Pattern Recognition and Machine Learning

Lab - 9 Assignment Report

Aryan Himmatlal Prajapati (B21EE012)

Question 1.

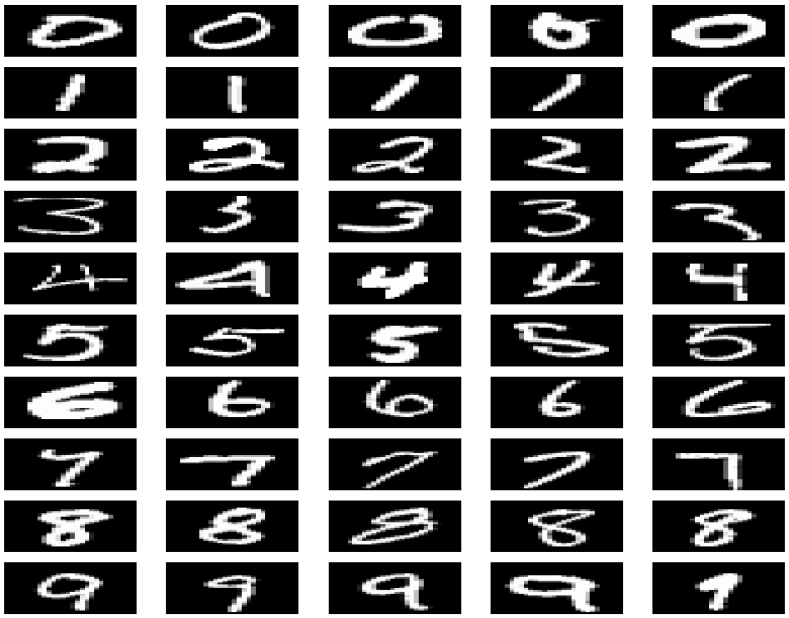
Part A

Downloaded the MNIST dataset using torch-vision. Splitted into train, test and validation dataset. Applied Augmentations to images:

a. Training dataset: RandomRotation(5 degrees), RandomCrop(size=28, padding =2), ToTensor and Normalize.
b. Testing dataset and validation dataset: ToTensor and Normalize

Part B

Plotted a Few Images from each class. Created a data loader for the training dataset as well as the testing dataset.



Part C

Wrote a 3-Layer MLP using PyTorch all using Linear layers. Printed the number of trainable parameters of the model.

```
Number of Trainable Parameters: 261510
```

