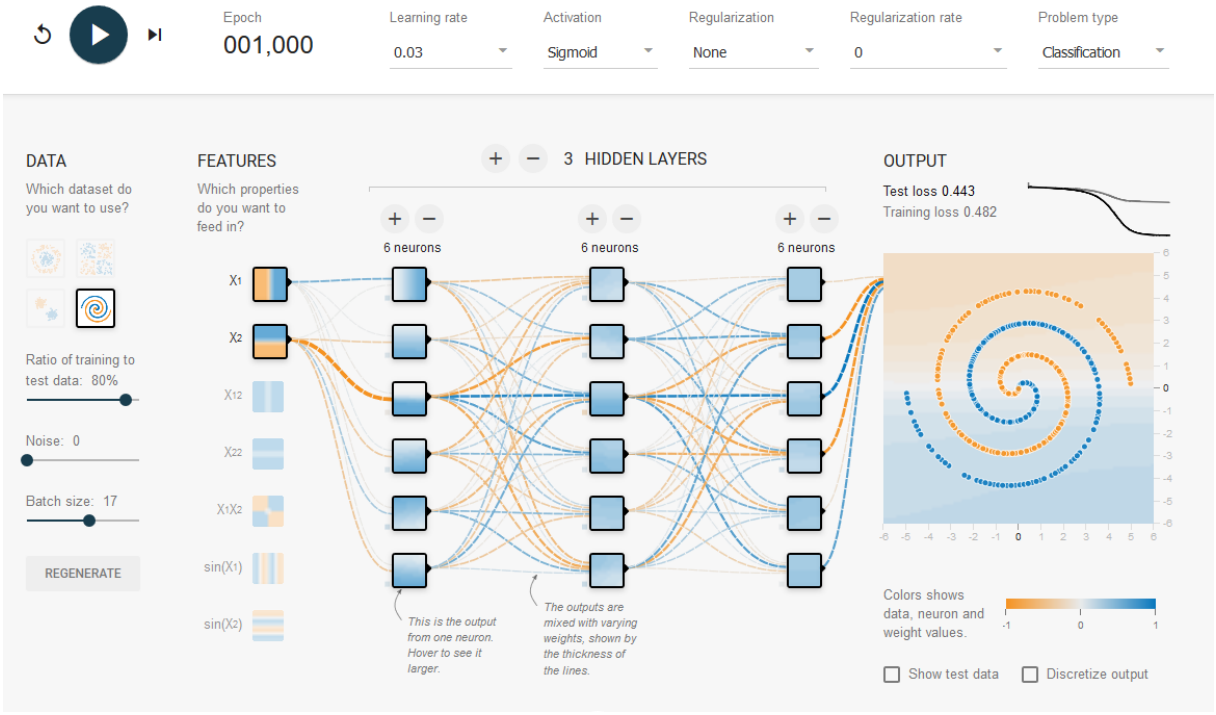


These results give an hint that the number of layers (the deepness) is more important than the number of neurons per layer. It can also be seen that the good results start to jump when reaching a high degree of precision. Reducing the learning rate when reaching low loss values maybe an option to avoid this.

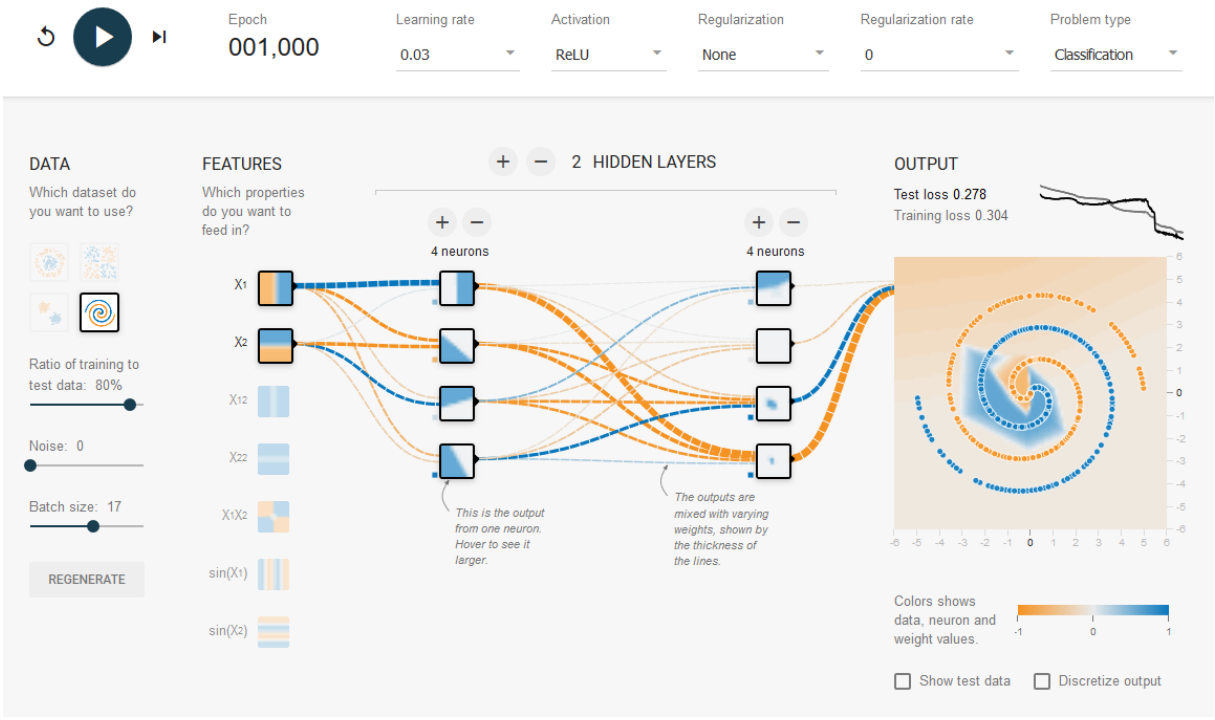
The results above are all achieved with the tanh as the activation function. Now the influence of the activation function on the result will be analyzed. For this, 2 Networks were built to avoid, that the network size has a big influence on the results.

