# 1 Output Layer

## 1

## 2

,

## 3

# 2 Single Hidden Layer

## 1

## 2

## 3

# 3 UF Network

## 1

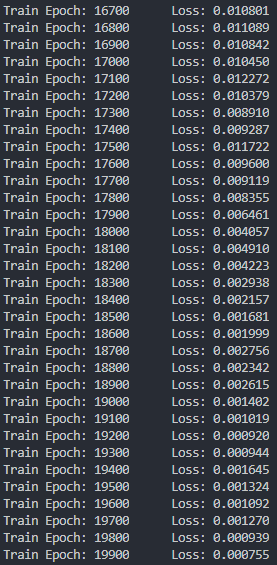
I will need both 80 units in the first and second hidden layers. Because if there are too few units, the training and testing result is not good (model fitting ability and accuracy are lower). But too many units may lead to overfitting and spend too much time. So, I choose 80 units after I tried to choose other number of units.

Figure 1 units=[70,70]

Figure 2 units=[80,80]

Figure 3 units=[90,90]

## 2

### a learning curve

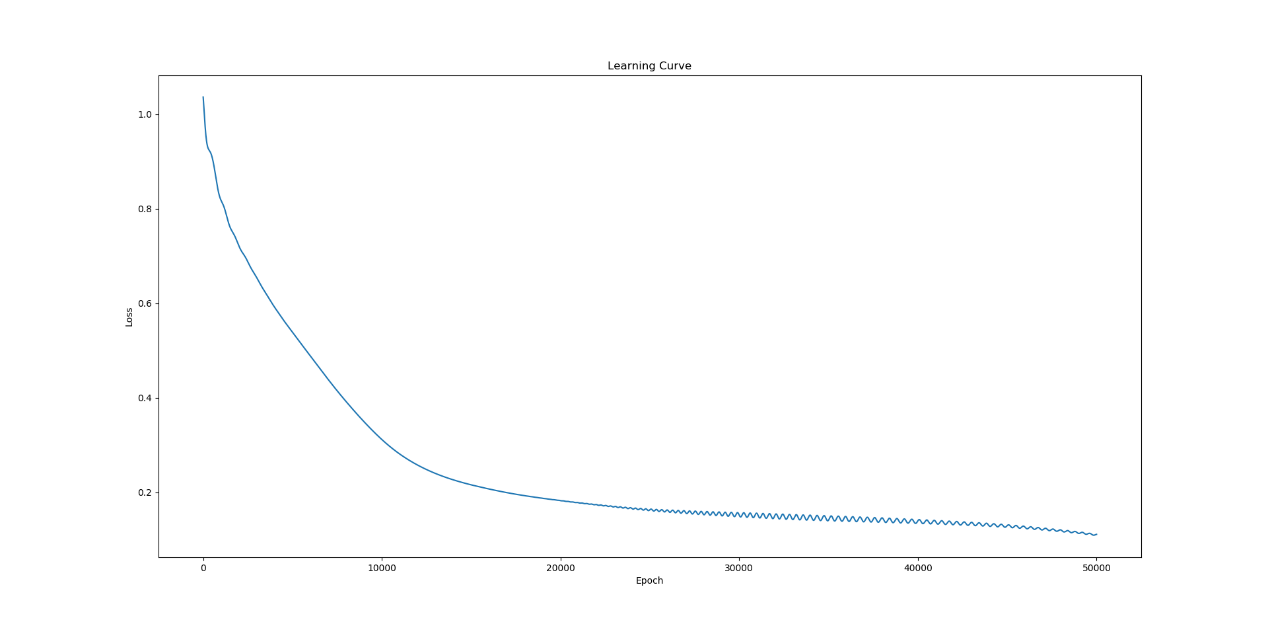


Figure 4 learning rate=0.00001 epochs=50000

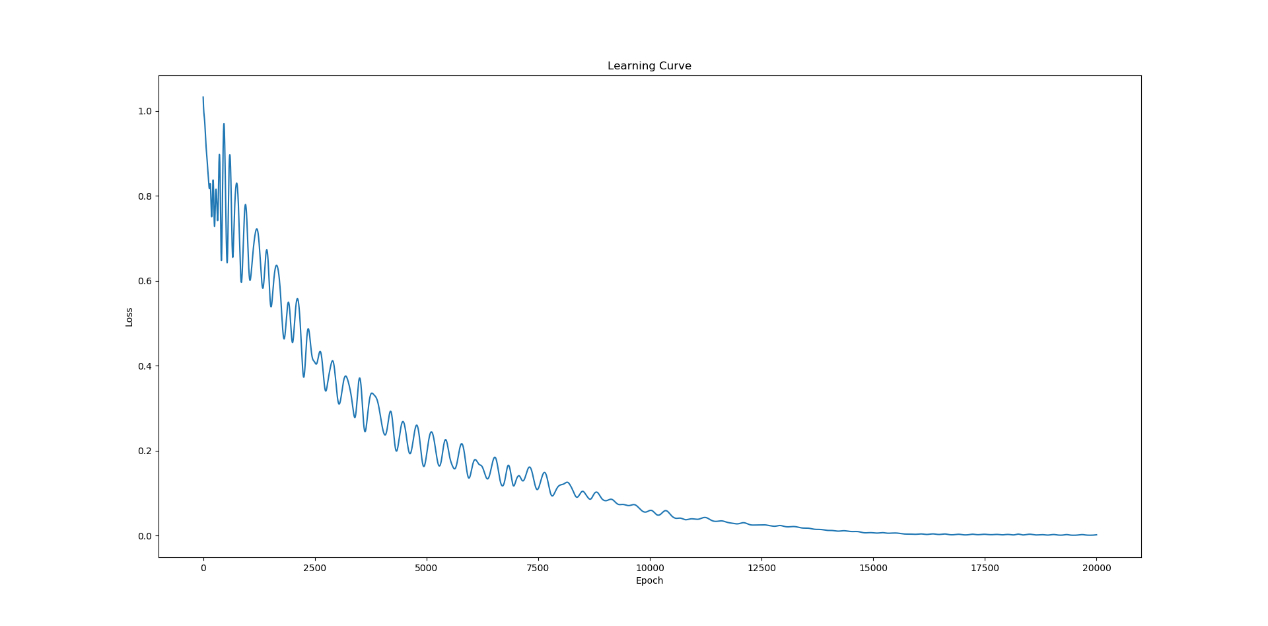