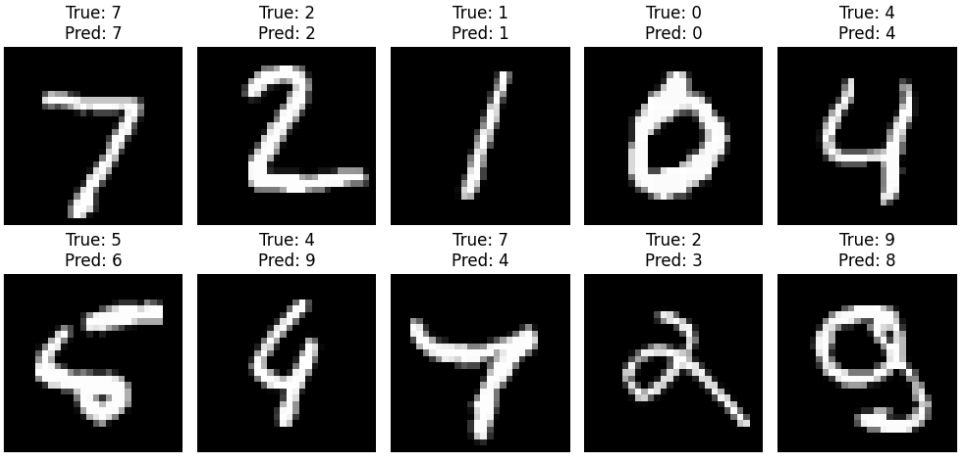
Part D

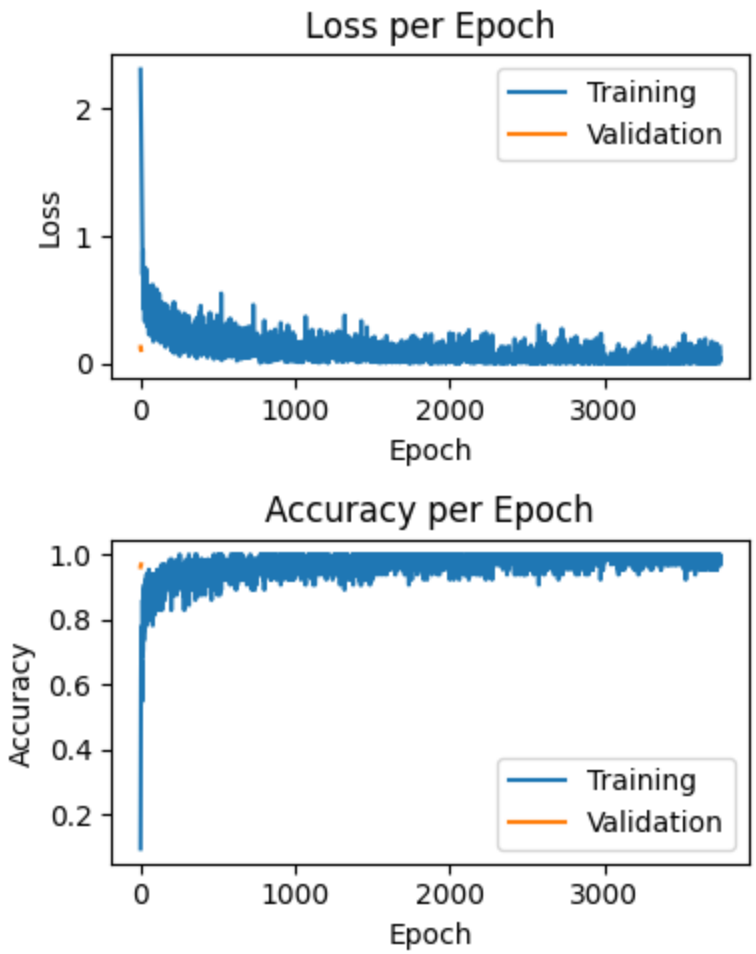
Trained the model for 5 epochs using Adam as the optimizer and CrossEntropyLoss as the Loss Function. Evaluated the model on the validation set after each epoch and saved the best model as well as log the accuracy and loss of the model on training and validation data at the end of each epoch.

Epoch 1 : Training Loss: 0.2039342224597931,Validation Loss: 0.13112191573538362 , Validation Accuracy: 0.961
Epoch 2 : Training Loss: 0.19848671555519104 , Validation Loss: 0.12033595695616082 , Validation Accuracy: 0.9655
Epoch 3 : Training Loss: 0.0499301515519619 , Validation Loss: 0.10879878449362722 , Validation Accuracy: 0.9668333333333333
Epoch 4 : Training Loss: 0.021326759830117226 , Validation Loss: 0.12004944905415772 , Validation Accuracy: 0.96775
Epoch 5 : Training Loss: 0.13232024013996124 , Validation Loss: 0.10849056286771999 , Validation Accuracy: 0.969

Part E

Visualized correct and Incorrect predictions along with Loss-Epoch and Accuracy-Epoch graphs for both training and validation.





Question 2.

Part A

Preprocessed & visualized the data. Created train, val, and test splits.

Part B & C

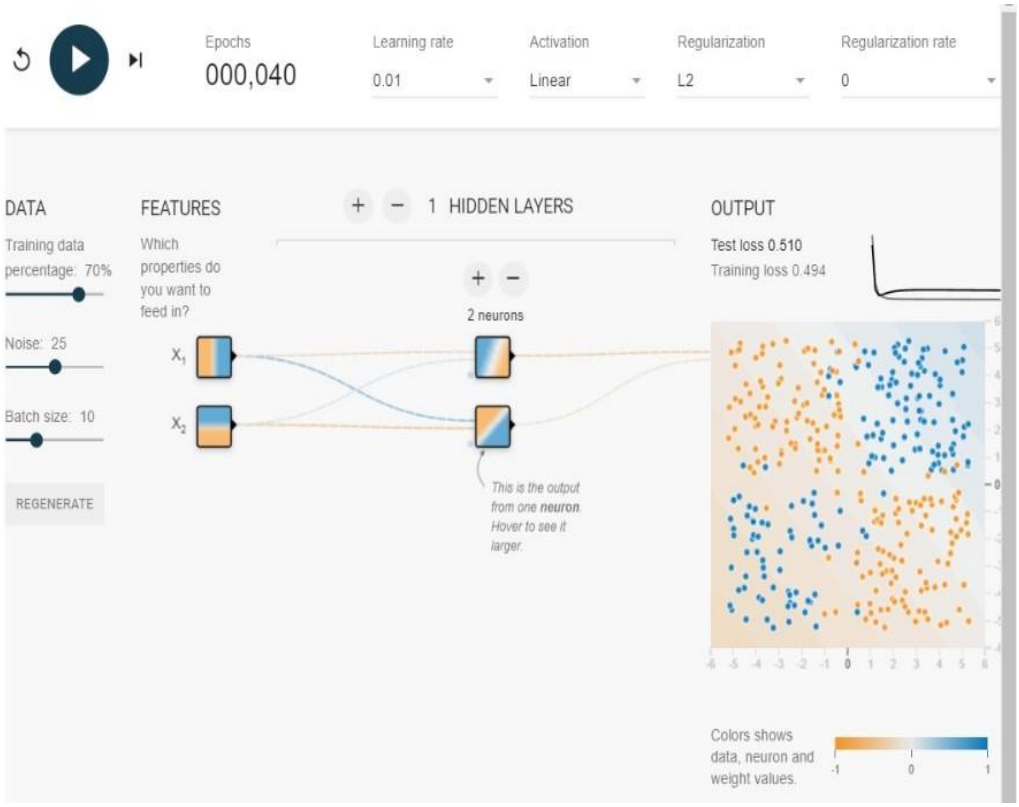
Implemented a multi-layer perceptron from scratch. Which includes the following:

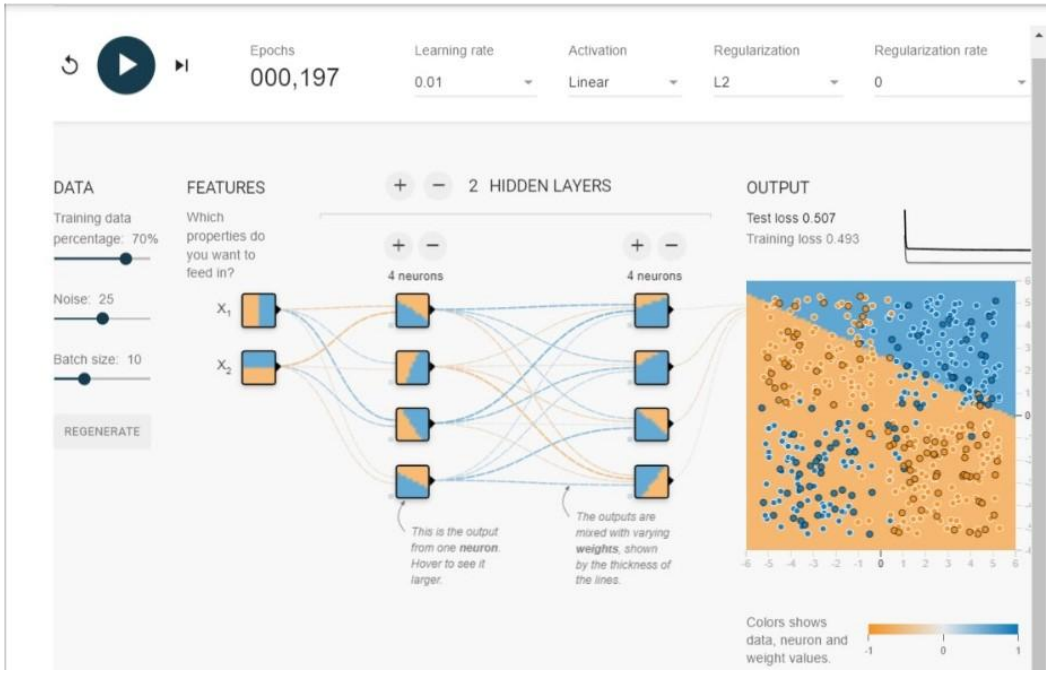
a. Activation functions.
b. Forward propagate the input.
c. Back propagate the error.
d. Train the network using stochastic gradient descent.
e. Predict the output for a given test sample and compute the accuracy.

