**Hypothesis: Deep network represents the features better.**

1. Analysis of different division of number of nodes in the hidden layer:

**For two hidden layer model:**

**Model trained on various datasets and concluded from the result that division of number of nodes in hidden layers like this [60,40]% i.e., 60% nodes in first layer and 40% nodes in second layer achieves better accuracy.**

* Model trained on data shell radius (1,1.5) dimensions 128

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 256 total nodes | [10,90]% | [20,80]% | [30,70]% | [40,60]% | [50,50]% | [60,40]% | [70,30]% | [80,20]% | [90,10]% |
| train (700) | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| test(200) | 77 | 85.5 | 89 | 90.5 | 89.5 | 90.5 | 93 | 93 | 89.5 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 512 total nodes | [10,90]% | [20,80]% | [30,70]% | [40,60]% | [50,50]% | [60,40]% | [70,30]% | [80,20]% | [90,10]% |
| train(700) | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| test(200) | 83.5 | 87 | 92 | 94.5 | 94 | 95.5 | 94.5 | 95 | 95 |

* Model trained on data shell radius (1,1.1) dimensions 128

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 512 total nodes | [10,90]% | [20,80]% | [30,70]% | [40,60]% | [50,50]% | [60,40]% | [70,30]% | [80,20]% | [90,10]% |
| train (700) | 100 | 100 | 100 | 100 | 100 | 99.86 | 99.86 | 100 | 100 |
| test(200) | 58 | 57.5 | 54.00 | 56 | 53.5 | 57.5 | 53.5 | 56 | 54.5 |

**For three hidden layer model:**

**Model trained on various datasets and concluded from the result that division of number of nodes in hidden layers like this [60,30,10]% i.e., 60% nodes in first layer and 30% nodes in second layer and 10% nodes in third layer achieves better accuracy.**

|  |  |
| --- | --- |
| **Source Domain:**  Training = 700  Testing = 200 | **Target Domain:**  Training = 200  Testing = 200 |

Source:

(1,1.5) 128 dim 512 nodes

tensor(97.5714, device='cuda:0') tensor(85.5000, device='cuda:0')

tensor(100., device='cuda:0') tensor(95.5000, device='cuda:0')

tensor(100., device='cuda:0') tensor(100., device='cuda:0')

Target:

(1,r) 128 dim

1. **Decision region plots by model trained on K-fold shell data**

[0.4, 0.8] radius

48 total hidden nodes

Number of hidden layers 1 : 100%

Number of hidden layers 2 : 60%,40%

Number of hidden layers 3 : 50%,30%,20%

**0 => output layer node 1, 1 => output layer node 2, -1 => final output layer(after softmax) result**

|  |  |  |
| --- | --- | --- |
| Number of hidden layers = 1 | Number of hidden layers = 2 | Number of hidden layers = 3 |
|  |  |  |
|  |  |  |
|  |  |  |

|  |  |  |
| --- | --- | --- |
| Number of hidden layers = 1 | Number of hidden layers = 2 | Number of hidden layers = 3 |
|  |  |  |
|  |  |  |
|  |  |  |

|  |  |  |
| --- | --- | --- |
| Number of hidden layers = 1 | Number of hidden layers = 2 | Number of hidden layers = 3 |
|  |  |  |
|  |  |  |
|  |  |  |

|  |  |  |
| --- | --- | --- |
| Number of hidden layers = 1 | Number of hidden layers = 2 | Number of hidden layers = 3 |
|  |  |  |
|  |  |  |
|  |  |  |