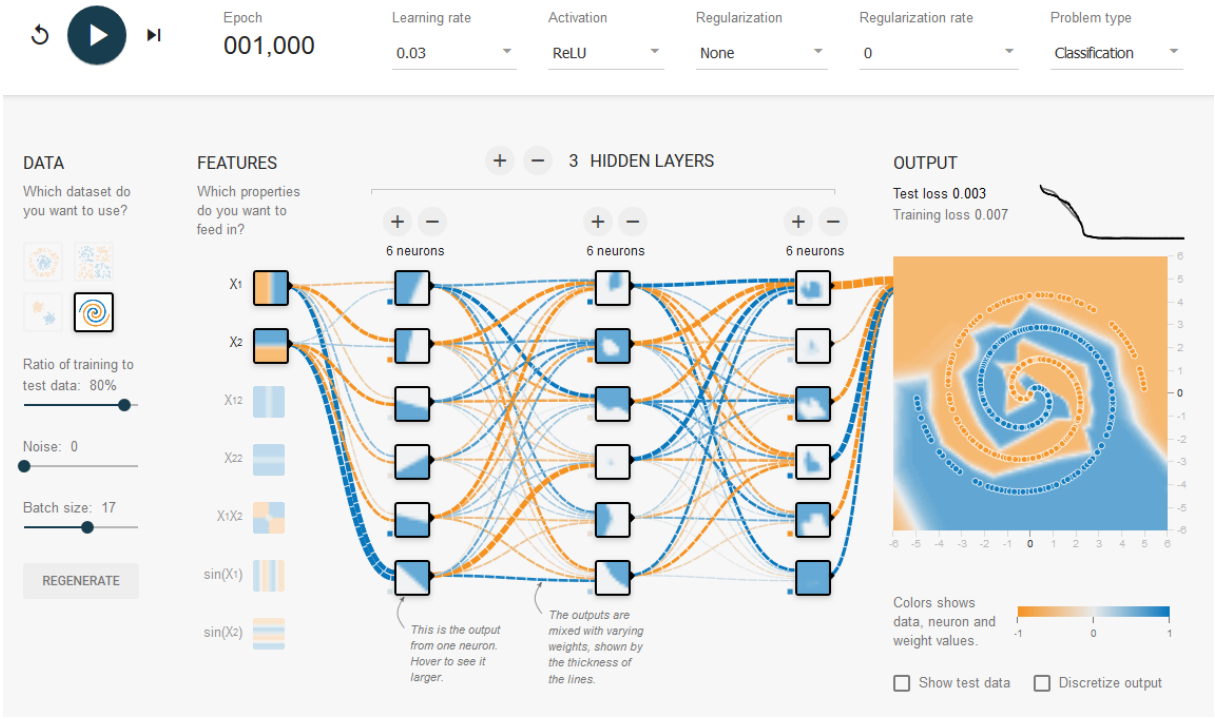
In the following the results of the sigmoid function are presented. With the sigmoid function, the network was not able to fit the dataset at all in both cases





I contrast the ReLu function worked very well on the dataset





There is no hint, why the results are so different. My personal guess would b, that the ReLu function is more sharp and is able to cut different regions on a short range what could be important for this dataset.