**Question 1**

**Part a**

* The Neural Network initially has 3 hidden layers with 3 neurons each. The hidden layers have a RELU activation function and output layer has a SoftMax activation function.
* Weights are initialized from a uniform distribution in the range [0, 1].

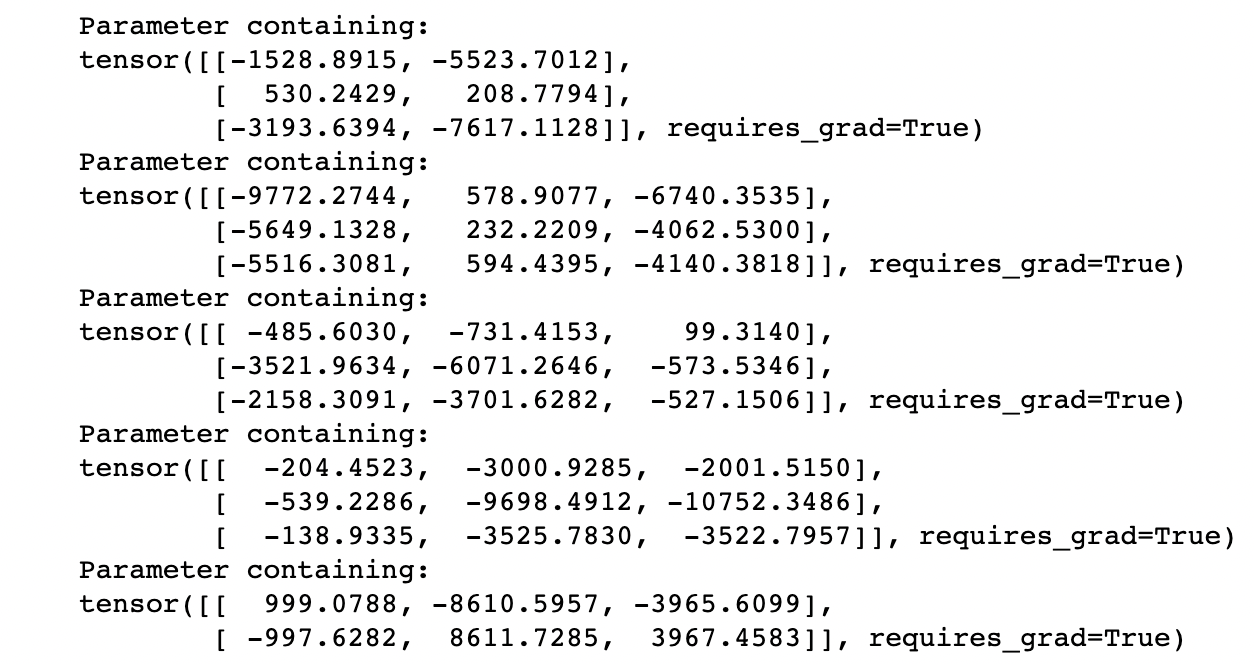
The XOR data is not linearly separable. The result is as follows:



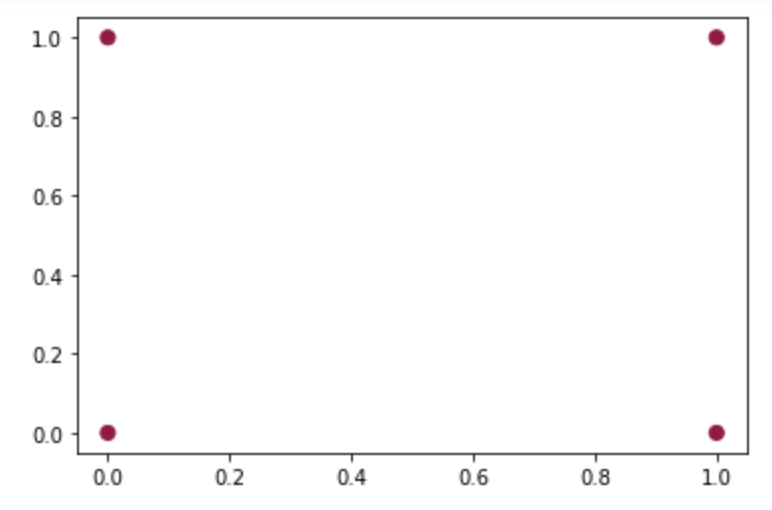
We specified the stopping criteria as 20,000 epochs or an accuracy of 1, we trained the neural network. These were the results:

* The training stops at a little over 1200 epochs.
* Loss = 0.7884
* Accuracy = 1.0

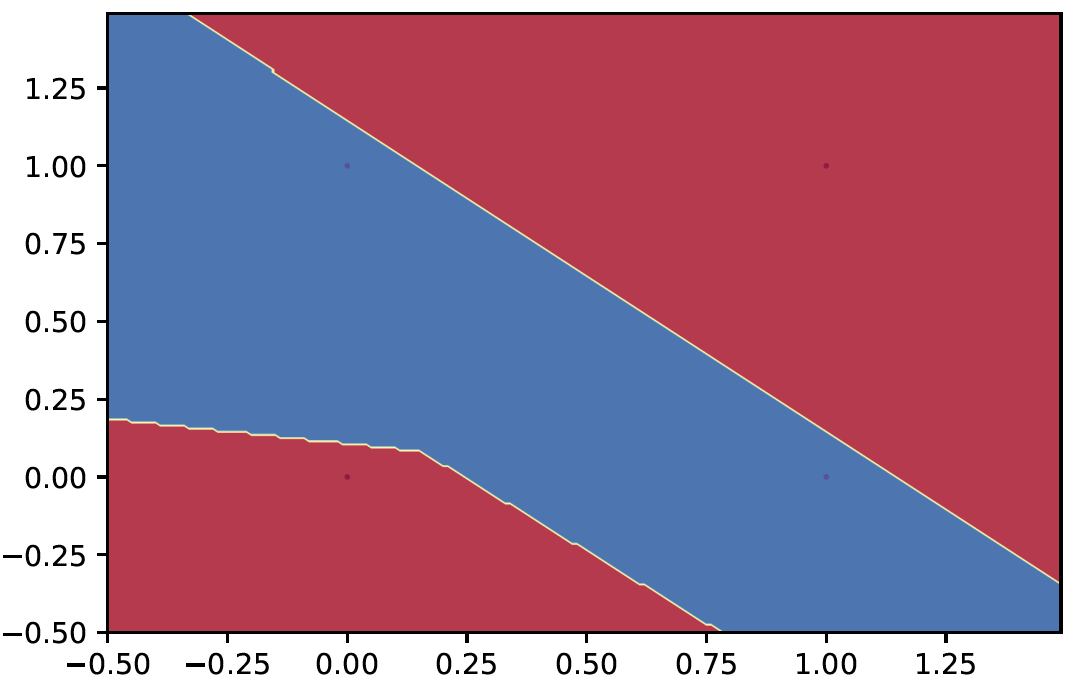
Parameters



Predicted values:



Decision Boundary:



**Part b**

* We used the same Neural Network architecture.
* Weights are initialized from a uniform distribution in the range [0, 1].
* Stopping criteria is 40,000 epochs or loss is less than 10-4

