Homework 02-1: due 2022/03/31 23:59 (30%)

- In this part, you should calculate the forward pass and backpropagation manually and there is no need for any coding.

- Please scan your hand-writting calculation and save it as HW2-1.pdf

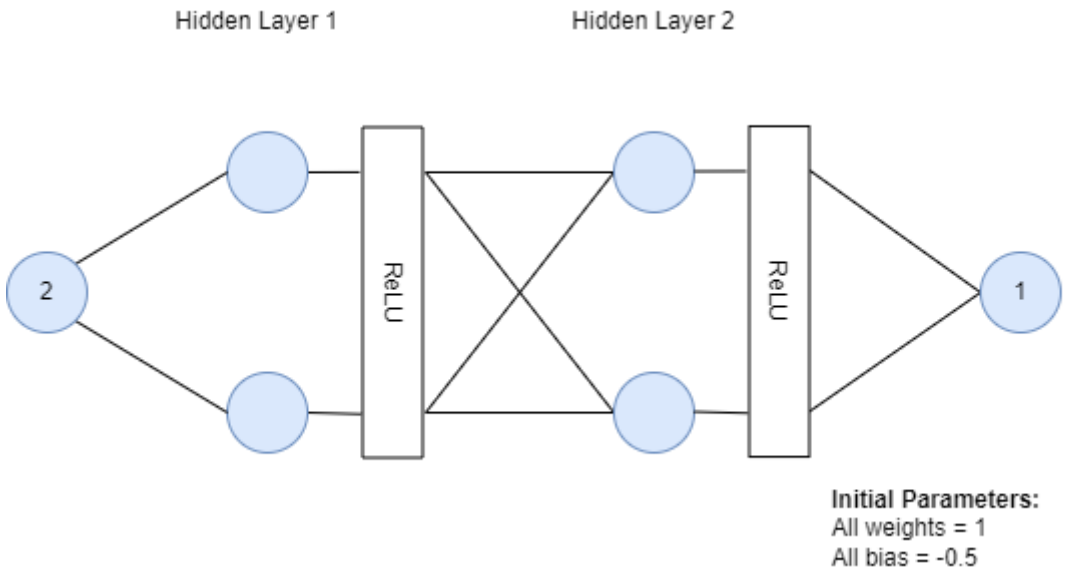
- By running the following script, you can check your answer and observe how to do the backpropagation in PyTorch.

- You can change the iterations in script to observe how will the loss change.

1. Please do the forward pass and backpropagation with a neural network as below, the input is 2 and the target is 1. Also, calculate the quadratic loss,

i.e, $$Loss = \frac{1}{2}(y-y^*)^2$$

Please update the parameters twice, i.e., do two back propagation operations, and use the learning rate 0.01.



In [1]:

```
import torch
import torch.nn as nn
import torch.optim as optim
from collections import OrderedDict
```

In [2]:

```
X = torch.tensor([2], dtype= torch.float32)
y = torch.tensor([1], dtype= torch.float32)
```

In [3]:

```
# Half of the sum square error
def loss(y, pred):
    return ((pred-y)**2).sum()/2
```

In [4]:

```
# Show parameters
def show_parameters(i, X, model):
    print(f"Iters {i}")
    print("Input:")
    print(X)
    for layer_name, layers in model.named_modules():
        print("-----------------------")
        if not isinstance(layers, nn.Sequential):
            for param_name, param in layers.named_parameters():
                print(f"{layer_name} {param_name} {param}")
                print(f"{layer_name} {param_name} Gradient")
                print(param.grad)
            print(f"{layer_name} output:")
            X = layers(X)
            print(X)

    print("================================================")
```

In [5]:

```
def initialize_weights(model):
    for name, i in model.named_modules():
        if isinstance(i, nn.Linear):
            nn.init.constant_(i.weight.data, 1)
            nn.init.constant_(i.bias.data, -0.5)
```

In [9]:

```
model = nn.Sequential(OrderedDict([("Layer1", nn.Linear(1, 2)),
                                   ("ReLU1", nn.ReLU()),
                                   ("Layer2", nn.Linear(2, 2)),
                                   ("ReLU2", nn.ReLU()),
                                   ("Layer3", nn.Linear(2, 1))]))
initialize_weights(model)
lr = 0.01
n_iters = 100
optimizer = optim.SGD(model.parameters(), lr=lr, momentum=0)
loss_list = []
for i in range(n_iters+1):
    optimizer.zero_grad()
    pred = model(X)
    l = loss(pred, y)
    loss_list.append(l.detach().numpy())
    l.backward()
    show_parameters(i, X, model)
    optimizer.step()
```

```
Iters 0
Input:
tensor([2.])
-----------------------
-----------------------
Layer1 weight Parameter containing:
tensor([[1.],
        [1.]], requires_grad=True)
Layer1 weight Gradient
tensor([[14.],
        [14.]])
Layer1 bias Parameter containing:
tensor([-0.5000, -0.5000], requires_grad=True)
Layer1 bias Gradient
tensor([7., 7.])
