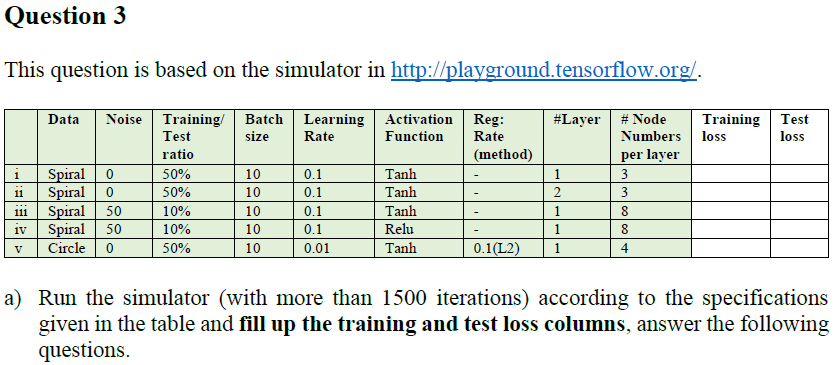


MSE isn’t a hideously bad approach but if you think about how MSE is computed you’ll see that, compared to ACE, MSE gives too much emphasis to the incorrect outputs.

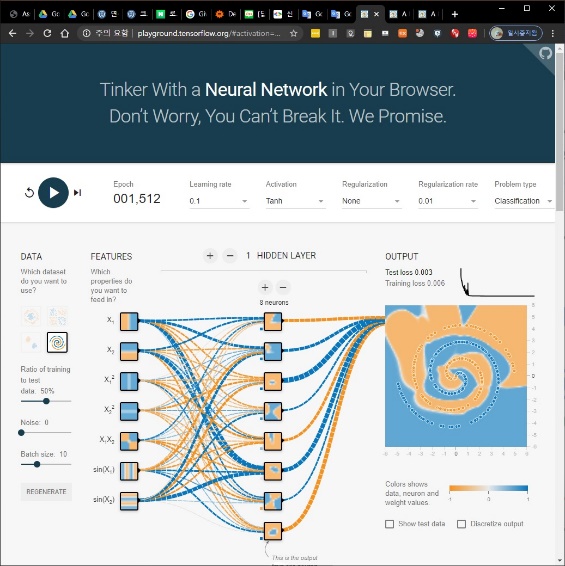


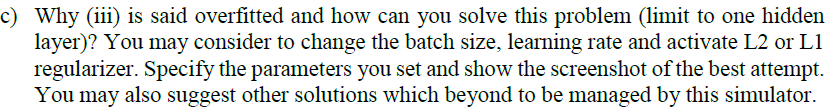
|  |  |  |
| --- | --- | --- |
|  | Training Loss | Test Loss |
| 1 | 0.323 | 0.52 |
| 2 | 0.341 | 0.459 |
| 3 | 0.154 | 0.672 |
| 4 | 0.382 | 0.883 |
| 5 | 0.25 | 0.261 |



is underfitted Because Training Loss is Big

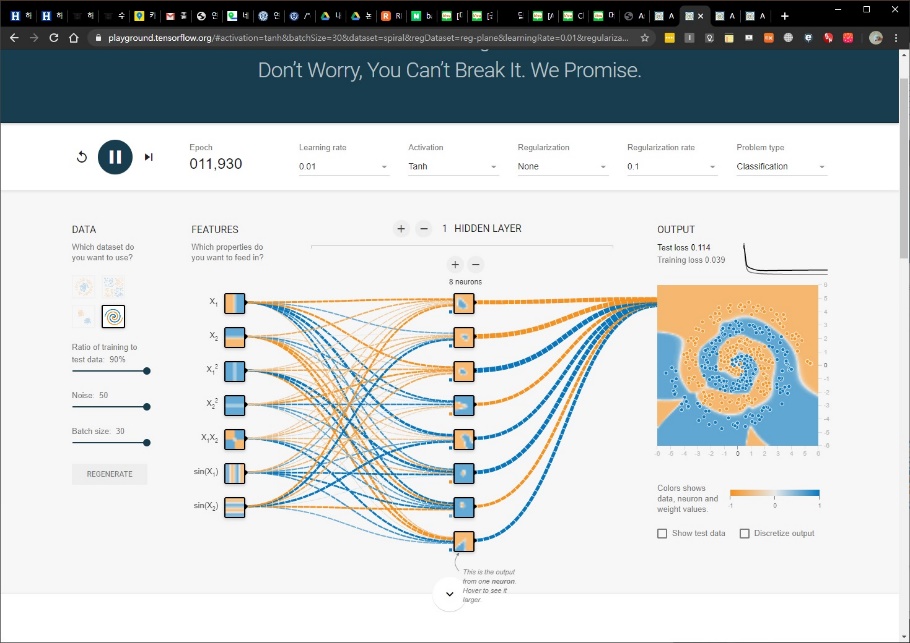
we can solve this problem by using variety of features.