Figure 5 learning rate=0.0001 epochs=20000

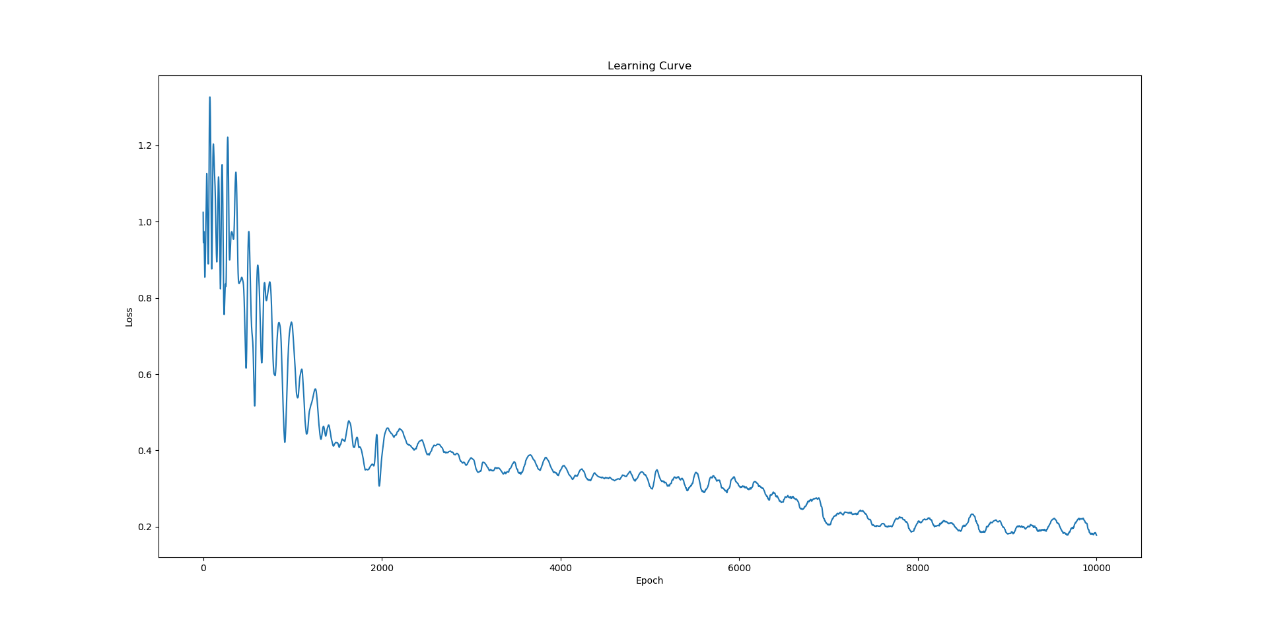


Figure 6 learning rate=0.001 epochs=10000

### b decision boundary

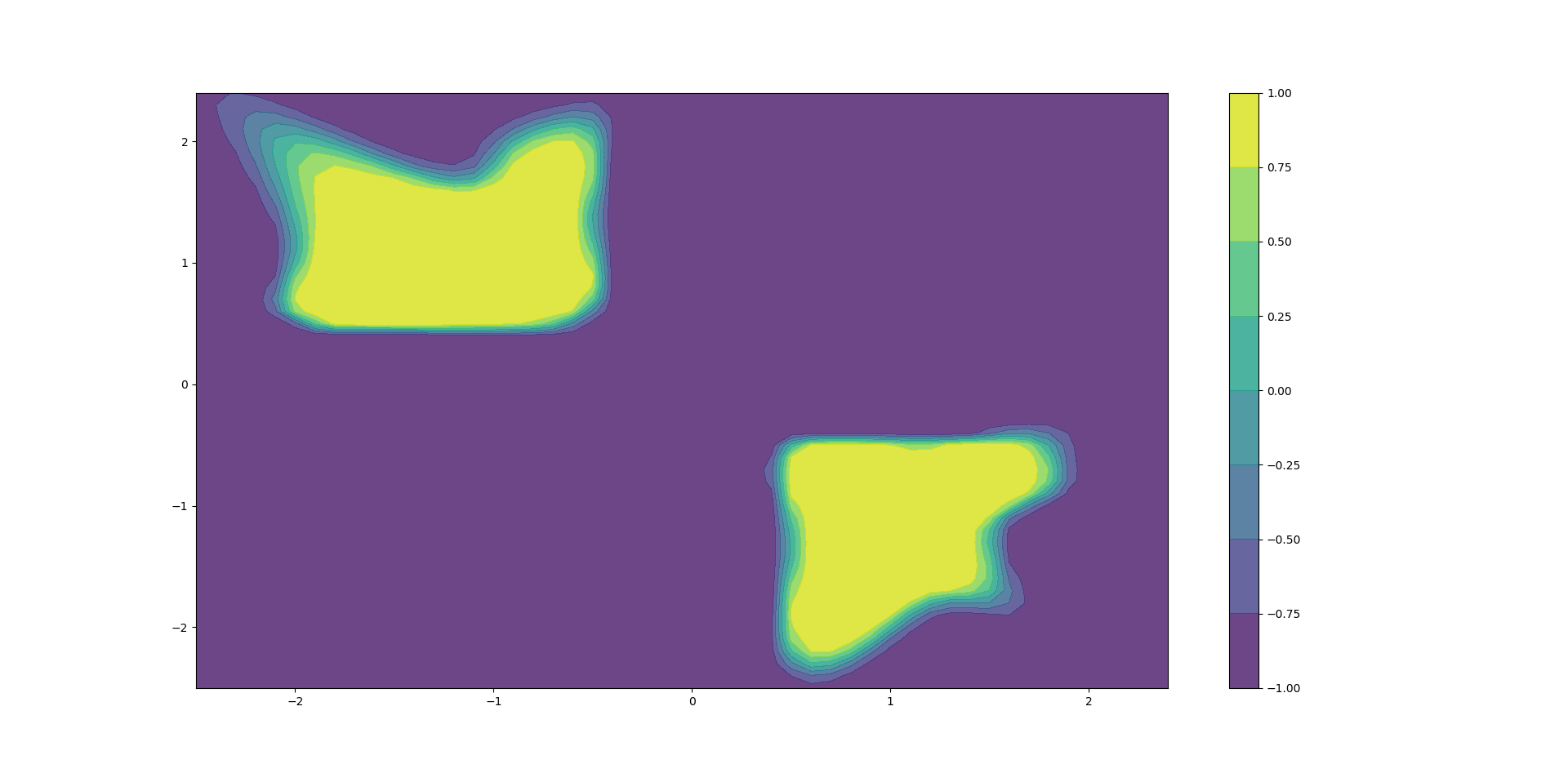


Figure 7 learning rate=0.00001 epochs=50000

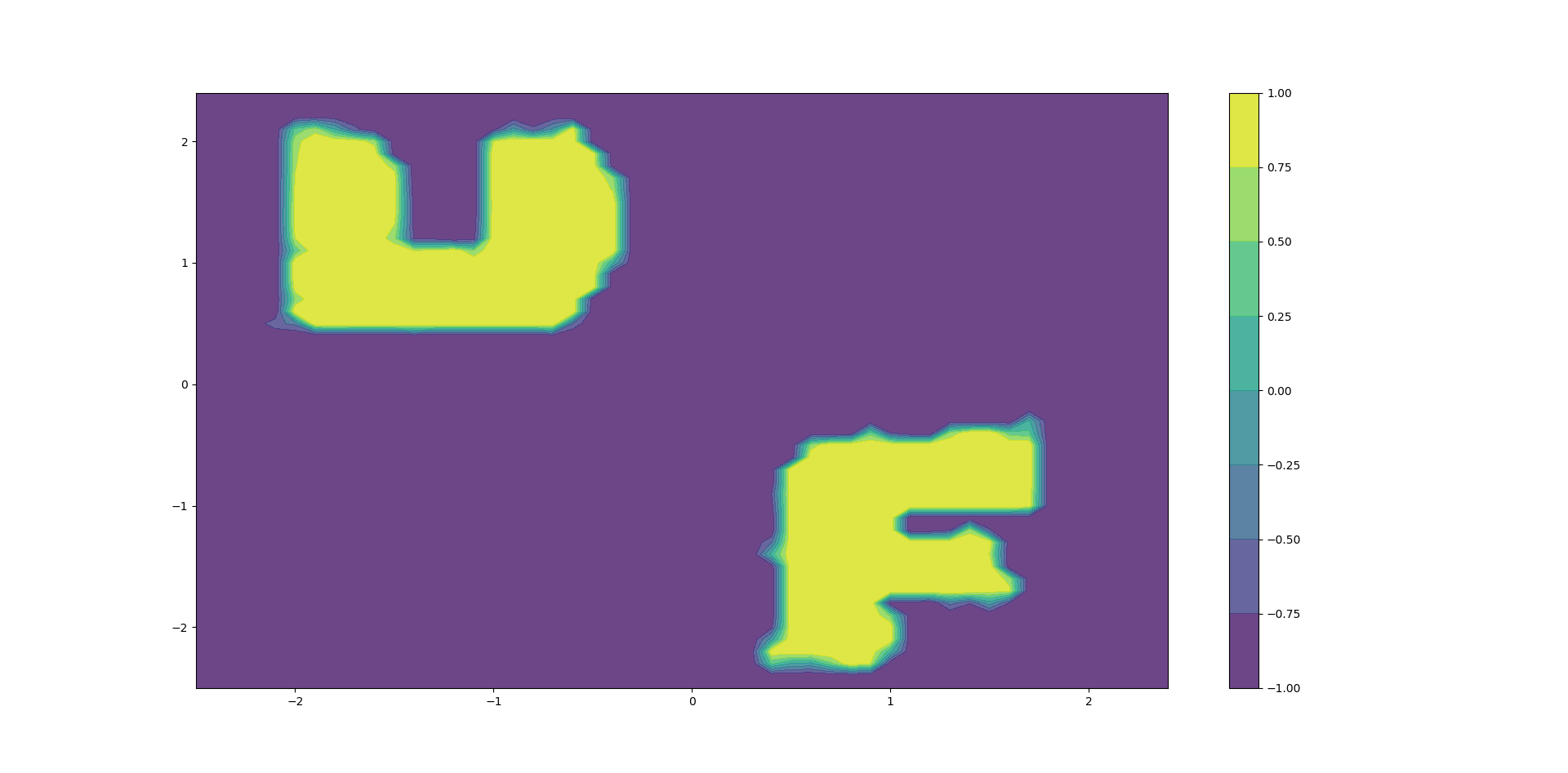


Figure 8 learning rate=0.0001 epochs=20000

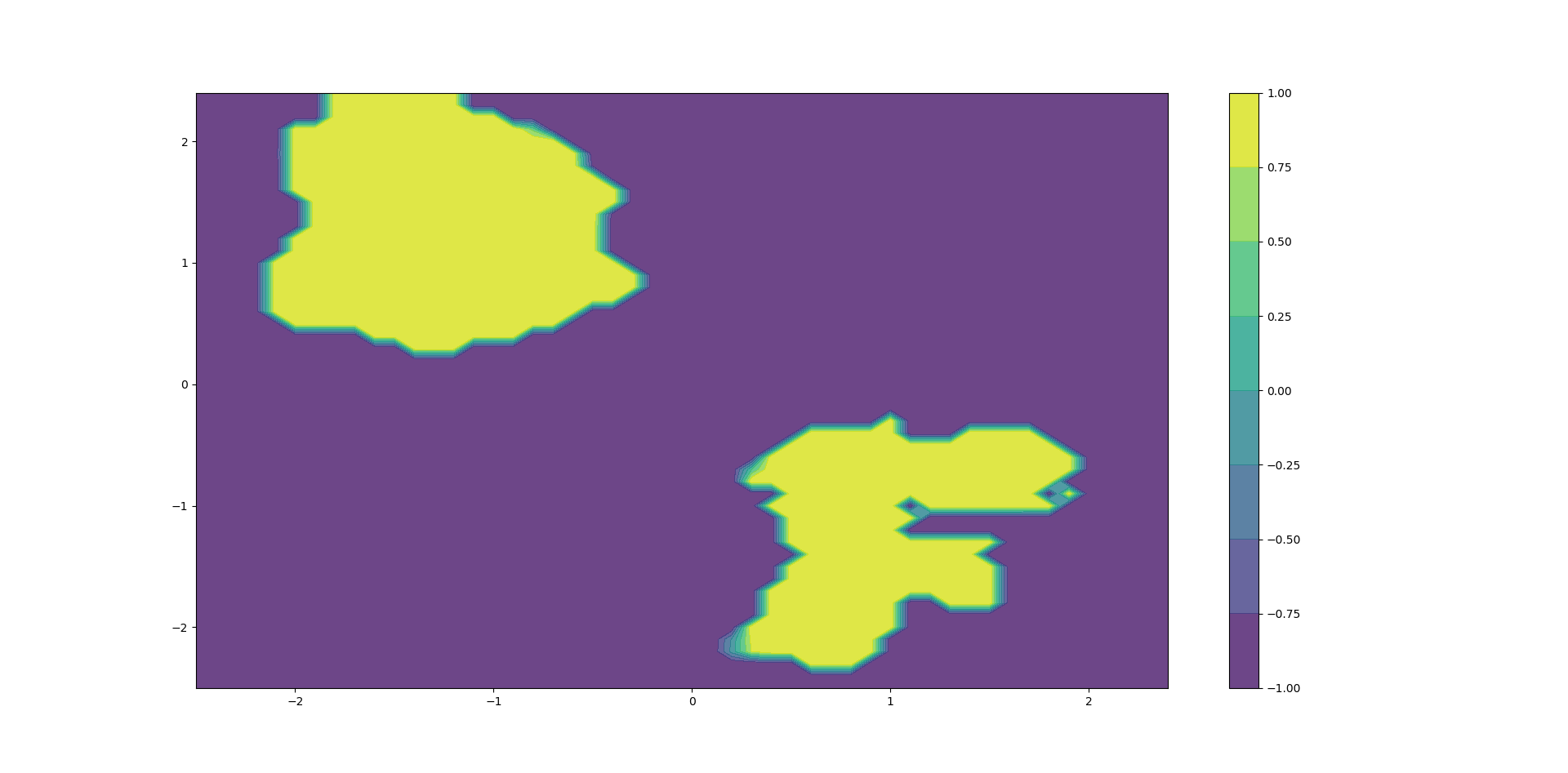


Figure 9 learning rate=0.001 epochs=10000

## 3

### a

If learning rate is small, gradient descent can be slow. And it will be cost much more time to optimize the model. If learning rate is high, gradient descent can overshoot the minimum. It may fail to converge, or even diverge.

If the number of epochs is small, the updates of the weights in the neural network is small so that the model is underfitting. If the number of epochs is big, the updates of the weights in the neural network is big and the model may overfitting. Also, it will cost too much time to training the model.

### b

Because I have tried many values and selected the best learning rate and number epochs whose learning curve and decision boundary look better.