Layer1 output:
tensor([1.5000, 1.5000], grad_fn=<AddBackward0>)
-----------------------
ReLU1 output:
tensor([1.5000, 1.5000], grad_fn=<ReluBackward0>)
-----------------------
Layer2 weight Parameter containing:
tensor([[1., 1.],
        [1., 1.]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.2500, 5.2500],
        [5.2500, 5.2500]])
Layer2 bias Parameter containing:
tensor([-0.5000, -0.5000], requires_grad=True)
Layer2 bias Gradient
tensor([3.5000, 3.5000])
Layer2 output:
tensor([2.5000, 2.5000], grad_fn=<AddBackward0>)
-----------------------
ReLU2 output:
tensor([2.5000, 2.5000], grad_fn=<ReluBackward0>)
-----------------------
Layer3 weight Parameter containing:
tensor([[1., 1.]], requires_grad=True)
Layer3 weight Gradient
tensor([[8.7500, 8.7500]])
Layer3 bias Parameter containing:
tensor([-0.5000], requires_grad=True)
Layer3 bias Gradient
tensor([3.5000])
Layer3 output:
tensor([4.5000], grad_fn=<AddBackward0>)
================================================
Iters 1
Input:
tensor([2.])
-----------------------
-----------------------
Layer1 weight Parameter containing:
tensor([[0.8600],
        [0.8600]], requires_grad=True)
Layer1 weight Gradient
tensor([[5.0691],
        [5.0691]])
Layer1 bias Parameter containing:
tensor([-0.5700, -0.5700], requires_grad=True)
```

```
Layer1 bias Gradient
tensor([2.5346, 2.5346])
Layer1 output:
tensor([1.1500, 1.1500], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([1.1500, 1.1500], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9475, 0.9475],
        [0.9475, 0.9475]], requires_grad=True)
Layer2 weight Gradient
tensor([[1.5381, 1.5381],
        [1.5381, 1.5381]])
Layer2 bias Parameter containing:
tensor([-0.5350, -0.5350], requires_grad=True)
Layer2 bias Gradient
tensor([1.3375, 1.3375])
Layer2 output:
tensor([1.6443, 1.6443], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([1.6443, 1.6443], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.9125, 0.9125]], requires_grad=True)
Layer3 weight Gradient
tensor([[2.4101, 2.4101]])
Layer3 bias Parameter containing:
tensor([-0.5350], requires_grad=True)
Layer3 bias Gradient
tensor([1.4658])
Layer3 output:
tensor([2.4658], grad_fn=<AddBackward0>)
================================================
Iters 2
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.8093],
        [0.8093]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.8667],
        [2.8667]])
Layer1 bias Parameter containing:
tensor([-0.5953, -0.5953], requires_grad=True)
Layer1 bias Gradient
tensor([1.4333, 1.4333])
Layer1 output:
tensor([1.0233, 1.0233], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([1.0233, 1.0233], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9321, 0.9321],
        [0.9321, 0.9321]], requires_grad=True)
Layer2 weight Gradient
tensor([[0.7868, 0.7868],
```

```
        [0.7868, 0.7868]])
Layer2 bias Parameter containing:
tensor([-0.5484, -0.5484], requires_grad=True)
Layer2 bias Gradient
tensor([0.7689, 0.7689])
Layer2 output:
tensor([1.3592, 1.3592], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([1.3592, 1.3592], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8884, 0.8884]], requires_grad=True)
Layer3 weight Gradient
tensor([[1.1764, 1.1764]])
Layer3 bias Parameter containing:
tensor([-0.5497], requires_grad=True)
Layer3 bias Gradient
tensor([0.8654])
Layer3 output:
tensor([1.8654], grad_fn=<AddBackward0>)
================================================
Iters 3
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7806],
        [0.7806]], requires_grad=True)
Layer1 weight Gradient
tensor([[1.7852],
        [1.7852]])
Layer1 bias Parameter containing:
tensor([-0.6097, -0.6097], requires_grad=True)
Layer1 bias Gradient
tensor([0.8926, 0.8926])
Layer1 output:
tensor([0.9516, 0.9516], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.9516, 0.9516], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9243, 0.9243],
        [0.9243, 0.9243]], requires_grad=True)
Layer2 weight Gradient
tensor([[0.4595, 0.4595],
        [0.4595, 0.4595]])
Layer2 bias Parameter containing:
tensor([-0.5561, -0.5561], requires_grad=True)
Layer2 bias Gradient
tensor([0.4829, 0.4829])
Layer2 output:
tensor([1.2030, 1.2030], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([1.2030, 1.2030], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8766, 0.8766]], requires_grad=True)
```

```
Layer3 weight Gradient
tensor([[0.6626, 0.6626]])
Layer3 bias Parameter containing:
tensor([-0.5583], requires_grad=True)
Layer3 bias Gradient
tensor([0.5508])
Layer3 output:
tensor([1.5508], grad_fn=<AddBackward0>)
================================================
Iters 4
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7628],
        [0.7628]], requires_grad=True)
Layer1 weight Gradient
tensor([[1.1615],
        [1.1615]])
Layer1 bias Parameter containing:
tensor([-0.6186, -0.6186], requires_grad=True)
Layer1 bias Gradient
tensor([0.5808, 0.5808])
Layer1 output:
tensor([0.9070, 0.9070], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.9070, 0.9070], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9197, 0.9197],
        [0.9197, 0.9197]], requires_grad=True)
Layer2 weight Gradient
tensor([[0.2864, 0.2864],
        [0.2864, 0.2864]])
Layer2 bias Parameter containing:
tensor([-0.5609, -0.5609], requires_grad=True)
Layer2 bias Gradient
tensor([0.3158, 0.3158])
Layer2 output:
tensor([1.1073, 1.1073], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([1.1073, 1.1073], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8700, 0.8700]], requires_grad=True)
Layer3 weight Gradient
tensor([[0.4019, 0.4019]])
Layer3 bias Parameter containing:
tensor([-0.5638], requires_grad=True)
Layer3 bias Gradient
tensor([0.3629])
Layer3 output:
tensor([1.3629], grad_fn=<AddBackward0>)
================================================
Iters 5
Input:
tensor([2.])