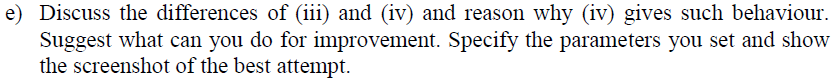
It has small Training Loss but has large Test Loss, So It is overfitted

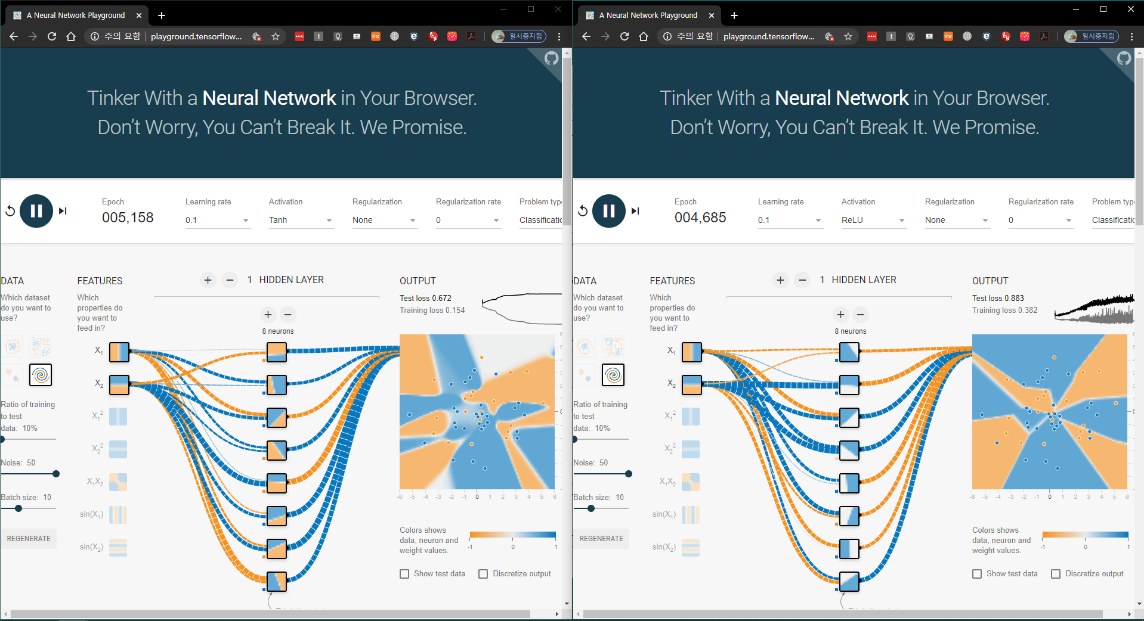
You should increase the batch size, Reduce the learning ratio