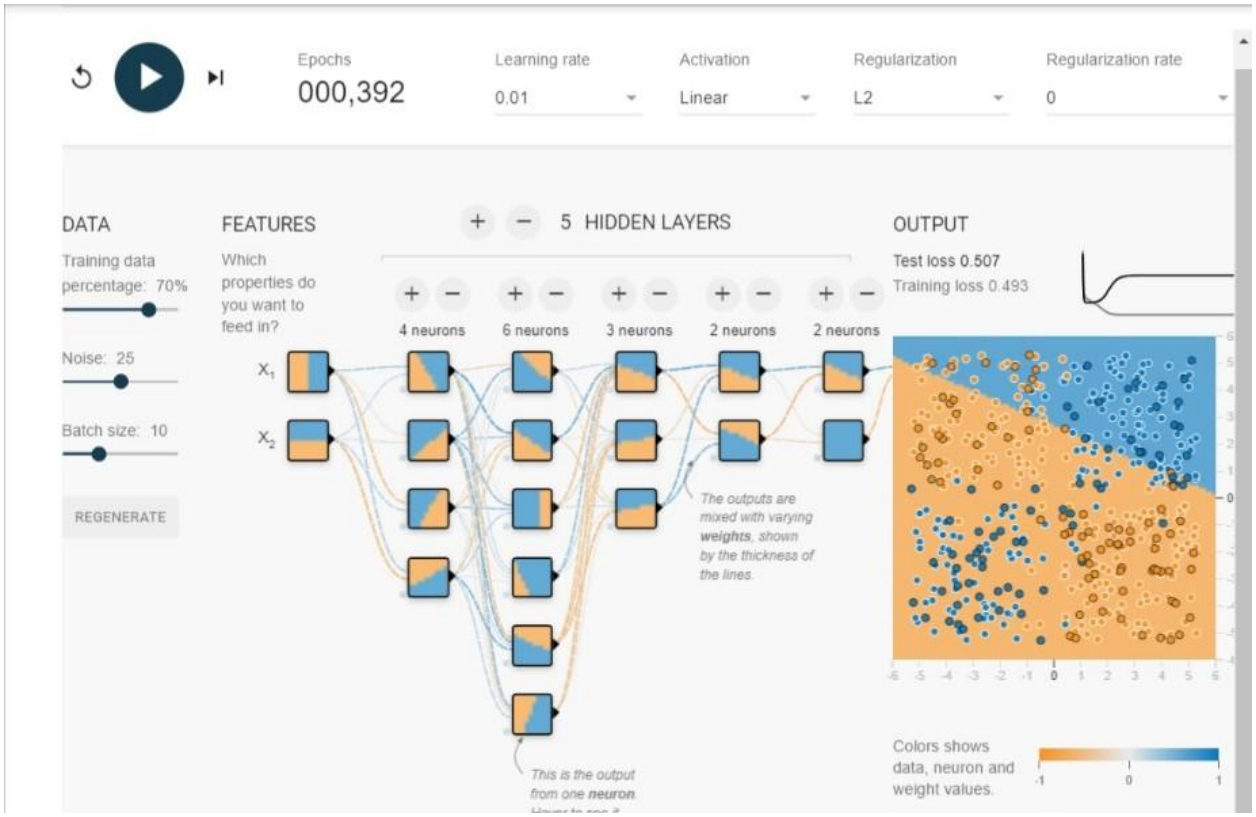
Question 3.

Part A

a.