----------------------
```

```
----------------------
Layer1 weight Parameter containing:
tensor([[0.7512],
        [0.7512]], requires_grad=True)
Layer1 weight Gradient
tensor([[0.7740],
        [0.7740]])
Layer1 bias Parameter containing:
tensor([-0.6244, -0.6244], requires_grad=True)
Layer1 bias Gradient
tensor([0.3870, 0.3870])
Layer1 output:
tensor([0.8779, 0.8779], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8779, 0.8779], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9168, 0.9168],
        [0.9168, 0.9168]], requires_grad=True)
Layer2 weight Gradient
tensor([[0.1853, 0.1853],
        [0.1853, 0.1853]])
Layer2 bias Parameter containing:
tensor([-0.5641, -0.5641], requires_grad=True)
Layer2 bias Gradient
tensor([0.2111, 0.2111])
Layer2 output:
tensor([1.0457, 1.0457], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([1.0457, 1.0457], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8660, 0.8660]], requires_grad=True)
Layer3 weight Gradient
tensor([[0.2549, 0.2549]])
Layer3 bias Parameter containing:
tensor([-0.5674], requires_grad=True)
Layer3 bias Gradient
tensor([0.2437])
Layer3 output:
tensor([1.2437], grad_fn=<AddBackward0>)
================================================
Iters 6
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7434],
        [0.7434]], requires_grad=True)
Layer1 weight Gradient
tensor([[0.5231],
        [0.5231]])
Layer1 bias Parameter containing:
tensor([-0.6283, -0.6283], requires_grad=True)
Layer1 bias Gradient
tensor([0.2616, 0.2616])
Layer1 output:
tensor([0.8586, 0.8586], grad_fn=<AddBackward0>)
```

```
----------------------
ReLU1 output:
tensor([0.8586, 0.8586], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9149, 0.9149],
        [0.9149, 0.9149]], requires_grad=True)
Layer2 weight Gradient
tensor([[0.1227, 0.1227],
        [0.1227, 0.1227]])
Layer2 bias Parameter containing:
tensor([-0.5662, -0.5662], requires_grad=True)
Layer2 bias Gradient
tensor([0.1429, 0.1429])
Layer2 output:
tensor([1.0049, 1.0049], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([1.0049, 1.0049], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8634, 0.8634]], requires_grad=True)
Layer3 weight Gradient
tensor([[0.1664, 0.1664]])
Layer3 bias Parameter containing:
tensor([-0.5699], requires_grad=True)
Layer3 bias Gradient
tensor([0.1655])
Layer3 output:
tensor([1.1655], grad_fn=<AddBackward0>)
================================================
Iters 7
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7382],
        [0.7382]], requires_grad=True)
Layer1 weight Gradient
tensor([[0.3567],
        [0.3567]])
Layer1 bias Parameter containing:
tensor([-0.6309, -0.6309], requires_grad=True)
Layer1 bias Gradient
tensor([0.1784, 0.1784])
Layer1 output:
tensor([0.8455, 0.8455], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8455, 0.8455], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9137, 0.9137],
        [0.9137, 0.9137]], requires_grad=True)
Layer2 weight Gradient
tensor([[0.0825, 0.0825],
        [0.0825, 0.0825]])
Layer2 bias Parameter containing:
tensor([-0.5676, -0.5676], requires_grad=True)
Layer2 bias Gradient
```

```
tensor([0.0976, 0.0976])
Layer2 output:
tensor([0.9775, 0.9775], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9775, 0.9775], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8618, 0.8618]], requires_grad=True)
Layer3 weight Gradient
tensor([[0.1107, 0.1107]])
Layer3 bias Parameter containing:
tensor([-0.5715], requires_grad=True)
Layer3 bias Gradient
tensor([0.1133])
Layer3 output:
tensor([1.1133], grad_fn=<AddBackward0>)
================================================
Iters 8
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7346],
        [0.7346]], requires_grad=True)
Layer1 weight Gradient
tensor([[0.2447],
        [0.2447]])
Layer1 bias Parameter containing:
tensor([-0.6327, -0.6327], requires_grad=True)
Layer1 bias Gradient
tensor([0.1223, 0.1223])
Layer1 output:
tensor([0.8366, 0.8366], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8366, 0.8366], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9129, 0.9129],
        [0.9129, 0.9129]], requires_grad=True)
Layer2 weight Gradient
tensor([[0.0561, 0.0561],
        [0.0561, 0.0561]])
Layer2 bias Parameter containing:
tensor([-0.5686, -0.5686], requires_grad=True)
Layer2 bias Gradient
tensor([0.0670, 0.0670])
Layer2 output:
tensor([0.9589, 0.9589], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9589, 0.9589], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8607, 0.8607]], requires_grad=True)
Layer3 weight Gradient
tensor([[0.0746, 0.0746]])
Layer3 bias Parameter containing:
tensor([-0.5727], requires_grad=True)
```

```
Layer3 bias Gradient
tensor([0.0778])
Layer3 output:
tensor([1.0778], grad_fn=<AddBackward0>)
================================================
Iters 9
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7322],
        [0.7322]], requires_grad=True)
Layer1 weight Gradient
tensor([[0.1685],
        [0.1685]])
Layer1 bias Parameter containing:
tensor([-0.6339, -0.6339], requires_grad=True)
Layer1 bias Gradient
tensor([0.0842, 0.0842])
Layer1 output:
tensor([0.8305, 0.8305], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8305, 0.8305], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9123, 0.9123],
        [0.9123, 0.9123]], requires_grad=True)
Layer2 weight Gradient
tensor([[0.0383, 0.0383],
        [0.0383, 0.0383]])