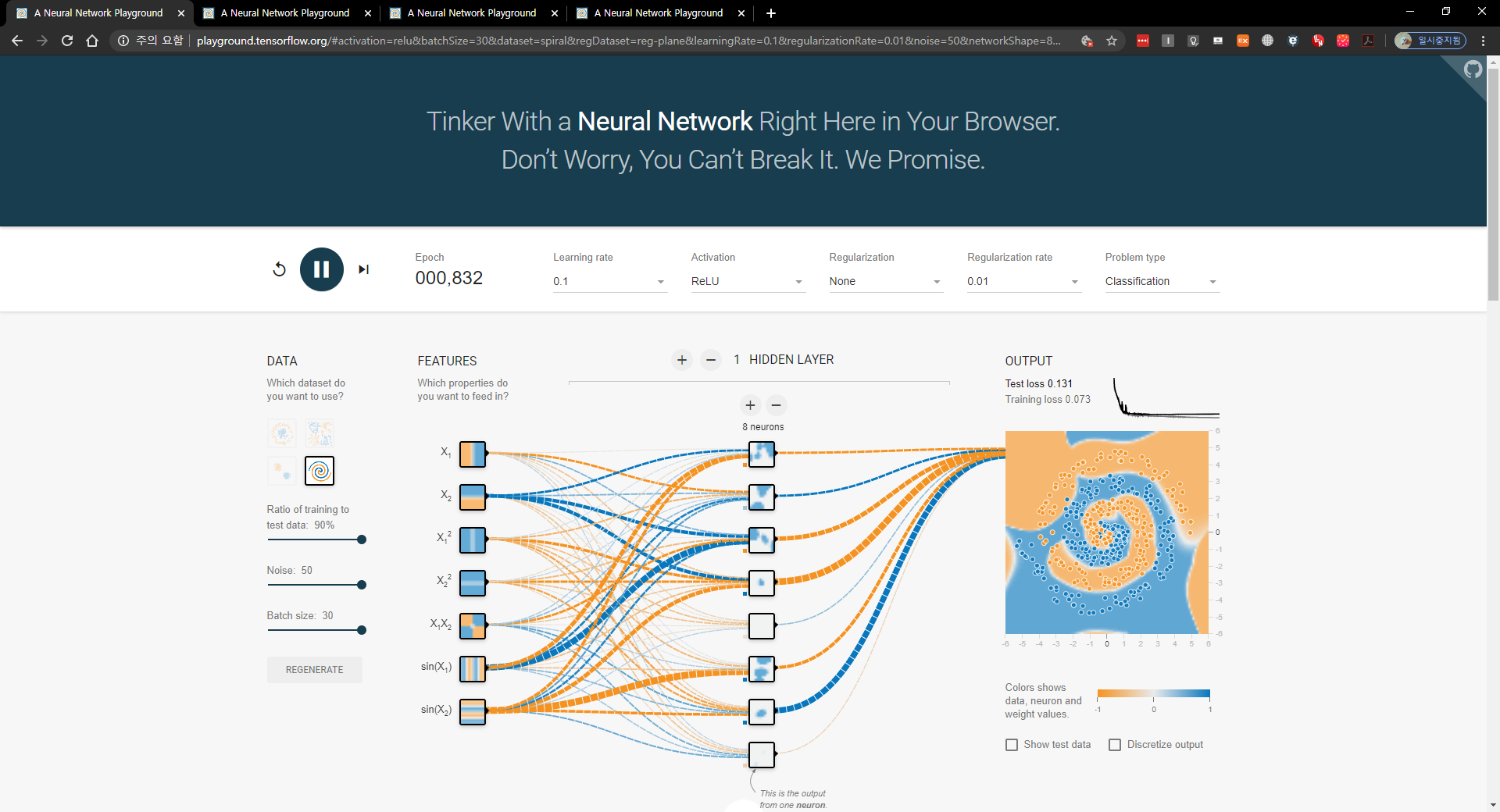


As the L2 coefficient increases, the underfitting becomes more severe, and accordingly, a lot of epoch training is possible.