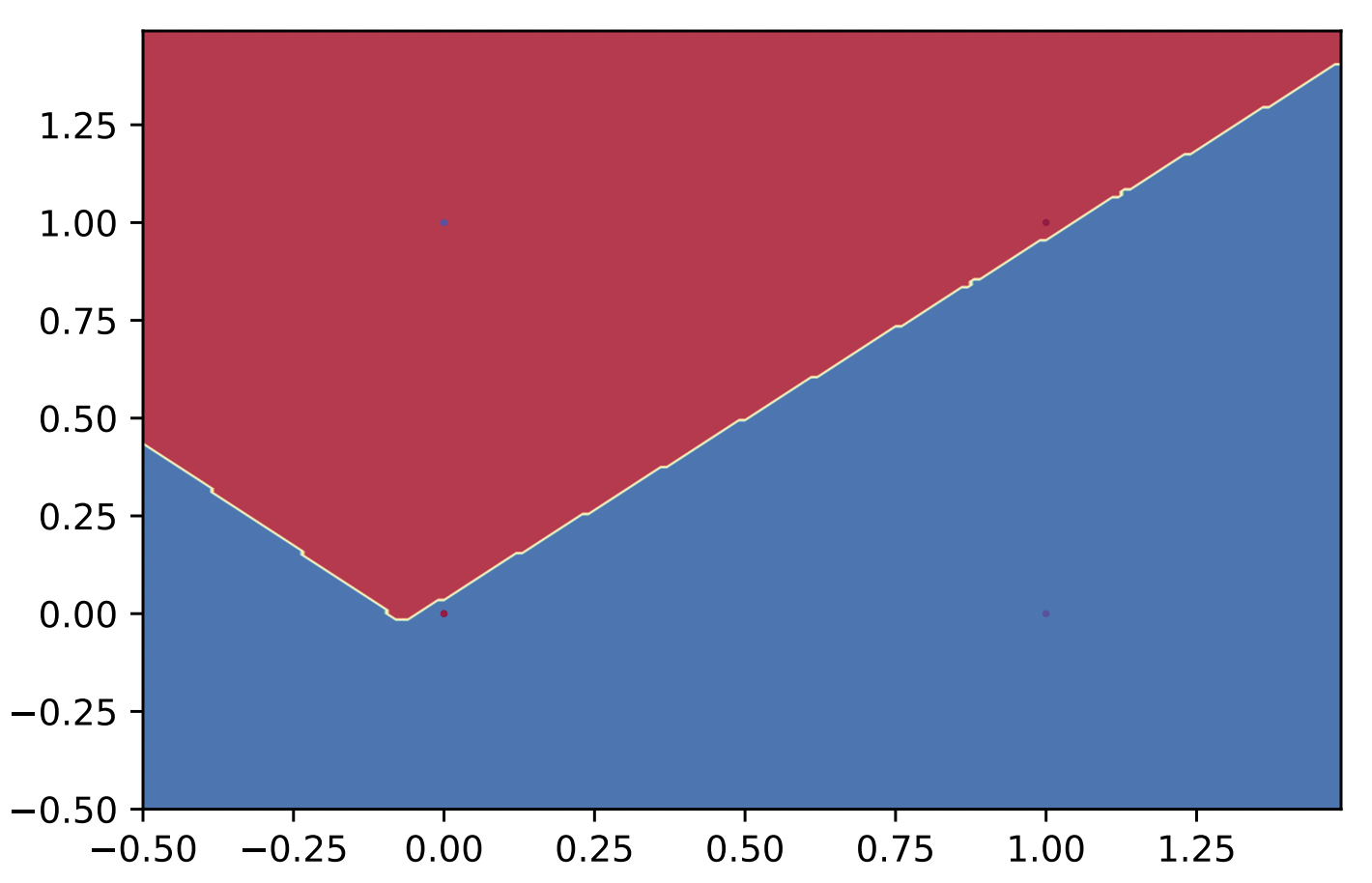
The results are as follows:

* The training stops at 7224 epochs.
* Loss = 9.997 x 10-5
* Accuracy = 1.0

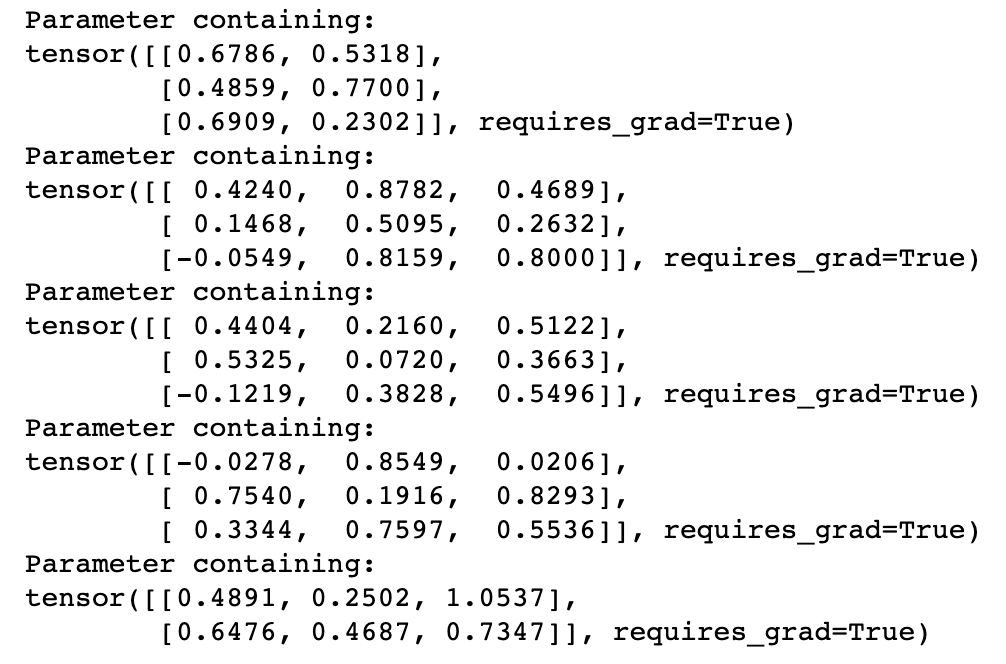
The predicted values can be plotted as follows.