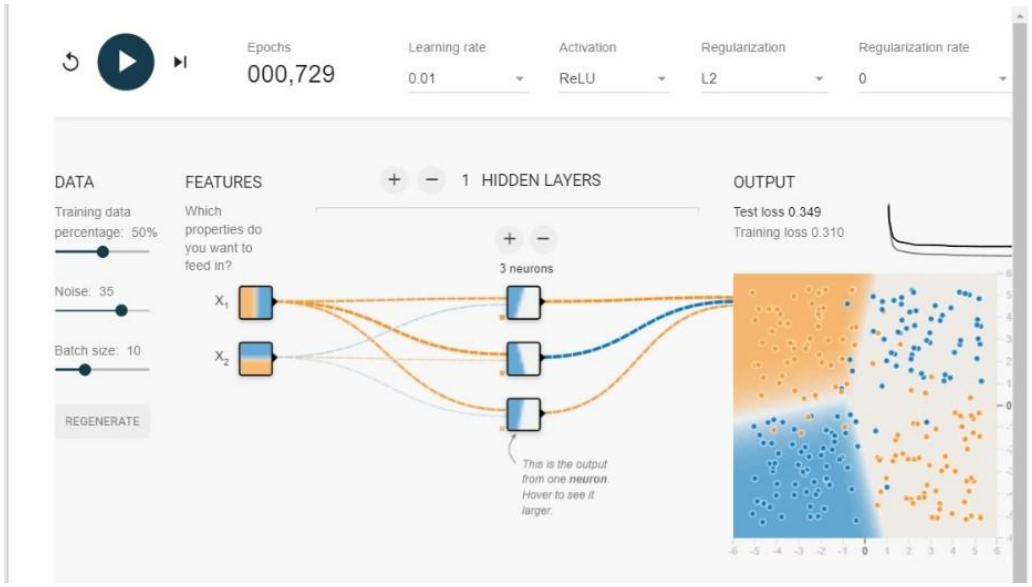


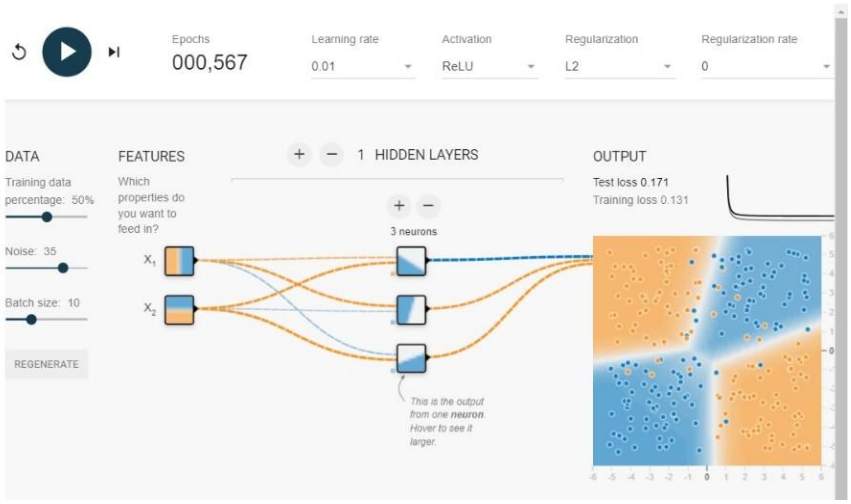


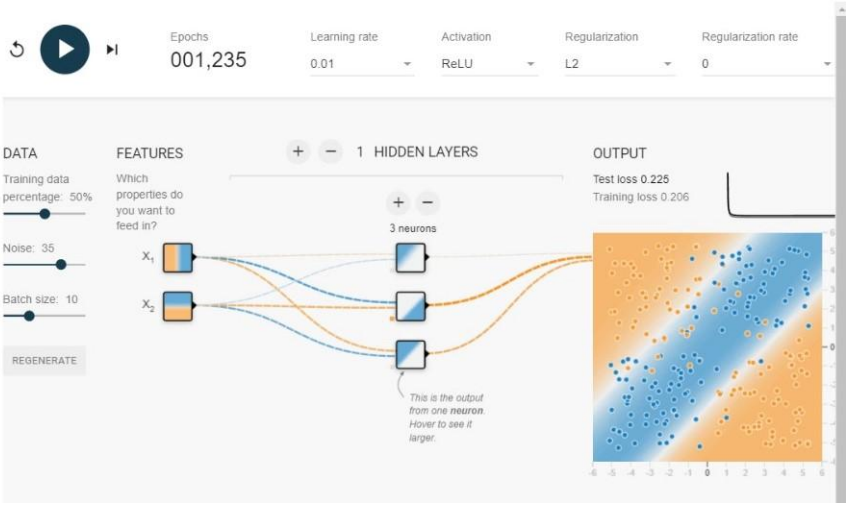


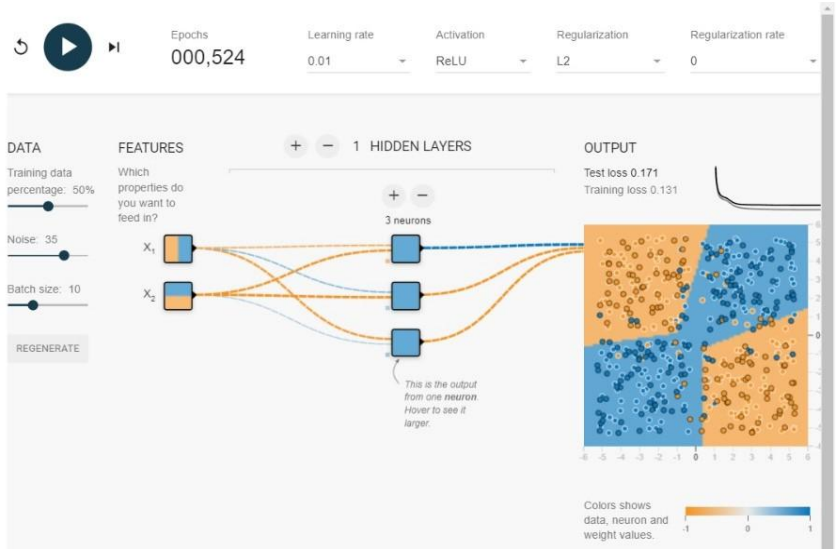
We are able to observe that as the size of the model increases, the model becomes more complex, and the amount of time it takes to run the model also increases; however, the loss does not decrease, and the boundary does not change significantly. In addition, the amount of time required to converge varies according to the size.

b.