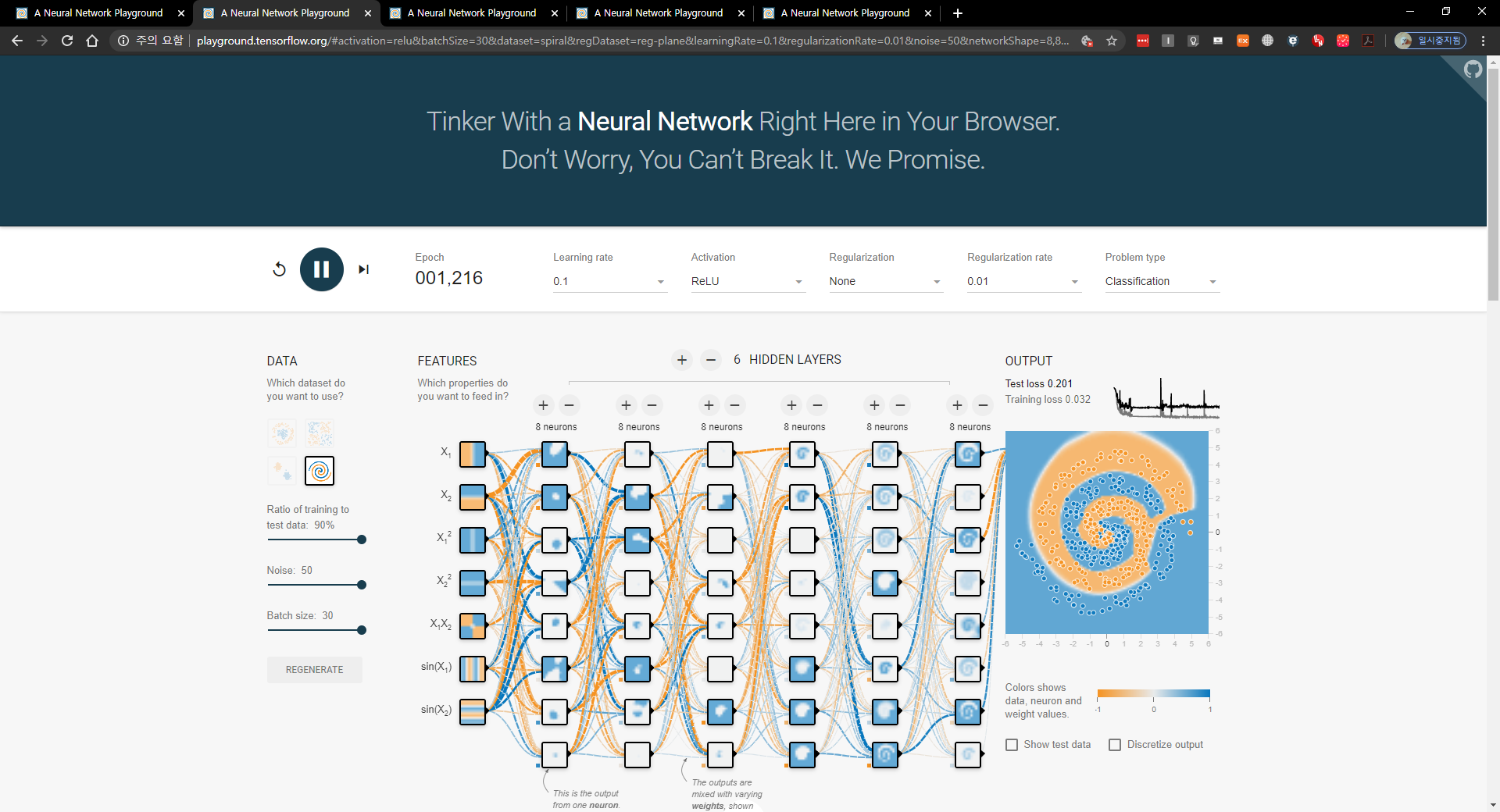


They both have high Test Loss but it has high Training Loss at ReLu

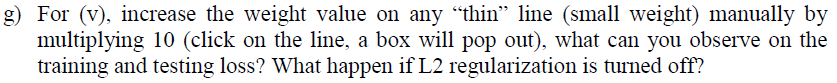