Decision boundary:



Parameters:

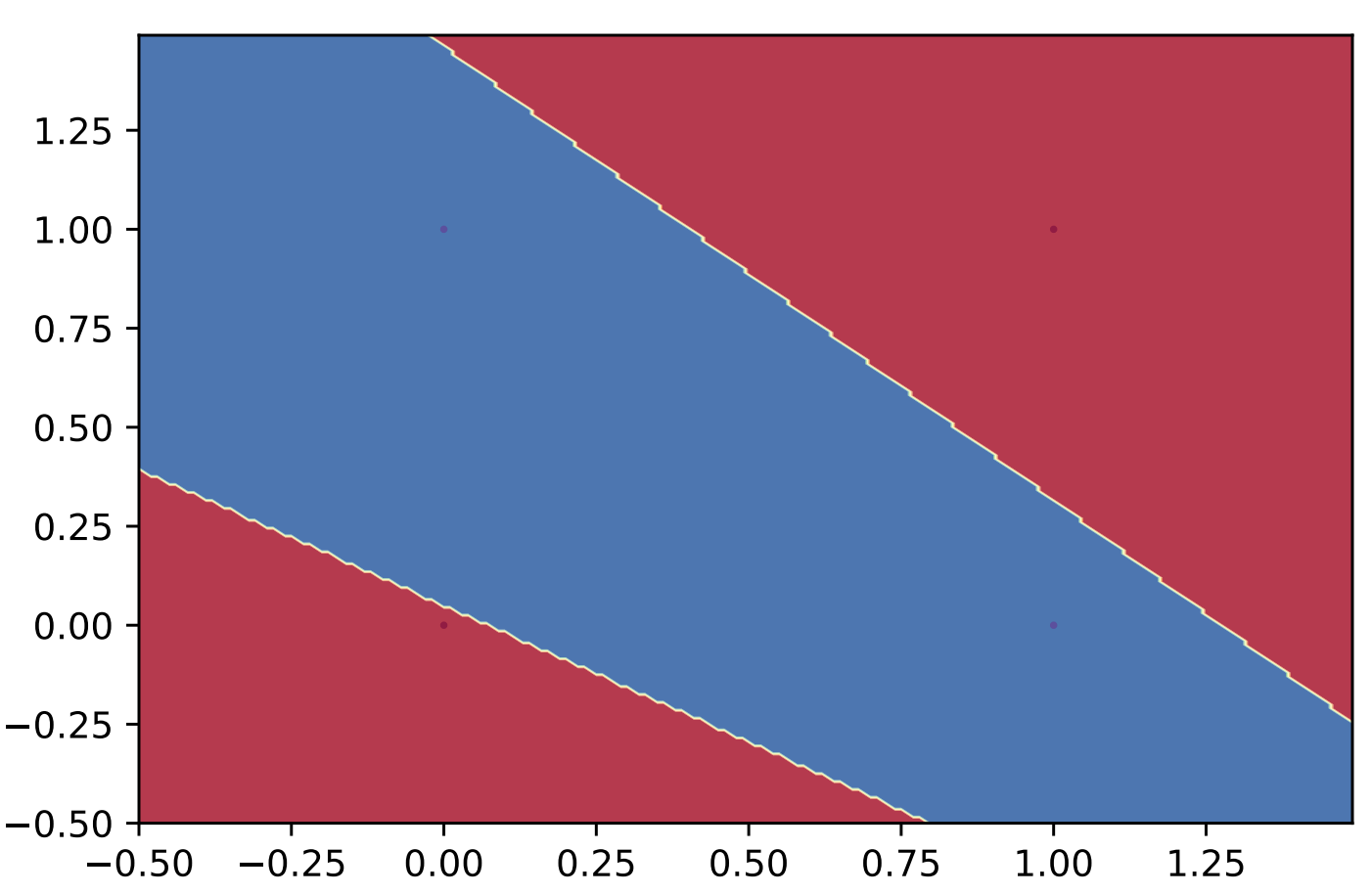


**Part c**

* We used a Neural Network with 2 layers with 2 neurons each.
* The input layer has a RELU activation function and the output layer a softmax activation function.
* Weights are initialized from a uniform distribution in the range [0, 1].
* Stopping criteria is 40,000 epochs or an accuracy of 1.

The model trains for about 270 epochs to a loss of 0.3557 and an accuracy of 1.

Decision boundary:



Parameters:

