

Name- Chitrang Gupta

Roll No- 170050024

### **Xor Data set**

Hyperparameters-random seed=42,lr=1e-3,batchsize=100,epochs=30

Test accuracy=95.7%

Architecture-

- 1)One input Layer (of size 2),
- 2) Fully connected Hidden Layer with Relu activation and 4 nodes.
- 3)One Output Layer Fullyconnected with Two nodes and Softmax activation.

Total number of Nodes(excluding input and output nodes)=4

Total Number of Nodes=4+4=8

Total Number of Parameters=2\*4+4\*2=16

Why is this minimal Topology?

Experimental Answers- Logistic Regression can separate (well) only linearly separable data and hence one non-linear hidden layer is necessary. With 3 nodes in the hidden Layers(and relu activation), I get accuracy 65% even after 100 epochs. With 4 nodes I get well above 90%.

### **Semicircle Data set**

Hyperparameters-random seed=42,lr=1e-3,batchsize=100,epochs=30

Test accuracy=96.8%

Architecture-

- 1)One input Layer (of size 2),
- 2) Fully connected Hidden Layer with Relu activation and 2 nodes.
- 3)One Output Layer Fullyconnected with Two nodes and Softmax activation.

Total number of Nodes(excluding input and output nodes)=2

Total Number of Nodes=2+4=6

Total Number of Parameters=2\*2+2\*2=8

Why is this minimal Topology?

Experimental Answers- Logistic Regression can separate (well) only linearly separable data and hence one non-linear hidden layer is necessary. With 1 nodes in the hidden layer I get accuracy as 78%. With 2 nodes, I get accuracy well above 90%.

### **Mnist Data set:**

Hyperparameters-random seed=42,lr=1e-3,batchsize=100,epochs=10

Test accuracy=91.8%

Architecture-

- 1)One input Layer (of size 784),
- 2)One Output Layer with 10 nodes and Softmax activation.

Total number of Nodes(excluding input and output nodes)=0

Total Number of Nodes=784+10=794

Total Number of Parameters=794\*10=7940

Why is this minimal Topology?

Experimental Answers- We cannot have a topology smaller than this.

**Cifar10 Data set:**

Hyperparameters-random seed=42,lr=1e-3,batchsize=100,epochs=33

Test accuracy=39.1%

Architecture-

- 1)One input Layer (of size 3,32,32),
- 2)One Conv Layer of with numfiltera=8, filter size=(12,12), and stride=4 and “relu activation”.
- 3)One AvgPool Layer of with filter size=(2,2), and stride=2.
- 4)One Flatten Layer.
- 6)One Fullyconnected Layer of with outnodes=10 and activation=softmax

Total number of Nodes(excluding input and output nodes)= $8*6*6+72=360$

Total Number of Parameters= $8*12*12+8*2*2+72*10=1904$

Why is this minimal Topology?

Experimental Answers- Experimentally I found out that Using a convolution, avgpooling, flatten and a fullyconnected layer give accuracy above 35%. I tried to use as many as possible number of feature maps to capture as much features of the images as possible in the convolution layer while keeping the size of output image low. This was ensured by using large filter maps (but not so large so as to increase the parameters) and stride somewhat half the size of the filter. And then a Fully connected Layer was necessary to calculate the probabilities of the 10 classes.