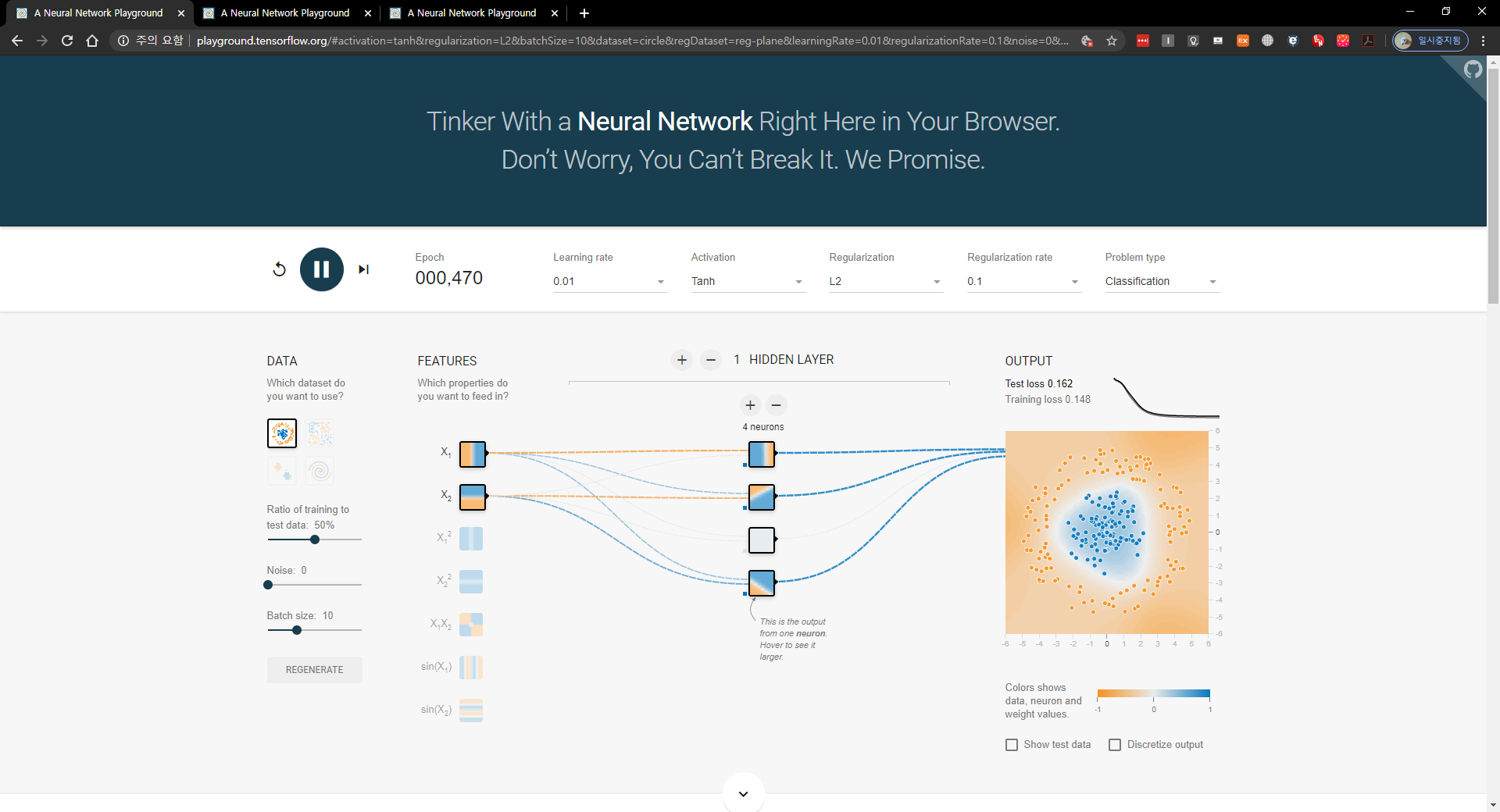
We can get better solution if we choose high Ratio of Training to test data and Batch Size