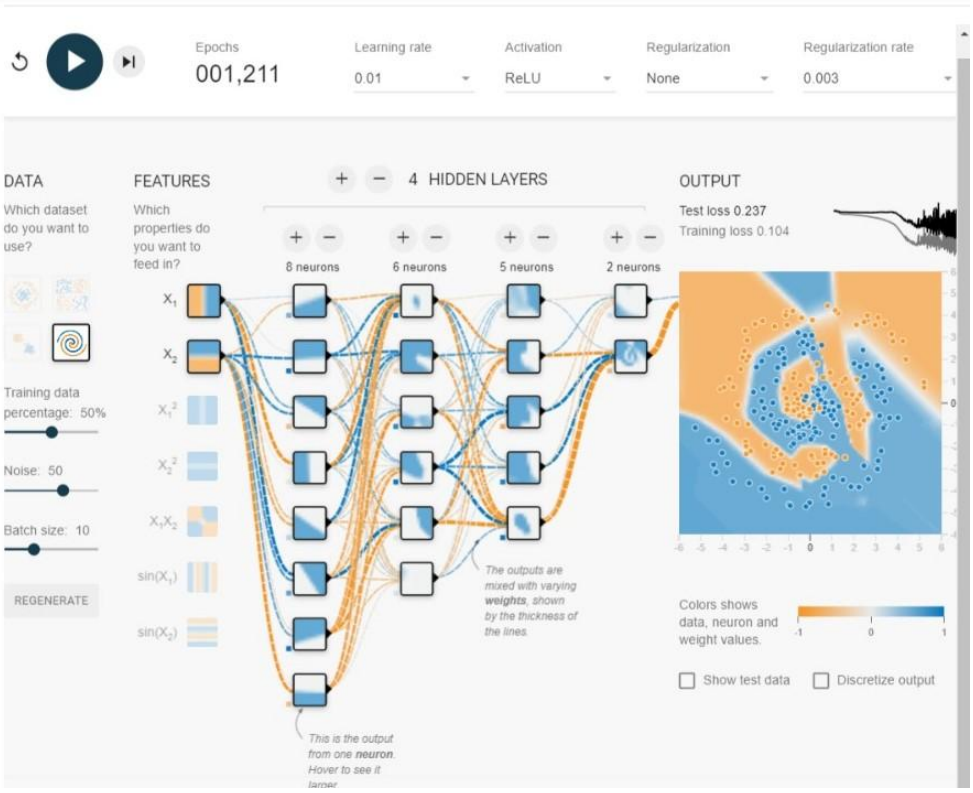

Epochs
000,567
Learning rate
0.01
Activation
ReLU
Regularization
L2
Regularization rate
0
DATA
Training data percentage: 50%
Noise: 35
Batch size: 10
REGENERATE
FEATURES
Which properties do you want to feed in?
1 HIDDEN LAYERS
3 neurons
This is the output from one neuron. Hover to see it larger.
OUTPUT
Test loss 0.171
Training loss 0.131

