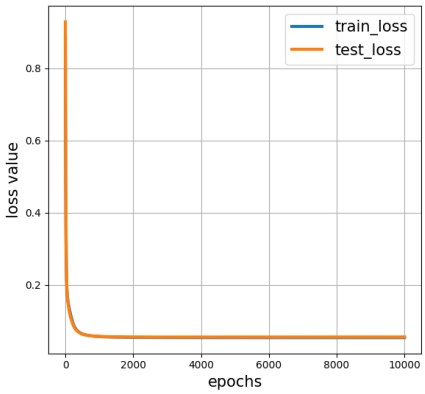
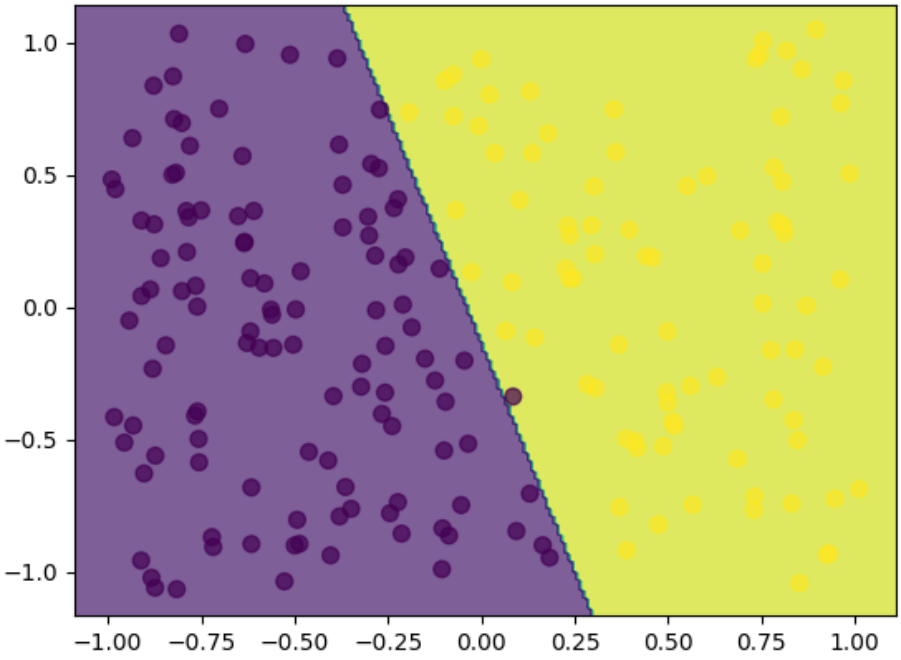
18786 HW2 - Training multilayer perceptrons

Andrea Vigano

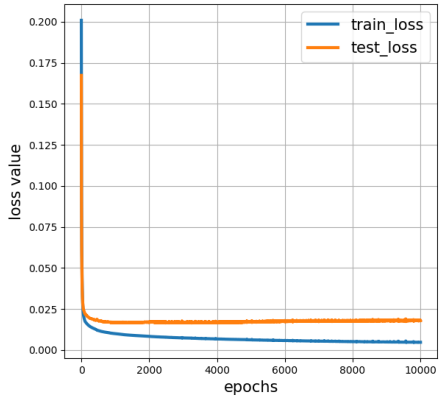
14 February 2025

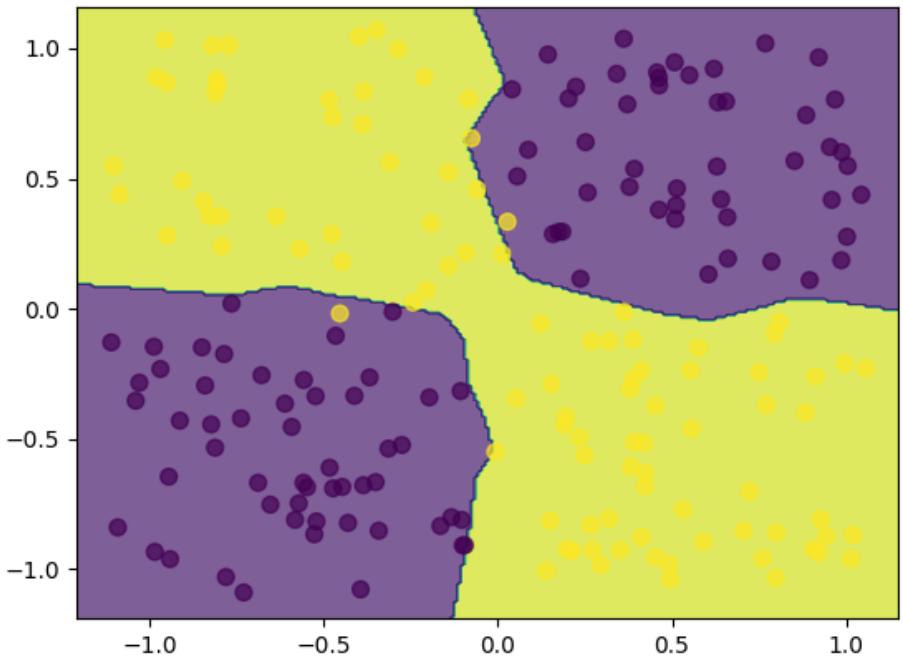
1 Deliverable 2 - Linear separable dataset





2 Deliverable 3 - XOR



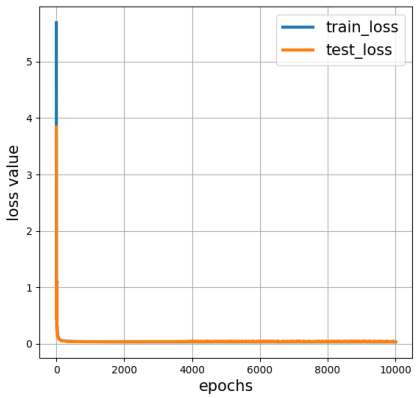


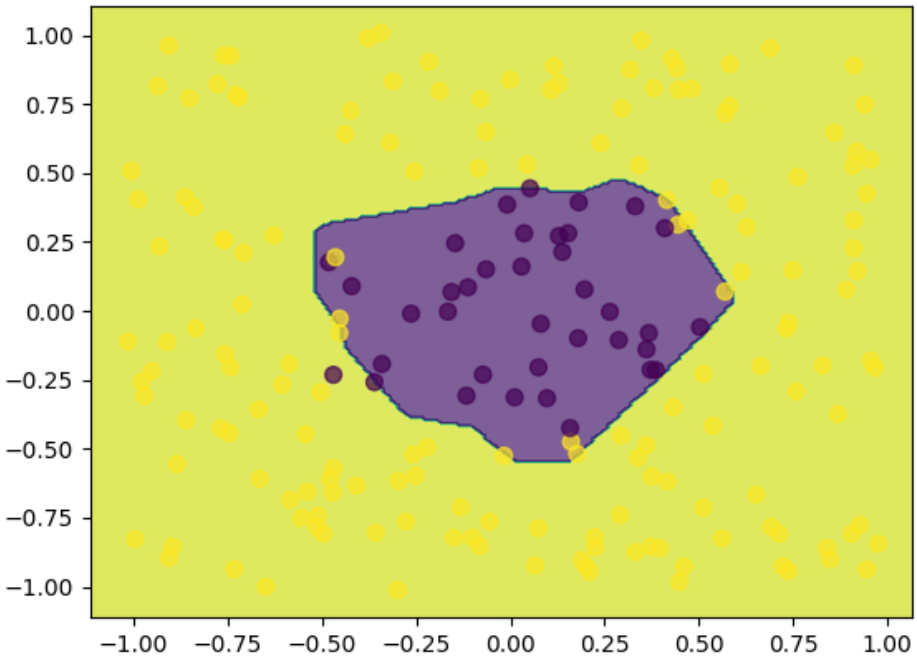
The network consisted of a single hidden layer with 64 hidden neurons. The hidden layer used ReLU activation and the output was a single neuron with Sigmoid activation. The weights were initialized using Xavier initialization, and the training was carried out for 10000 epochs using full batch Adam GD with

binary cross-entropy loss.

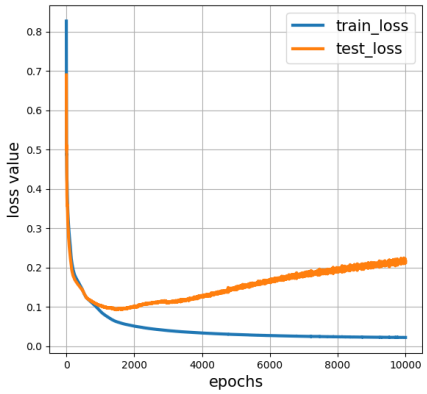
3 Deliverable 4 - Differences in cost function

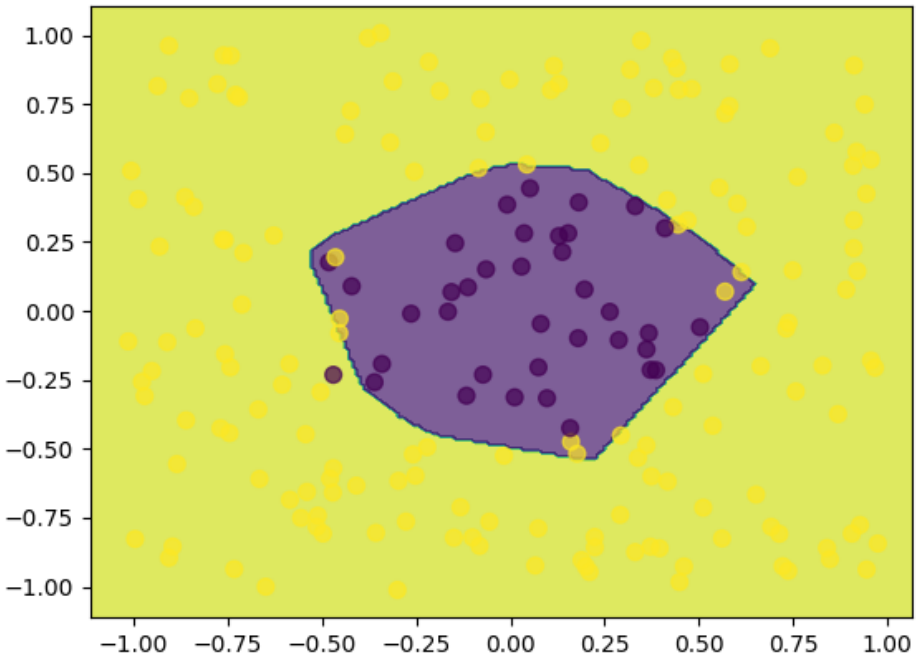
Regressor





Classifier

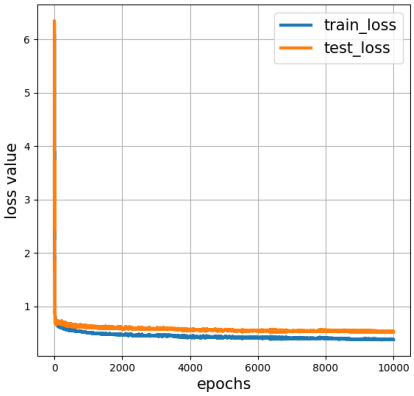


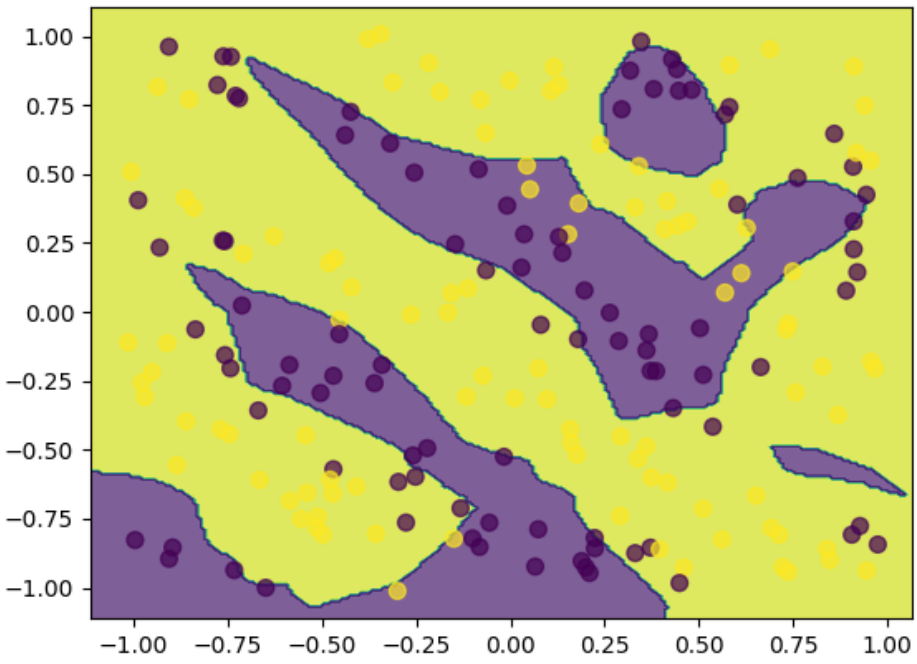


Both networks consisted of a single hidden layer with 32 units and ReLU activation. Output activation was Linear for the regressor and Sigmoid for the classifier. Both were trained for 10000 epochs with Adam. The regressor used L2 loss and the classified used binary cross-entropy.

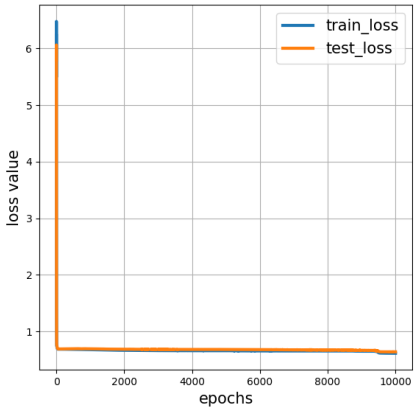
4 Deliverable 5 - Differences in optimizers

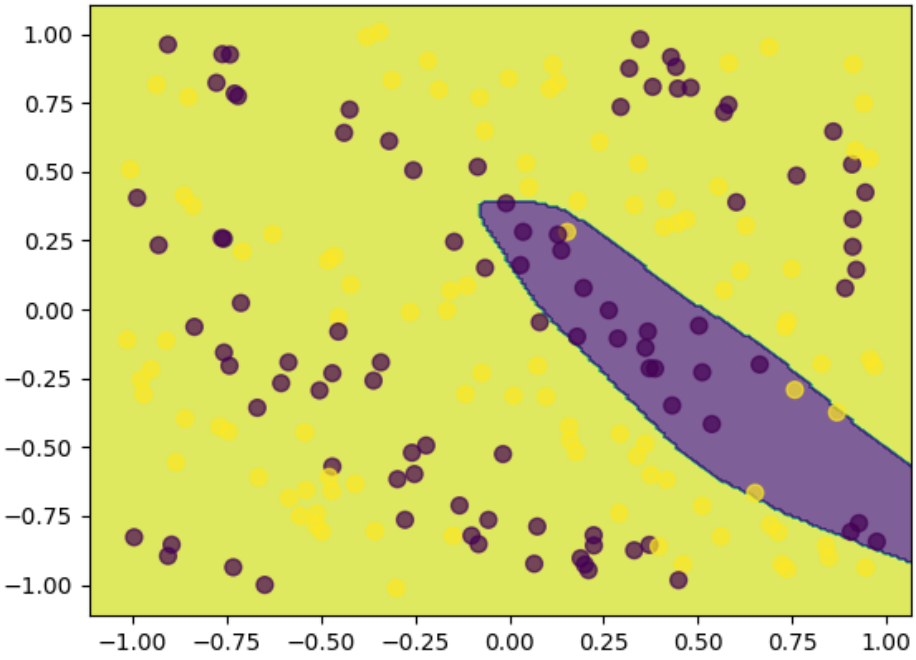
Vanilla gradient descent



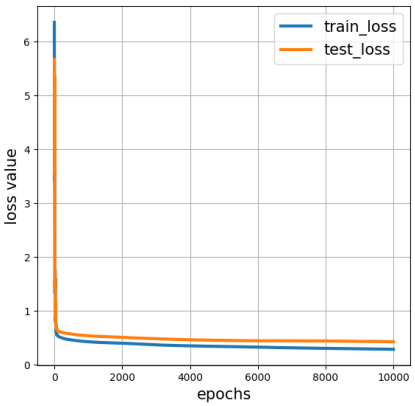


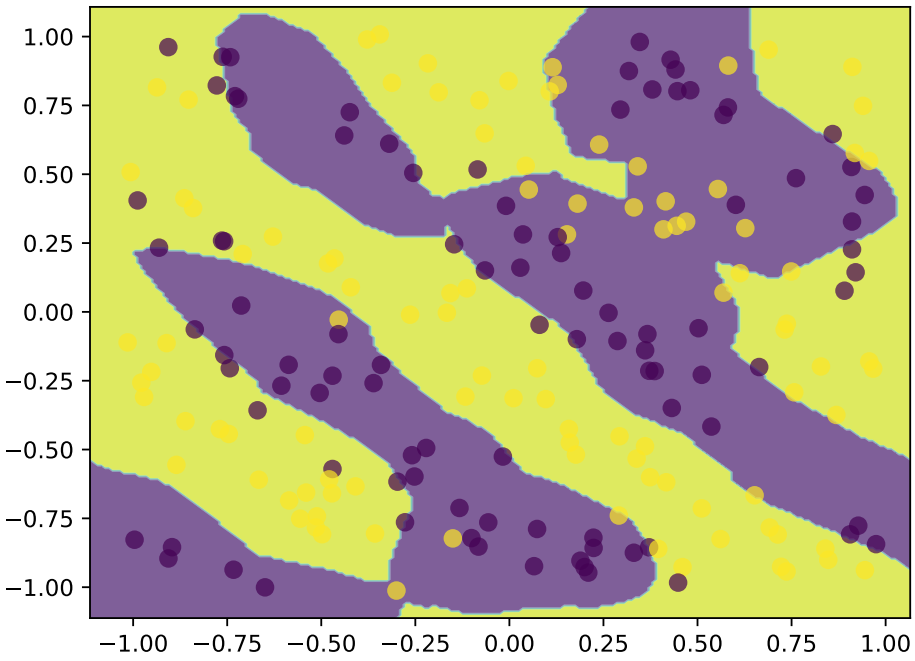
Momentum gradient descent



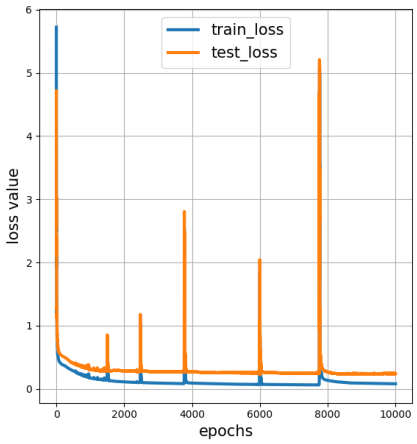


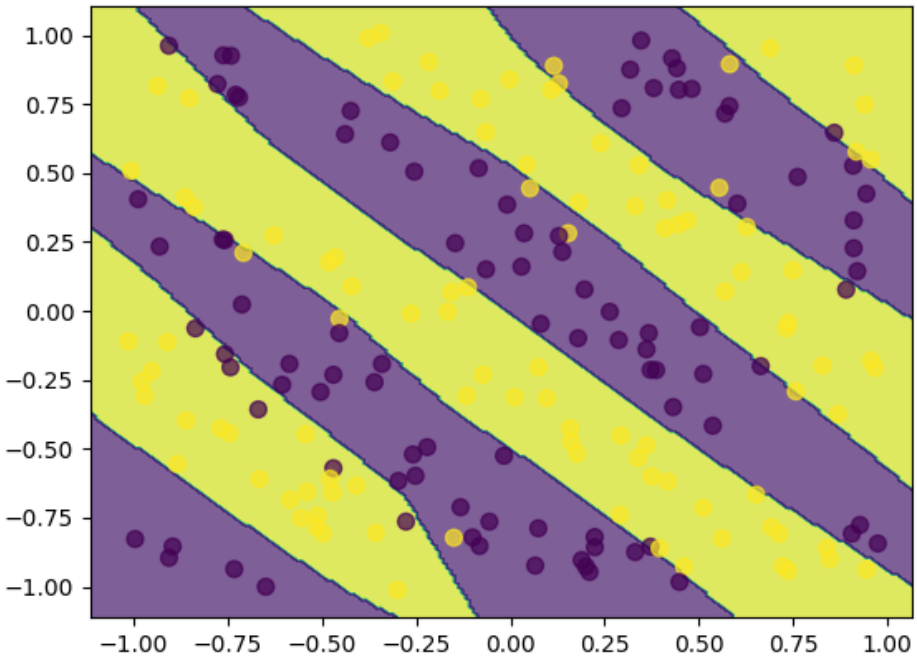
Note that after fine-tuning the learning rate, the performance of momentum GD improves significantly.



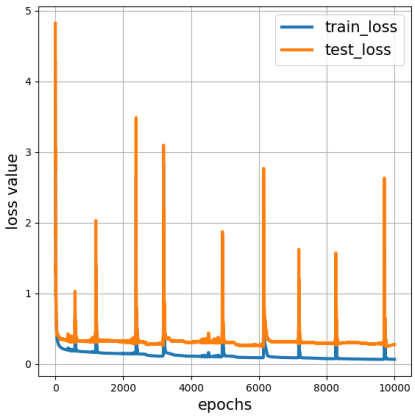


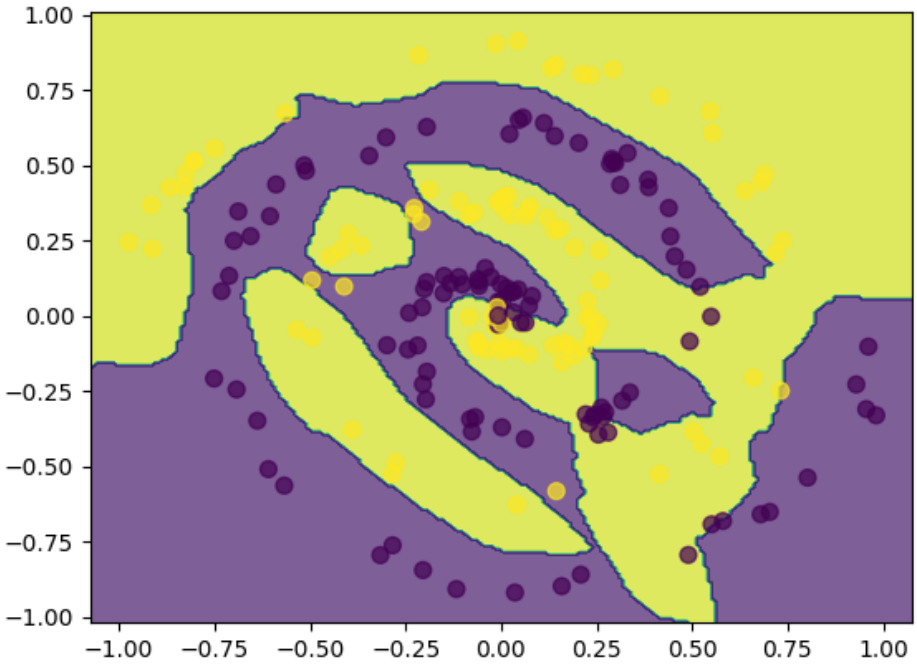
Adam gradient descent


train_loss
test_loss
loss value
epochs
0
1
2
3
4
5
6
0
2000
4000
6000
8000
10000


1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00

5 Deliverable 6 - Swiss roll





Below is the architecture of the network used:
NUMBER_OF_LAYERS: 5,

LAYERS_IO: ((2, 32), (32, 64), (64, 16), (16, 4), (4, 1)),

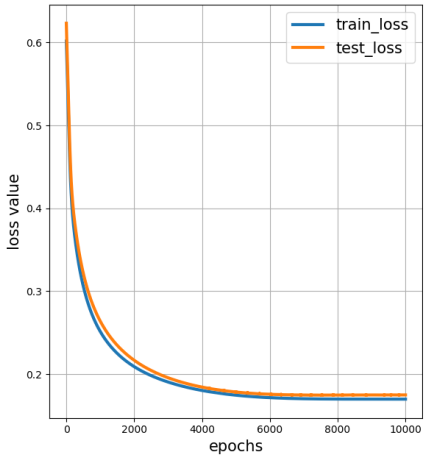
LAYERS_ACTIVATIONS: (ActivationType.RELU, ActivationType.RELU, ActivationType.RELU, ActivationType.RELU, ActivationType.SIGMOID),

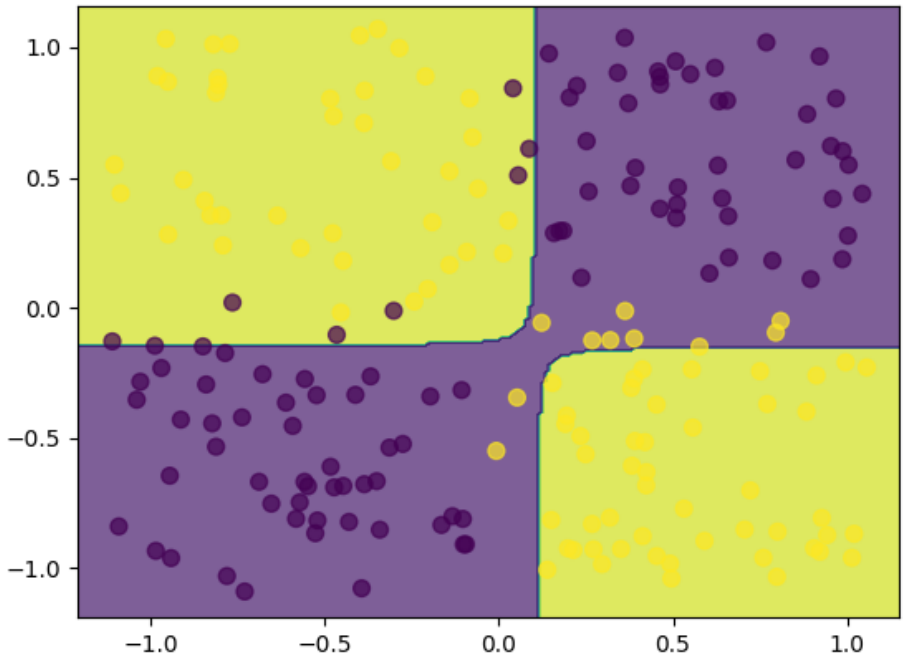
LAYERS_INITIALIZATIONS: (Initialization.HE, Initialization.HE, Initialization.HE, Initialization.HE, Initialization.HE),

It was trained for 10000 epochs with binary cross-entropy loss and Adam optimizer.

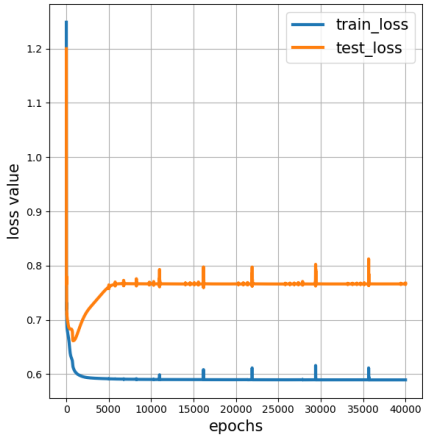
6 Deliverable 7

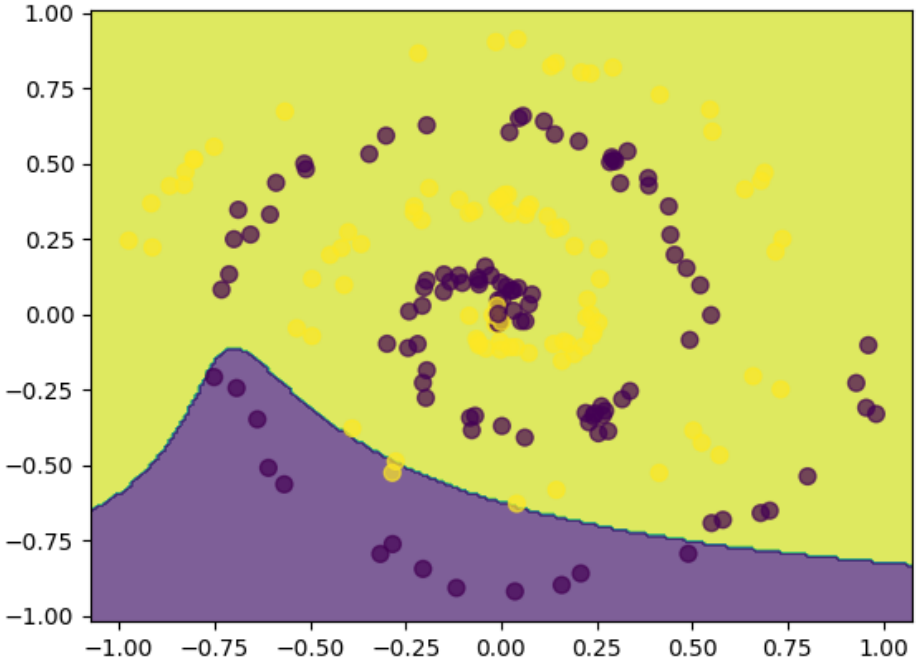
XOR




1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0

Swiss-roll


train_loss
test_loss
loss value
epochs
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0
5000
10000
15000
20000
25000
30000
35000
40000


1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00