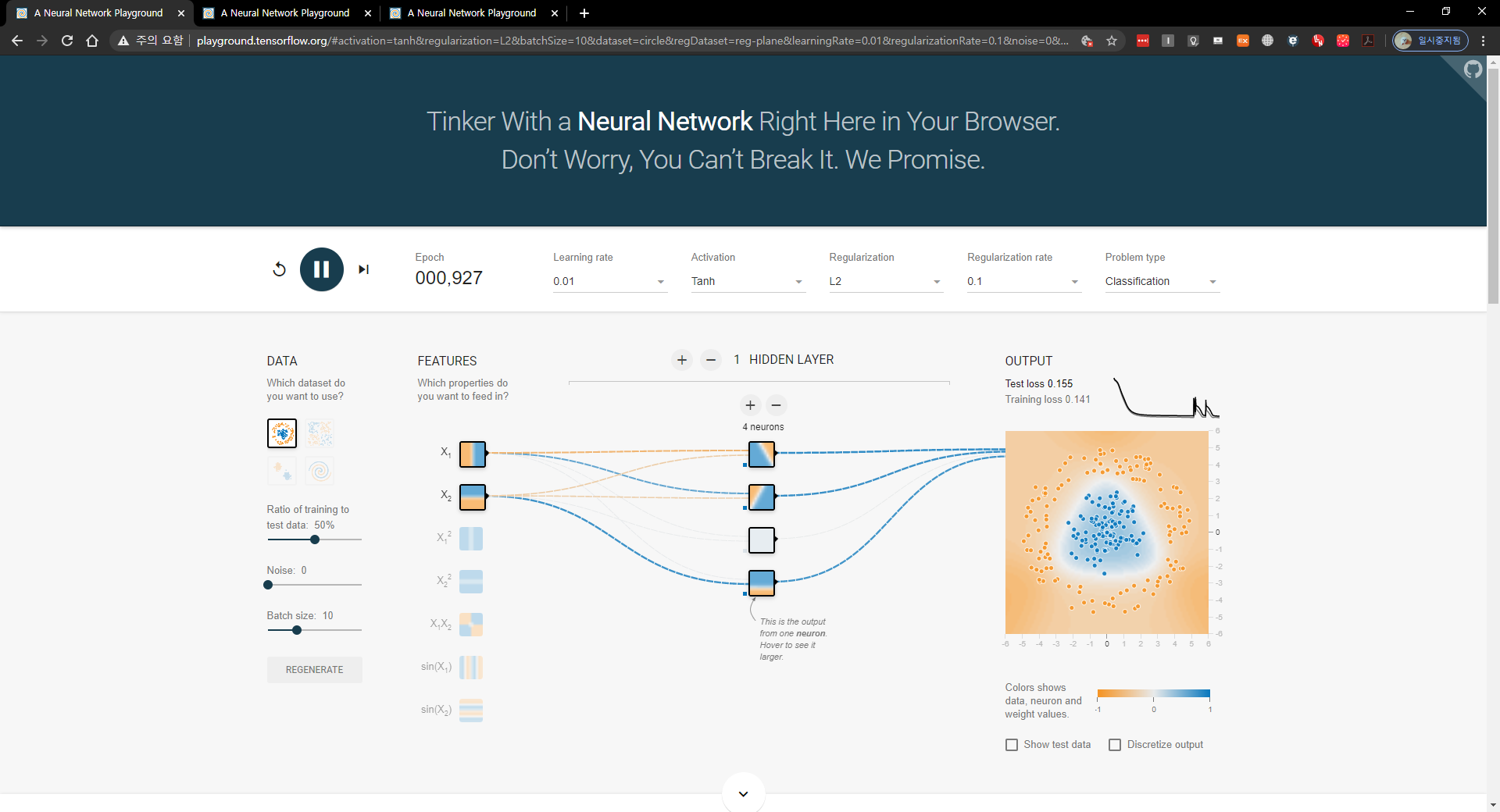




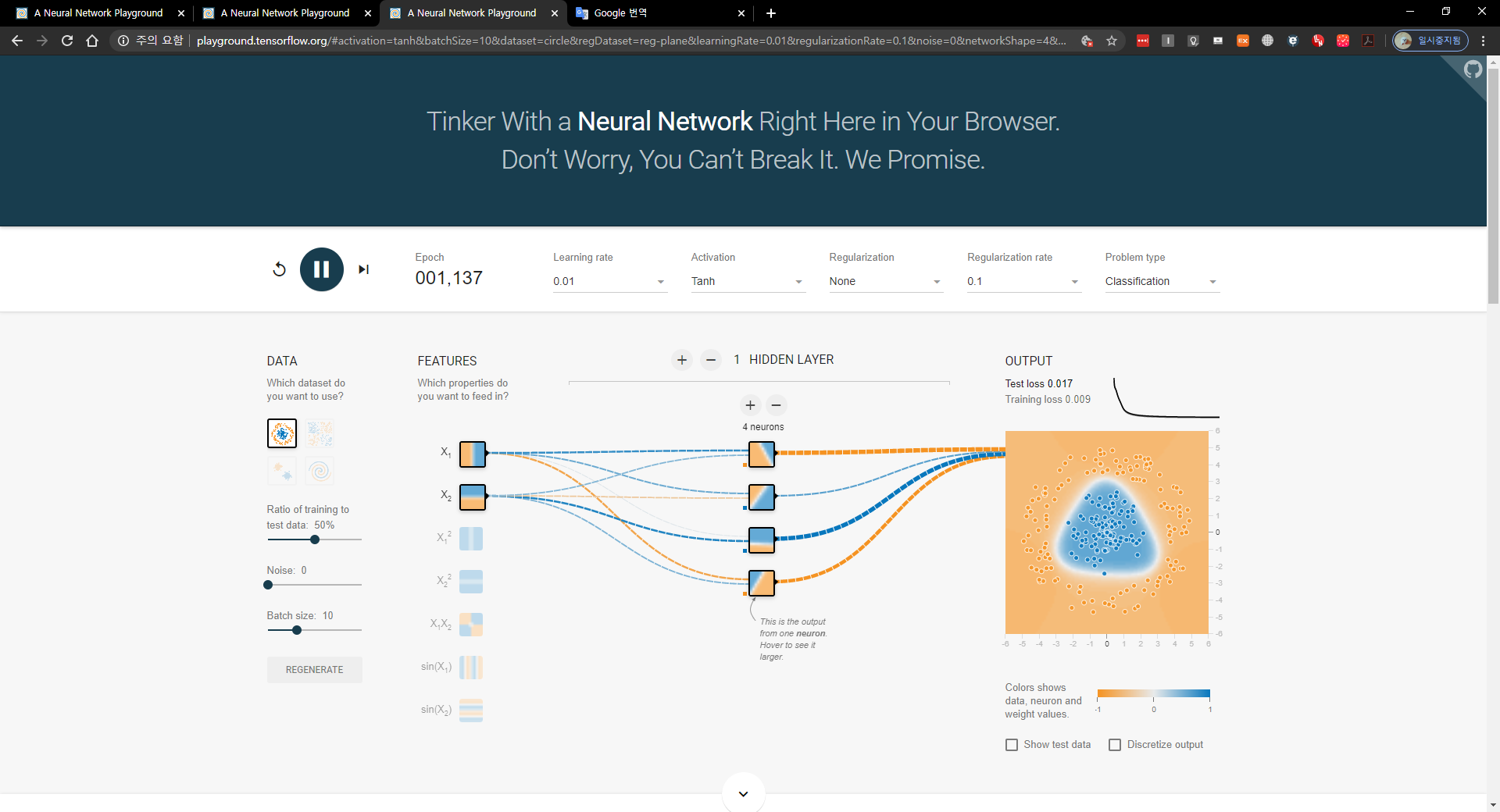
We have better result because of using more hidden layer as back-propagation





 weight changed

As you can see an output, Test loss and Training loss changed higher momentarily. But they Falls quickly.

 L2 turned off

Their boundary becomes more clear and Test loss and Training loss are gets lower.