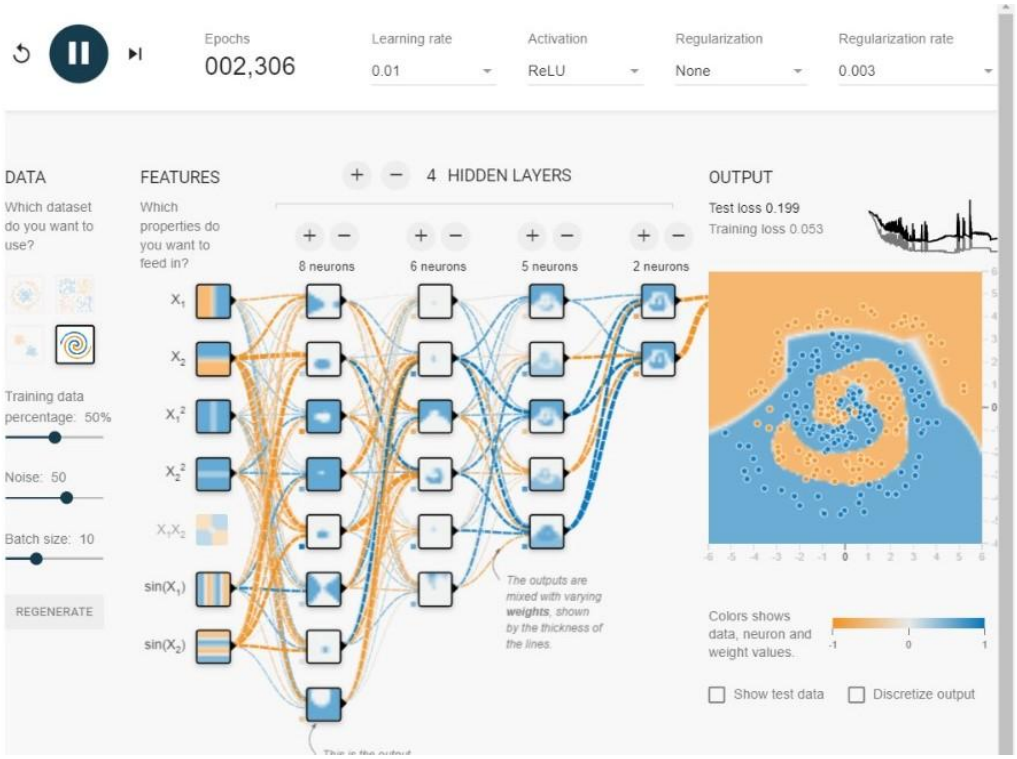

Epochs
001,235
Learning rate
0.01
Activation
ReLU
Regularization
L2
Regularization rate
0
DATA
Training data percentage: 50%
Noise: 35
Batch size: 10
REGENERATE
FEATURES
Which properties do you want to feed in?
1 HIDDEN LAYERS
3 neurons
This is the output from one neuron. Hover to see it larger.
OUTPUT
Test loss 0.225
Training loss 0.206

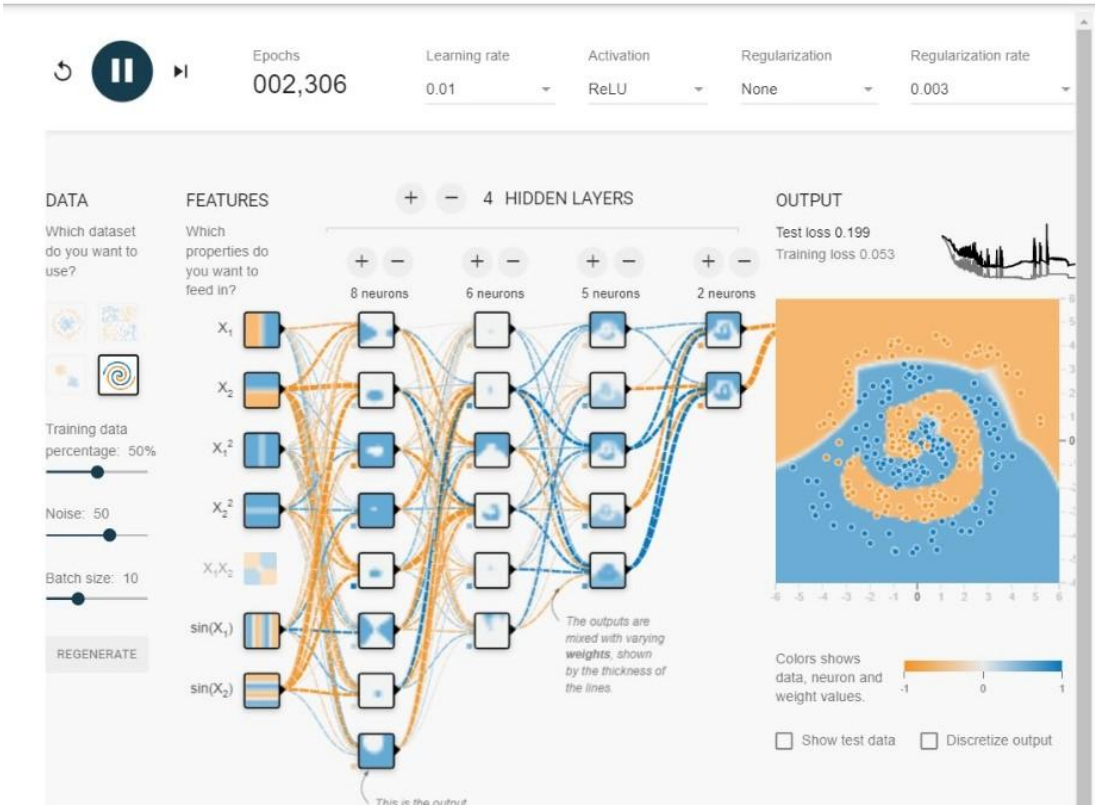

Epochs
000,524
Learning rate
0.01
Activation
ReLU
Regularization
L2
Regularization rate
0
DATA
Training data percentage: 50%
Noise: 35
Batch size: 10
REGENERATE
FEATURES
Which properties do you want to feed in?
1 HIDDEN LAYERS
3 neurons
This is the output from one neuron. Hover to see it larger.
OUTPUT
Test loss 0.171
Training loss 0.131
Colors shows data, neuron and weight values.

Each run produces a distinctively different shape, and the amount of test loss also varies significantly.

C.

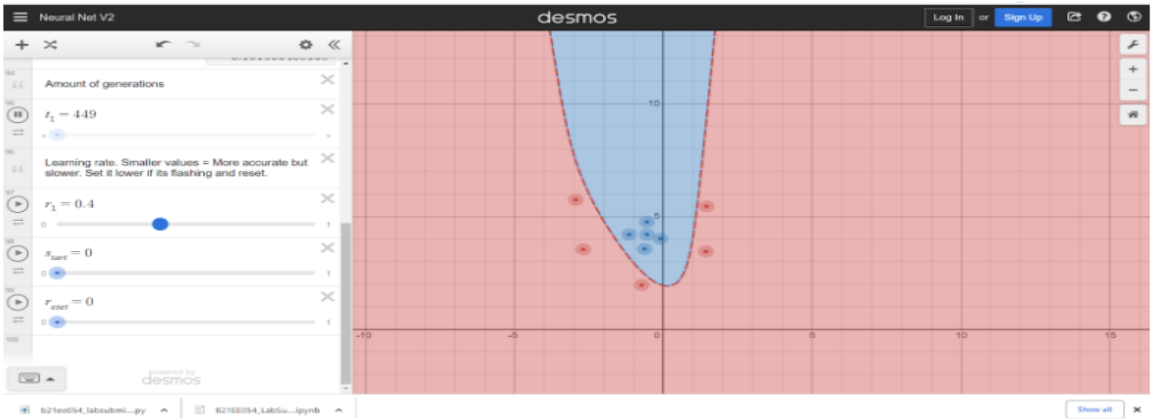


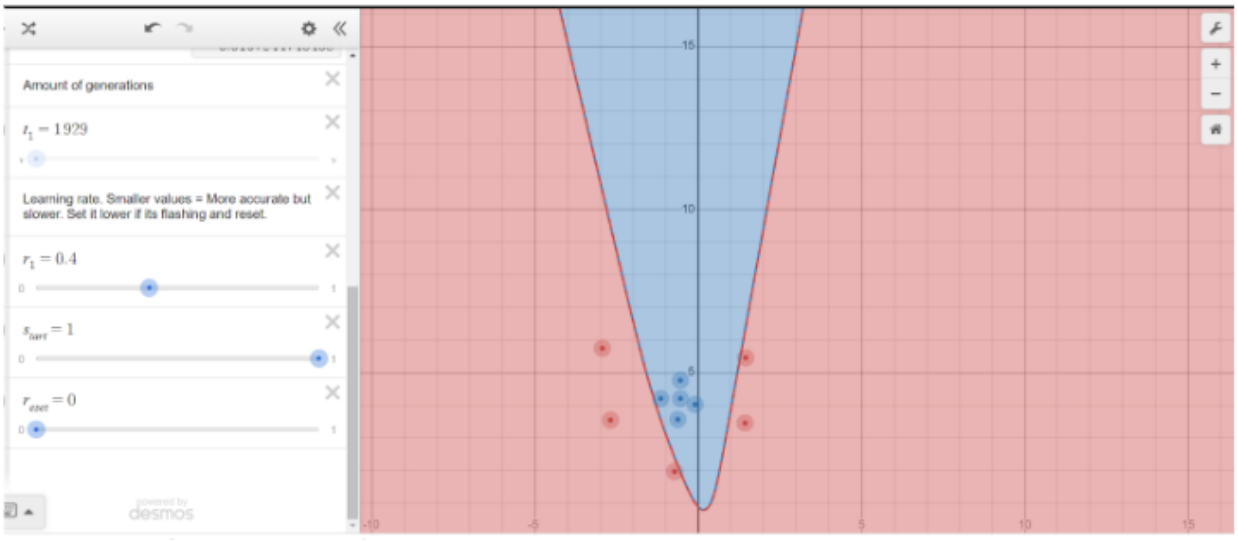




It is clear that the amount of loss is not particularly low, and that the boundaries are also not particularly smooth. It is the best possible model that can be obtained with the conditions that are given.
As the number of features is increased, the overall test loss is decreased, and the boundaries also become more smooth.

Part B





With a learning rate of 0.4, it reaches a good boundary earlier but then begins to diverge as the final output.
Despite the fact that they were too time-consuming to run for a lower learning rate, they provided a superior final boundary.