Layer2 bias Parameter containing:
tensor([-0.5692, -0.5692], requires_grad=True)
Layer2 bias Gradient
tensor([0.0462, 0.0462])
Layer2 output:
tensor([0.9461, 0.9461], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9461, 0.9461], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8599, 0.8599]], requires_grad=True)
Layer3 weight Gradient
tensor([[0.0508, 0.0508]])
Layer3 bias Parameter containing:
tensor([-0.5735], requires_grad=True)
Layer3 bias Gradient
tensor([0.0537])
Layer3 output:
tensor([1.0537], grad_fn=<AddBackward0>)
================================================
Iters 10
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7305],
        [0.7305]], requires_grad=True)
```

```
Layer1 weight Gradient
tensor([[0.1163],
        [0.1163]])
Layer1 bias Parameter containing:
tensor([-0.6347, -0.6347], requires_grad=True)
Layer1 bias Gradient
tensor([0.0581, 0.0581])
Layer1 output:
tensor([0.8263, 0.8263], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8263, 0.8263], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9119, 0.9119],
        [0.9119, 0.9119]], requires_grad=True)
Layer2 weight Gradient
tensor([[0.0263, 0.0263],
        [0.0263, 0.0263]])
Layer2 bias Parameter containing:
tensor([-0.5697, -0.5697], requires_grad=True)
Layer2 bias Gradient
tensor([0.0319, 0.0319])
Layer2 output:
tensor([0.9373, 0.9373], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9373, 0.9373], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8594, 0.8594]], requires_grad=True)
Layer3 weight Gradient
tensor([[0.0348, 0.0348]])
Layer3 bias Parameter containing:
tensor([-0.5740], requires_grad=True)
Layer3 bias Gradient
tensor([0.0371])
Layer3 output:
tensor([1.0371], grad_fn=<AddBackward0>)
================================================
Iters 11
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7293],
        [0.7293]], requires_grad=True)
Layer1 weight Gradient
tensor([[0.0804],
        [0.0804]])
Layer1 bias Parameter containing:
tensor([-0.6353, -0.6353], requires_grad=True)
Layer1 bias Gradient
tensor([0.0402, 0.0402])
Layer1 output:
tensor([0.8234, 0.8234], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8234, 0.8234], grad_fn=<ReluBackward0>)
----------------------
```

```
Layer2 weight Parameter containing:
tensor([[0.9117, 0.9117],
        [0.9117, 0.9117]], requires_grad=True)
Layer2 weight Gradient
tensor([[0.0182, 0.0182],
        [0.0182, 0.0182]])
Layer2 bias Parameter containing:
tensor([-0.5700, -0.5700], requires_grad=True)
Layer2 bias Gradient
tensor([0.0220, 0.0220])
Layer2 output:
tensor([0.9313, 0.9313], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9313, 0.9313], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8591, 0.8591]], requires_grad=True)
Layer3 weight Gradient
tensor([[0.0239, 0.0239]])
Layer3 bias Parameter containing:
tensor([-0.5744], requires_grad=True)
Layer3 bias Gradient
tensor([0.0257])
Layer3 output:
tensor([1.0257], grad_fn=<AddBackward0>)
================================================
Iters 12
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7285],
        [0.7285]], requires_grad=True)
Layer1 weight Gradient
tensor([[0.0557],
        [0.0557]])
Layer1 bias Parameter containing:
tensor([-0.6357, -0.6357], requires_grad=True)
Layer1 bias Gradient
tensor([0.0278, 0.0278])
Layer1 output:
tensor([0.8213, 0.8213], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8213, 0.8213], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9115, 0.9115],
        [0.9115, 0.9115]], requires_grad=True)
Layer2 weight Gradient
tensor([[0.0125, 0.0125],
        [0.0125, 0.0125]])
Layer2 bias Parameter containing:
tensor([-0.5702, -0.5702], requires_grad=True)
Layer2 bias Gradient
tensor([0.0153, 0.0153])
Layer2 output:
tensor([0.9271, 0.9271], grad_fn=<AddBackward0>)
----------------------
```

```
ReLU2 output:
tensor([0.9271, 0.9271], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8588, 0.8588]], requires_grad=True)
Layer3 weight Gradient
tensor([[0.0165, 0.0165]])
Layer3 bias Parameter containing:
tensor([-0.5746], requires_grad=True)
Layer3 bias Gradient
tensor([0.0178])
Layer3 output:
tensor([1.0178], grad_fn=<AddBackward0>)
================================================
Iters 13
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7280],
        [0.7280]], requires_grad=True)
Layer1 weight Gradient
tensor([[0.0386],
        [0.0386]])
Layer1 bias Parameter containing:
tensor([-0.6360, -0.6360], requires_grad=True)
Layer1 bias Gradient
tensor([0.0193, 0.0193])
Layer1 output:
tensor([0.8200, 0.8200], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8200, 0.8200], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9114, 0.9114],
        [0.9114, 0.9114]], requires_grad=True)
Layer2 weight Gradient
tensor([[0.0087, 0.0087],
        [0.0087, 0.0087]])
Layer2 bias Parameter containing:
tensor([-0.5704, -0.5704], requires_grad=True)
Layer2 bias Gradient
tensor([0.0106, 0.0106])
Layer2 output:
tensor([0.9242, 0.9242], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9242, 0.9242], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8587, 0.8587]], requires_grad=True)
Layer3 weight Gradient
tensor([[0.0114, 0.0114]])
Layer3 bias Parameter containing:
tensor([-0.5748], requires_grad=True)
Layer3 bias Gradient
tensor([0.0123])
Layer3 output:
tensor([1.0123], grad_fn=<AddBackward0>)
```

```
================================================
Iters 14
Input:
tensor([2.])
-----------------------
-----------------------
Layer1 weight Parameter containing:
tensor([[0.7276],
        [0.7276]], requires_grad=True)
Layer1 weight Gradient
tensor([[0.0267],
        [0.0267]])
Layer1 bias Parameter containing:
tensor([-0.6362, -0.6362], requires_grad=True)
Layer1 bias Gradient
tensor([0.0134, 0.0134])
Layer1 output:
tensor([0.8190, 0.8190], grad_fn=<AddBackward0>)
-----------------------
ReLU1 output:
tensor([0.8190, 0.8190], grad_fn=<ReluBackward0>)
-----------------------
Layer2 weight Parameter containing:
tensor([[0.9113, 0.9113],
        [0.9113, 0.9113]], requires_grad=True)
Layer2 weight Gradient
tensor([[0.0060, 0.0060],
        [0.0060, 0.0060]])
Layer2 bias Parameter containing:
tensor([-0.5705, -0.5705], requires_grad=True)
Layer2 bias Gradient
tensor([0.0073, 0.0073])
Layer2 output:
tensor([0.9222, 0.9222], grad_fn=<AddBackward0>)
-----------------------
ReLU2 output:
tensor([0.9222, 0.9222], grad_fn=<ReluBackward0>)
-----------------------
Layer3 weight Parameter containing:
tensor([[0.8586, 0.8586]], requires_grad=True)
Layer3 weight Gradient
tensor([[0.0079, 0.0079]])
Layer3 bias Parameter containing:
tensor([-0.5749], requires_grad=True)
Layer3 bias Gradient
tensor([0.0085])
Layer3 output:
tensor([1.0085], grad_fn=<AddBackward0>)
================================================
Iters 15
Input:
tensor([2.])
-----------------------
-----------------------
Layer1 weight Parameter containing:
tensor([[0.7273],
        [0.7273]], requires_grad=True)
Layer1 weight Gradient
tensor([[0.0185],
        [0.0185]])
Layer1 bias Parameter containing:
```

```
tensor([-0.6363, -0.6363], requires_grad=True)
Layer1 bias Gradient
tensor([0.0093, 0.0093])
Layer1 output:
tensor([0.8183, 0.8183], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8183, 0.8183], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9112, 0.9112],
        [0.9112, 0.9112]], requires_grad=True)
Layer2 weight Gradient
tensor([[0.0042, 0.0042],
        [0.0042, 0.0042]])
Layer2 bias Parameter containing:
tensor([-0.5706, -0.5706], requires_grad=True)
Layer2 bias Gradient
tensor([0.0051, 0.0051])
Layer2 output:
tensor([0.9208, 0.9208], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9208, 0.9208], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8585, 0.8585]], requires_grad=True)
Layer3 weight Gradient
tensor([[0.0055, 0.0055]])
Layer3 bias Parameter containing:
tensor([-0.5750], requires_grad=True)
Layer3 bias Gradient
tensor([0.0059])
Layer3 output:
tensor([1.0059], grad_fn=<AddBackward0>)
================================================
Iters 16
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7271],
        [0.7271]], requires_grad=True)
Layer1 weight Gradient
tensor([[0.0129],
        [0.0129]])
Layer1 bias Parameter containing:
tensor([-0.6364, -0.6364], requires_grad=True)
Layer1 bias Gradient
tensor([0.0064, 0.0064])
Layer1 output:
tensor([0.8179, 0.8179], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8179, 0.8179], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9112, 0.9112],
        [0.9112, 0.9112]], requires_grad=True)
Layer2 weight Gradient
```

```
tensor([[0.0029, 0.0029],
        [0.0029, 0.0029]])
Layer2 bias Parameter containing:
tensor([-0.5706, -0.5706], requires_grad=True)
Layer2 bias Gradient
tensor([0.0035, 0.0035])
Layer2 output:
tensor([0.9198, 0.9198], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9198, 0.9198], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8584, 0.8584]], requires_grad=True)
Layer3 weight Gradient
tensor([[0.0038, 0.0038]])
Layer3 bias Parameter containing:
tensor([-0.5751], requires_grad=True)
Layer3 bias Gradient
tensor([0.0041])
Layer3 output:
tensor([1.0041], grad_fn=<AddBackward0>)
==============================================
Iters 17
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7270],
        [0.7270]], requires_grad=True)
Layer1 weight Gradient
tensor([[0.0089],
        [0.0089]])
Layer1 bias Parameter containing:
tensor([-0.6365, -0.6365], requires_grad=True)
Layer1 bias Gradient
tensor([0.0045, 0.0045])
Layer1 output:
tensor([0.8175, 0.8175], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8175, 0.8175], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9112, 0.9112],
        [0.9112, 0.9112]], requires_grad=True)
Layer2 weight Gradient
tensor([[0.0020, 0.0020],
        [0.0020, 0.0020]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([0.0024, 0.0024])
Layer2 output:
tensor([0.9191, 0.9191], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9191, 0.9191], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
```

```
tensor([[0.8584, 0.8584]], requires_grad=True)
Layer3 weight Gradient
tensor([[0.0026, 0.0026]])
Layer3 bias Parameter containing:
tensor([-0.5751], requires_grad=True)
Layer3 bias Gradient
tensor([0.0029])
Layer3 output:
tensor([1.0029], grad_fn=<AddBackward0>)
================================================
Iters 18
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7269],
        [0.7269]], requires_grad=True)
Layer1 weight Gradient
tensor([[0.0062],
        [0.0062]])
Layer1 bias Parameter containing:
tensor([-0.6365, -0.6365], requires_grad=True)
Layer1 bias Gradient
tensor([0.0031, 0.0031])
Layer1 output:
tensor([0.8173, 0.8173], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8173, 0.8173], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[0.0014, 0.0014],
        [0.0014, 0.0014]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([0.0017, 0.0017])
Layer2 output:
tensor([0.9187, 0.9187], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9187, 0.9187], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8584, 0.8584]], requires_grad=True)
Layer3 weight Gradient
tensor([[0.0018, 0.0018]])
Layer3 bias Parameter containing:
tensor([-0.5751], requires_grad=True)
Layer3 bias Gradient
tensor([0.0020])
Layer3 output:
tensor([1.0020], grad_fn=<AddBackward0>)
================================================
Iters 19
Input:
tensor([2.])
```

```
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7269],
        [0.7269]], requires_grad=True)
Layer1 weight Gradient
tensor([[0.0043],
        [0.0043]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([0.0021, 0.0021])
Layer1 output:
tensor([0.8172, 0.8172], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8172, 0.8172], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[0.0010, 0.0010],
        [0.0010, 0.0010]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([0.0012, 0.0012])
Layer2 output:
tensor([0.9184, 0.9184], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9184, 0.9184], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[0.0013, 0.0013]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([0.0014])
Layer3 output:
tensor([1.0014], grad_fn=<AddBackward0>)
==============================================
Iters 20
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7268],
        [0.7268]], requires_grad=True)
Layer1 weight Gradient
tensor([[0.0030],
        [0.0030]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([0.0015, 0.0015])
Layer1 output:
```

```
tensor([0.8170, 0.8170], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8170, 0.8170], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[0.0007, 0.0007],
        [0.0007, 0.0007]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([0.0008, 0.0008])
Layer2 output:
tensor([0.9181, 0.9181], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9181, 0.9181], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[0.0009, 0.0009]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([0.0010])
Layer3 output:
tensor([1.0010], grad_fn=<AddBackward0>)
================================================
Iters 21
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7268],
        [0.7268]], requires_grad=True)
Layer1 weight Gradient
tensor([[0.0021],
        [0.0021]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([0.0010, 0.0010])
Layer1 output:
tensor([0.8170, 0.8170], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8170, 0.8170], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[0.0005, 0.0005],
        [0.0005, 0.0005]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
```

```
Layer2 bias Gradient
tensor([0.0006, 0.0006])
Layer2 output:
tensor([0.9180, 0.9180], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9180, 0.9180], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[0.0006, 0.0006]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([0.0007])
Layer3 output:
tensor([1.0007], grad_fn=<AddBackward0>)
================================================
Iters 22
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7268],
        [0.7268]], requires_grad=True)
Layer1 weight Gradient
tensor([[0.0014],
        [0.0014]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([0.0007, 0.0007])
Layer1 output:
tensor([0.8169, 0.8169], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8169, 0.8169], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[0.0003, 0.0003],
        [0.0003, 0.0003]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([0.0004, 0.0004])
Layer2 output:
tensor([0.9179, 0.9179], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9179, 0.9179], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[0.0004, 0.0004]])
Layer3 bias Parameter containing:
```

```
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([0.0005])
Layer3 output:
tensor([1.0005], grad_fn=<AddBackward0>)
================================================
Iters 23
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7268],
        [0.7268]], requires_grad=True)
Layer1 weight Gradient
tensor([[0.0010],
        [0.0010]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([0.0005, 0.0005])
Layer1 output:
tensor([0.8169, 0.8169], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8169, 0.8169], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[0.0002, 0.0002],
        [0.0002, 0.0002]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([0.0003, 0.0003])
Layer2 output:
tensor([0.9178, 0.9178], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9178, 0.9178], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[0.0003, 0.0003]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([0.0003])
Layer3 output:
tensor([1.0003], grad_fn=<AddBackward0>)
================================================
Iters 24
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
```

```
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[0.0007],
        [0.0007]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([0.0003, 0.0003])
Layer1 output:
tensor([0.8169, 0.8169], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8169, 0.8169], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[0.0002, 0.0002],
        [0.0002, 0.0002]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([0.0002, 0.0002])
Layer2 output:
tensor([0.9178, 0.9178], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9178, 0.9178], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[0.0002, 0.0002]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([0.0002])
Layer3 output:
tensor([1.0002], grad_fn=<AddBackward0>)
================================================
Iters 25
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[0.0005],
        [0.0005]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([0.0002, 0.0002])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
```

```
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[0.0001, 0.0001],
        [0.0001, 0.0001]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([0.0001, 0.0001])
Layer2 output:
tensor([0.9177, 0.9177], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9177, 0.9177], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[0.0001, 0.0001]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([0.0002])
Layer3 output:
tensor([1.0002], grad_fn=<AddBackward0>)
==============================================
Iters 26
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[0.0003],
        [0.0003]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([0.0002, 0.0002])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[7.4717e-05, 7.4717e-05],
        [7.4717e-05, 7.4717e-05]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([9.1471e-05, 9.1471e-05])
Layer2 output:
tensor([0.9177, 0.9177], grad_fn=<AddBackward0>)
```

```
----------------------
ReLU2 output:
tensor([0.9177, 0.9177], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[9.7801e-05, 9.7801e-05]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([0.0001])
Layer3 output:
tensor([1.0001], grad_fn=<AddBackward0>)
==============================================
Iters 27
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[0.0002],
        [0.0002]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([0.0001, 0.0001])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.1900e-05, 5.1900e-05],
        [5.1900e-05, 5.1900e-05]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.3539e-05, 6.3539e-05])
Layer2 output:
tensor([0.9177, 0.9177], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9177, 0.9177], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.7934e-05, 6.7934e-05]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.4029e-05])
Layer3 output:
```

```
tensor([1.0001], grad_fn=<AddBackward0>)
================================================
Iters 28
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[0.0002],
        [0.0002]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([8.0356e-05, 8.0356e-05])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[3.6021e-05, 3.6021e-05],
        [3.6021e-05, 3.6021e-05]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([4.4099e-05, 4.4099e-05])
Layer2 output:
tensor([0.9177, 0.9177], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9177, 0.9177], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[4.7149e-05, 4.7149e-05]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([5.1379e-05])
Layer3 output:
tensor([1.0001], grad_fn=<AddBackward0>)
================================================
Iters 29
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[0.0001],
        [0.0001]])
```

```
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([5.5932e-05, 5.5932e-05])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[2.5072e-05, 2.5072e-05],
        [2.5072e-05, 2.5072e-05]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([3.0695e-05, 3.0695e-05])
Layer2 output:
tensor([0.9177, 0.9177], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9177, 0.9177], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[3.2818e-05, 3.2818e-05]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([3.5763e-05])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 30
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[7.7559e-05],
        [7.7559e-05]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([3.8779e-05, 3.8779e-05])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
```

```
Layer2 weight Gradient
tensor([[1.7383e-05, 1.7383e-05],
        [1.7383e-05, 1.7383e-05]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([2.1282e-05, 2.1282e-05])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[2.2754e-05, 2.2754e-05]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([2.4796e-05])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 31
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[5.3695e-05],
        [5.3695e-05]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([2.6847e-05, 2.6847e-05])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[1.2035e-05, 1.2035e-05],
        [1.2035e-05, 1.2035e-05]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([1.4734e-05, 1.4734e-05])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
```

```
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[1.5752e-05, 1.5752e-05]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([1.7166e-05])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 32
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[3.6542e-05],
        [3.6542e-05]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.8271e-05, 1.8271e-05])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[8.1902e-06, 8.1902e-06],
        [8.1902e-06, 8.1902e-06]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([1.0027e-05, 1.0027e-05])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[1.0720e-05, 1.0720e-05]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([1.1683e-05])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 33
Input:
```

```
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.5356e-05],
        [2.5356e-05]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.2678e-05, 1.2678e-05])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.6830e-06, 5.6830e-06],
        [5.6830e-06, 5.6830e-06]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.9575e-06, 6.9575e-06])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[7.4386e-06, 7.4386e-06]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([8.1062e-06])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 34
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[1.8644e-05],
        [1.8644e-05]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([9.3220e-06, 9.3220e-06])
```

```
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[4.1787e-06, 4.1787e-06],
        [4.1787e-06, 4.1787e-06]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([5.1158e-06, 5.1158e-06])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[5.4695e-06, 5.4695e-06]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([5.9605e-06])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 35
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[1.1932e-05],
        [1.1932e-05]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([5.9661e-06, 5.9661e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[2.6743e-06, 2.6743e-06],
        [2.6743e-06, 2.6743e-06]])
Layer2 bias Parameter containing:
```

```
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([3.2741e-06, 3.2741e-06])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[3.5005e-06, 3.5005e-06]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([3.8147e-06])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 36
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[8.9491e-06],
        [8.9491e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([4.4745e-06, 4.4745e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[2.0058e-06, 2.0058e-06],
        [2.0058e-06, 2.0058e-06]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([2.4556e-06, 2.4556e-06])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[2.6254e-06, 2.6254e-06]])
```

```
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([2.8610e-06])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 37
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[5.2203e-06],
        [5.2203e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([2.6102e-06, 2.6102e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[1.1700e-06, 1.1700e-06],
        [1.1700e-06, 1.1700e-06]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([1.4324e-06, 1.4324e-06])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[1.5315e-06, 1.5315e-06]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([1.6689e-06])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 38
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
```

```
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[5.2203e-06],
        [5.2203e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([2.6102e-06, 2.6102e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[1.1700e-06, 1.1700e-06],
        [1.1700e-06, 1.1700e-06]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([1.4324e-06, 1.4324e-06])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[1.5315e-06, 1.5315e-06]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([1.6689e-06])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 39
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[3.7288e-06],
        [3.7288e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.8644e-06, 1.8644e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
```

```
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
-----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[8.3573e-07, 8.3573e-07],
        [8.3573e-07, 8.3573e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([1.0232e-06, 1.0232e-06])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
-----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
-----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[1.0939e-06, 1.0939e-06]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([1.1921e-06])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 40
Input:
tensor([2.])
-----------------------
-----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
-----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
-----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
```

```
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
-----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
-----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 41
Input:
tensor([2.])
-----------------------
-----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
-----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
-----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
-----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
-----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
```

```
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 42
Input:
tensor([2.])
-----------------------
-----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
-----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
-----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
-----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
-----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 43
Input:
tensor([2.])
-----------------------
-----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
```

```
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 44
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
```

```
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 45
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
```

```
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
==============================================
Iters 46
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
==============================================
Iters 47
```

```
Input:
tensor([2.])
-----------------------
-----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
-----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
-----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
-----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
-----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 48
Input:
tensor([2.])
-----------------------
-----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
```

```
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 49
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
```

```
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 50
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
```

```
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 51
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 52
Input:
tensor([2.])
----------------------
----------------------
```

```
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 53
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
```

```
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 54
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
```

```
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 55
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
```

```
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
==========================================
Iters 56
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
==========================================
Iters 57
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
```

```
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 58
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
```

```
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 59
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
```

```
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
-----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 60
Input:
tensor([2.])
-----------------------
-----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
-----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
-----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
-----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
-----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
```

```
Iters 61
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 62
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
```

```
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 63
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
```

```
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 64
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
```

```
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 65
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 66
Input:
tensor([2.])
----------------------
```

```
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 67
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
```

```
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 68
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
```

```
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 69
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
```

```
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 70
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 71
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
```

```
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 72
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
```

```
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 73
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
```

```
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
==============================================
Iters 74
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
```

```
==============================================
Iters 75
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
==============================================
Iters 76
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
```

```
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 77
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
```

```
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 78
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
```

```
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 79
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 80
Input:
tensor([2.])
```

```
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
==============================================
Iters 81
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
```

```
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
-----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
-----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
-----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
-----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 82
Input:
tensor([2.])
-----------------------
-----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
-----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
-----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
```

```
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 83
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
```

```
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 84
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 85
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
```

```
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 86
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
```

```
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 87
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
```

```
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
==============================================
Iters 88
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
```

```
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 89
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 90
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
```

```
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 91
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
```

```
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 92
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
```

```
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 93
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 94
Input:
```

```
tensor([2.])
-----------------------
-----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
-----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
-----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
-----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
-----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 95
Input:
tensor([2.])
-----------------------
-----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
```

```
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 96
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
```

```
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 97
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
```

```
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 98
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 99
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
```

```
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
Iters 100
Input:
tensor([2.])
----------------------
----------------------
Layer1 weight Parameter containing:
tensor([[0.7267],
        [0.7267]], requires_grad=True)
Layer1 weight Gradient
tensor([[2.2373e-06],
        [2.2373e-06]])
Layer1 bias Parameter containing:
tensor([-0.6366, -0.6366], requires_grad=True)
Layer1 bias Gradient
tensor([1.1186e-06, 1.1186e-06])
Layer1 output:
tensor([0.8168, 0.8168], grad_fn=<AddBackward0>)
----------------------
ReLU1 output:
```

```
tensor([0.8168, 0.8168], grad_fn=<ReluBackward0>)
-----------------------
Layer2 weight Parameter containing:
tensor([[0.9111, 0.9111],
        [0.9111, 0.9111]], requires_grad=True)
Layer2 weight Gradient
tensor([[5.0144e-07, 5.0144e-07],
        [5.0144e-07, 5.0144e-07]])
Layer2 bias Parameter containing:
tensor([-0.5707, -0.5707], requires_grad=True)
Layer2 bias Gradient
tensor([6.1390e-07, 6.1390e-07])
Layer2 output:
tensor([0.9176, 0.9176], grad_fn=<AddBackward0>)
-----------------------
ReLU2 output:
tensor([0.9176, 0.9176], grad_fn=<ReluBackward0>)
-----------------------
Layer3 weight Parameter containing:
tensor([[0.8583, 0.8583]], requires_grad=True)
Layer3 weight Gradient
tensor([[6.5634e-07, 6.5634e-07]])
Layer3 bias Parameter containing:
tensor([-0.5752], requires_grad=True)
Layer3 bias Gradient
tensor([7.1526e-07])
Layer3 output:
tensor([1.0000], grad_fn=<AddBackward0>)
================================================
```

In [ ]: