Part 1

1、Analysis of the number of hidden layers

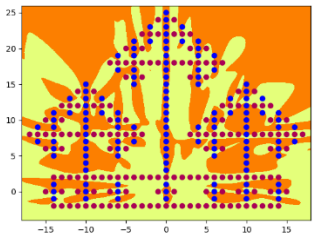
For the full3 network, when the number of training rounds is 10000, the following table shows the model accuracy, parameter quantity and running time under different hidden layer network parameters

	25	50	60	80
Accuracy	92.47	95.21	100	100
Param	751	2751	3901	6801
Time	3.25	3.34	3.41	3.43

Among them, the parameter calculation formula is:

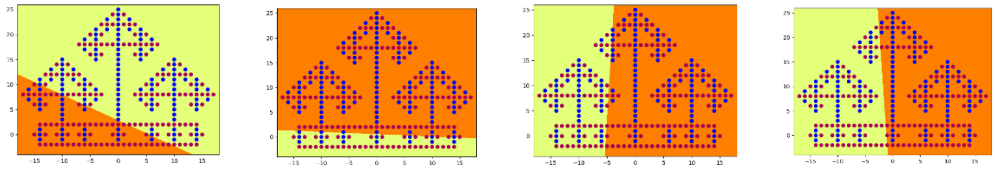
Param = (2+1) * hidden_size + (hidden_size+1)*hidden_size + (hidden_size+1)*1

Running time represents the time it took the model to run for 1000 epochs. (Use time.time() for timing, the running time of different hardware will be different). We found that with 60 neurons, almost all classifications were correct.

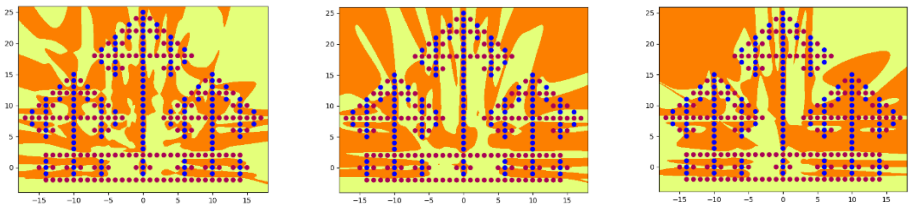


out_full3_60.png

The figure above shows the output of the network for different samples. The dynamics of each branch is shown in the figure below, which together form the final output result (there are 60 output results in 60, and only 4 are selected here for illustration):



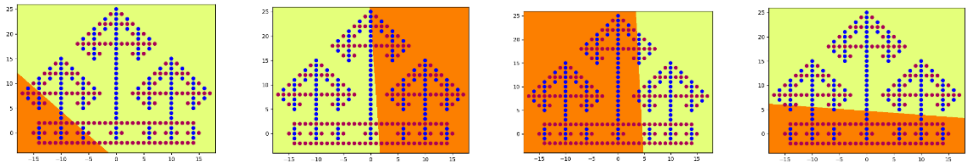
Partially hidden layer dynamics



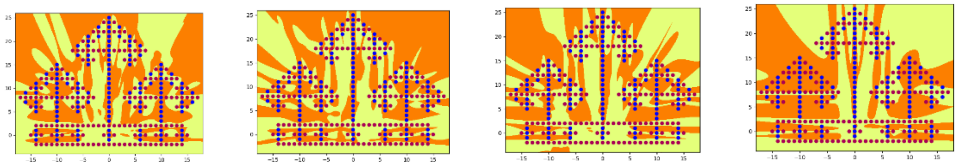
Decision boundary graphs for different hidden numbers. Left: 25 Middle: 50 Right: 80

For the full4 network

	25	50	60	80
Accurary	85.78	99.66	100	100
Param	1401	5301	7561	13281
Time	5.00	5.11	5.23	5.21



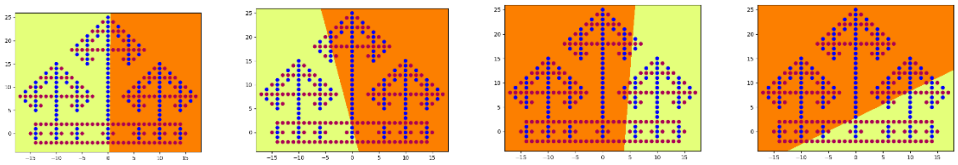
Partially hidden layer dynamics



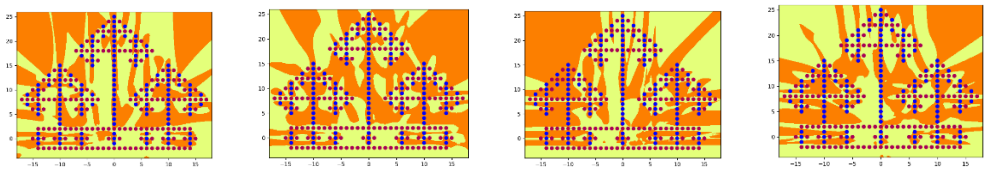
Decision boundary graphs for different hidden numbers.
From left to right: 25 50 60 80

For the dense network：

	25	50	60	80
Accuracy	93.84	100	100	100
Param	1453	5403	7683	13443
Time	5.12	5.42	5.82	6.03



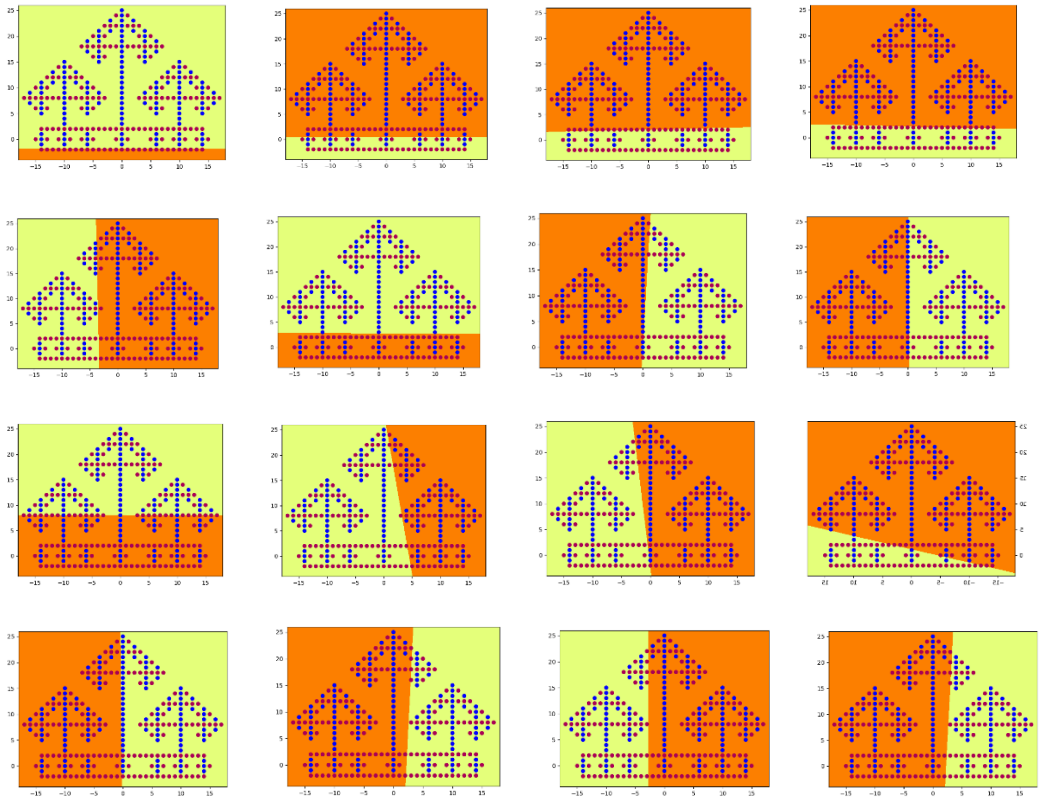
Partially hidden layer dynamics



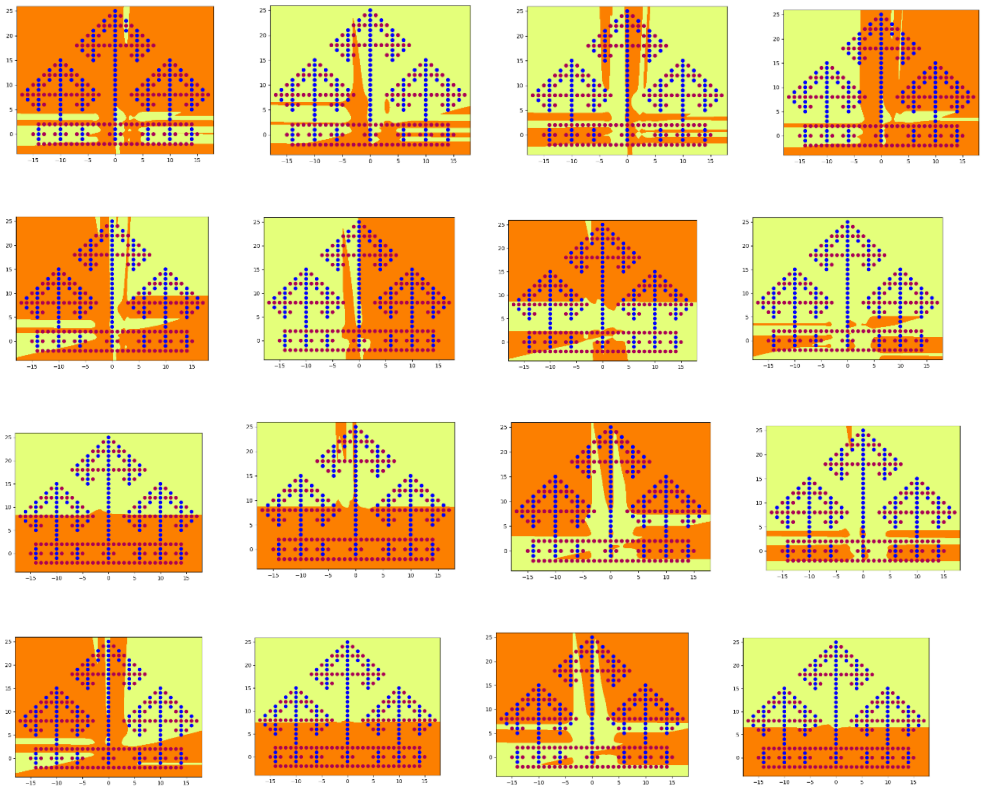
Decision boundary graphs for different hidden numbers.
From left to right: 25 50 60 80

2 Additional dynamic visualization of hidden layers

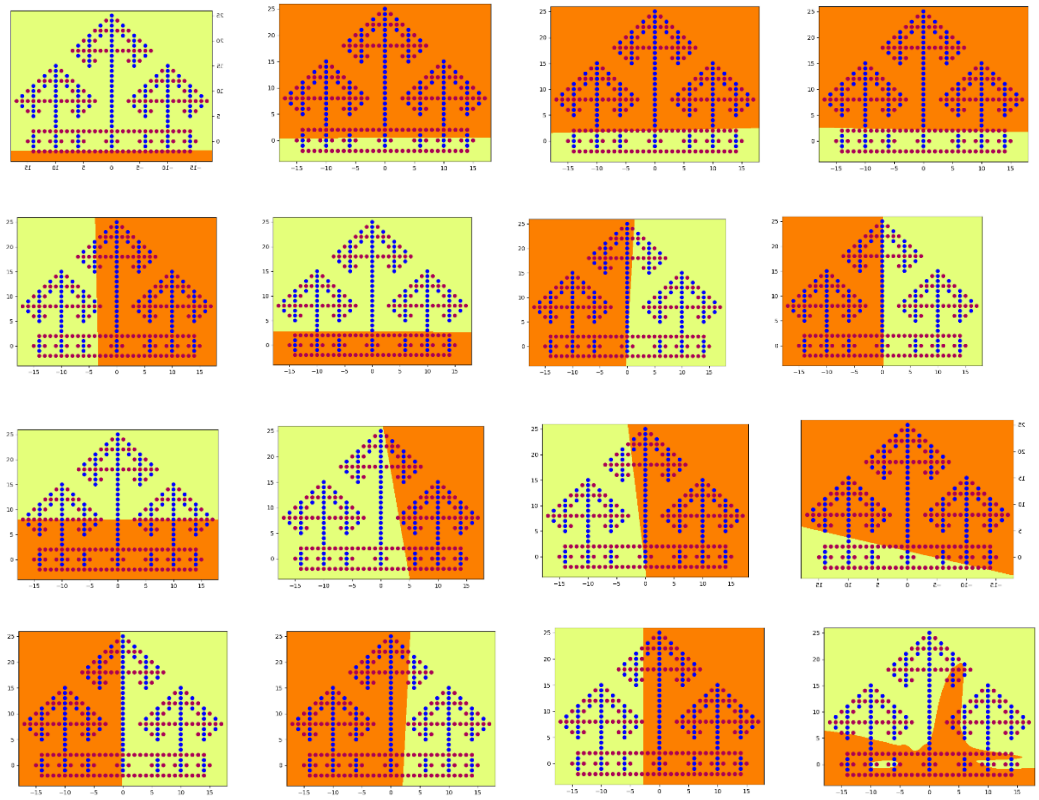
Full3 Net Hidden 1



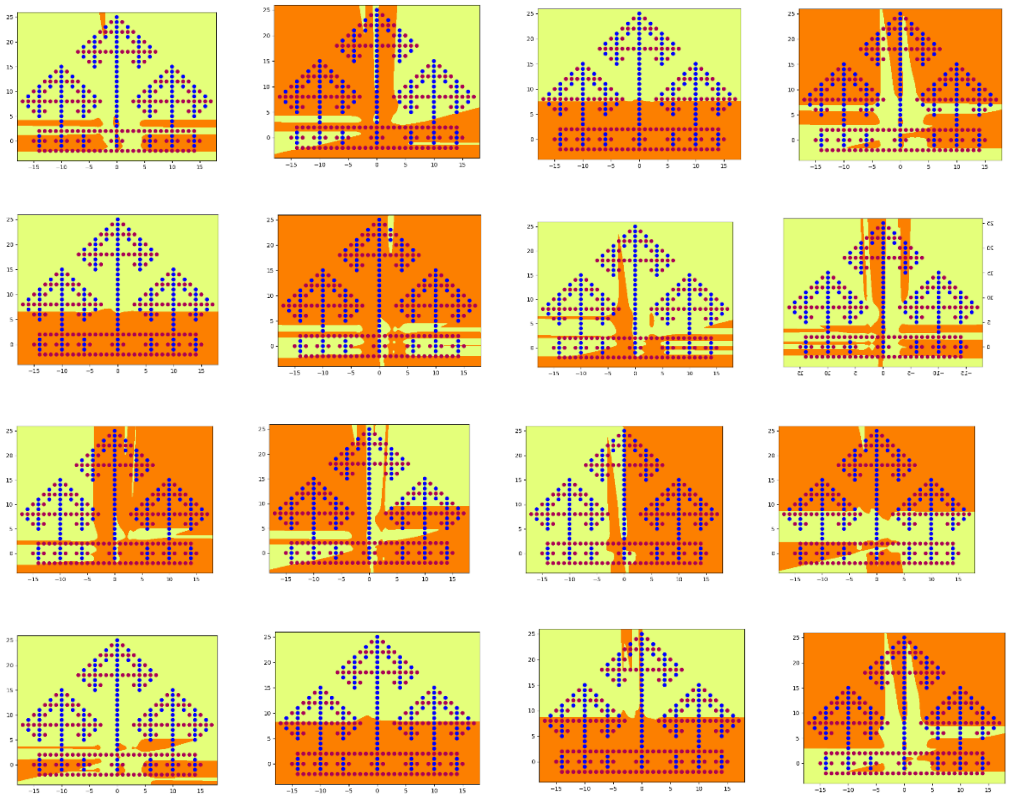
Full3 Net Hidden 2



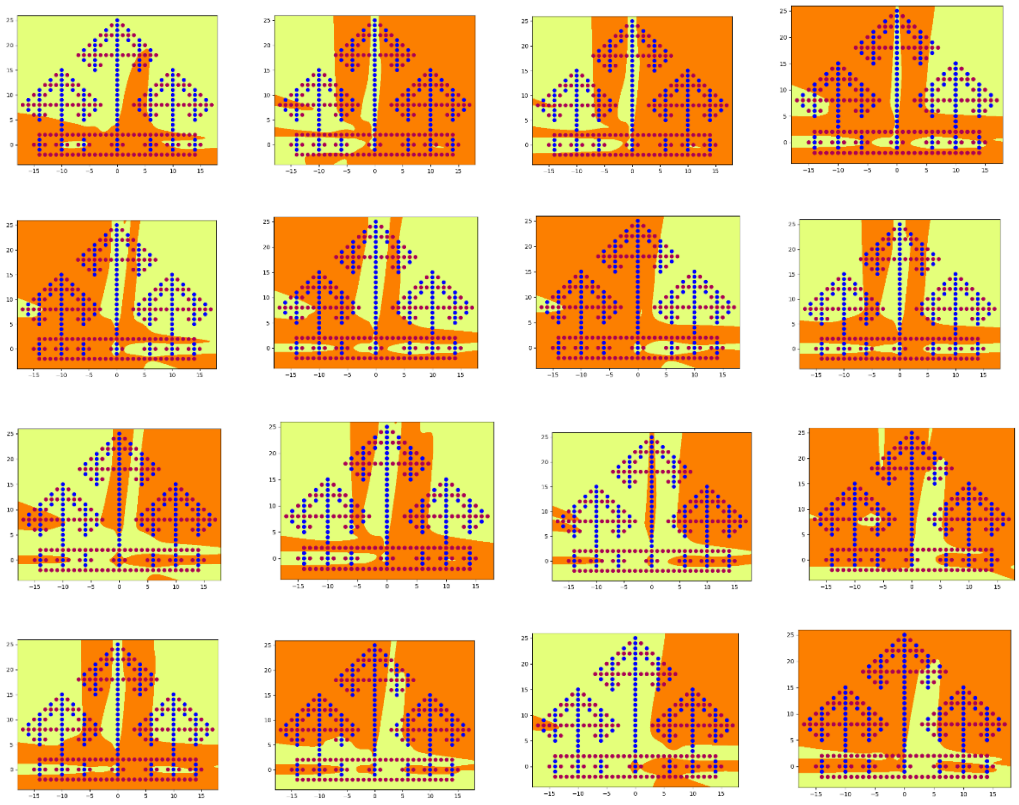
Full4 Net Hidden 1



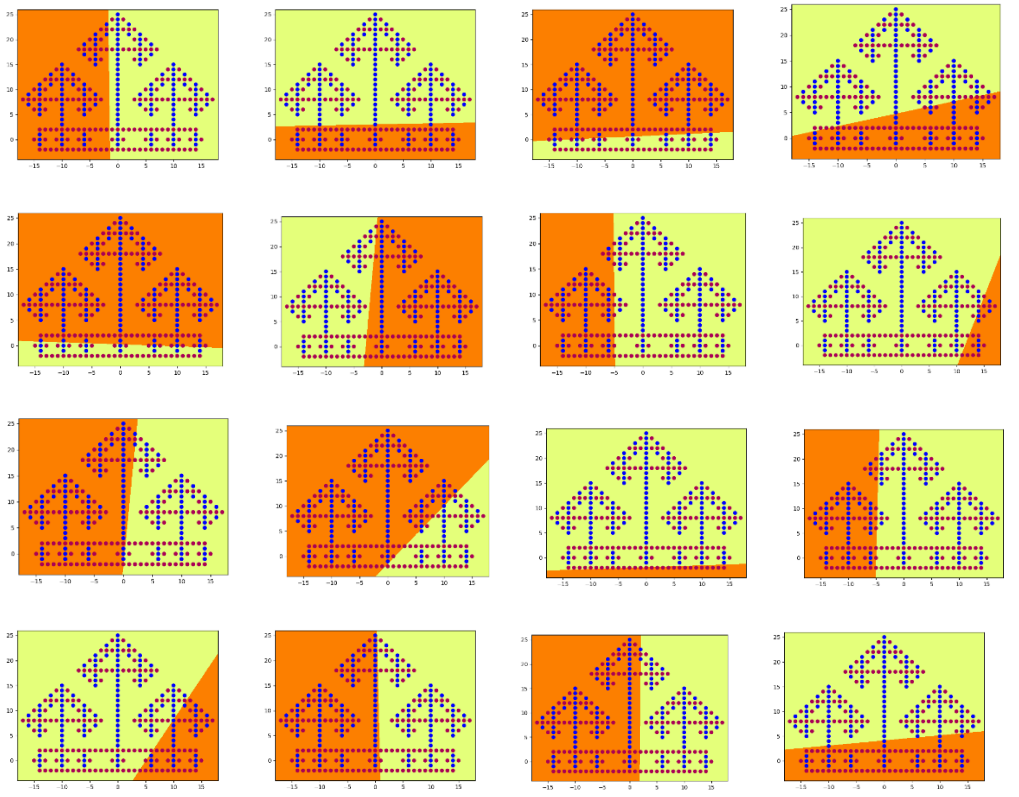
Full4 Net Hidden 2



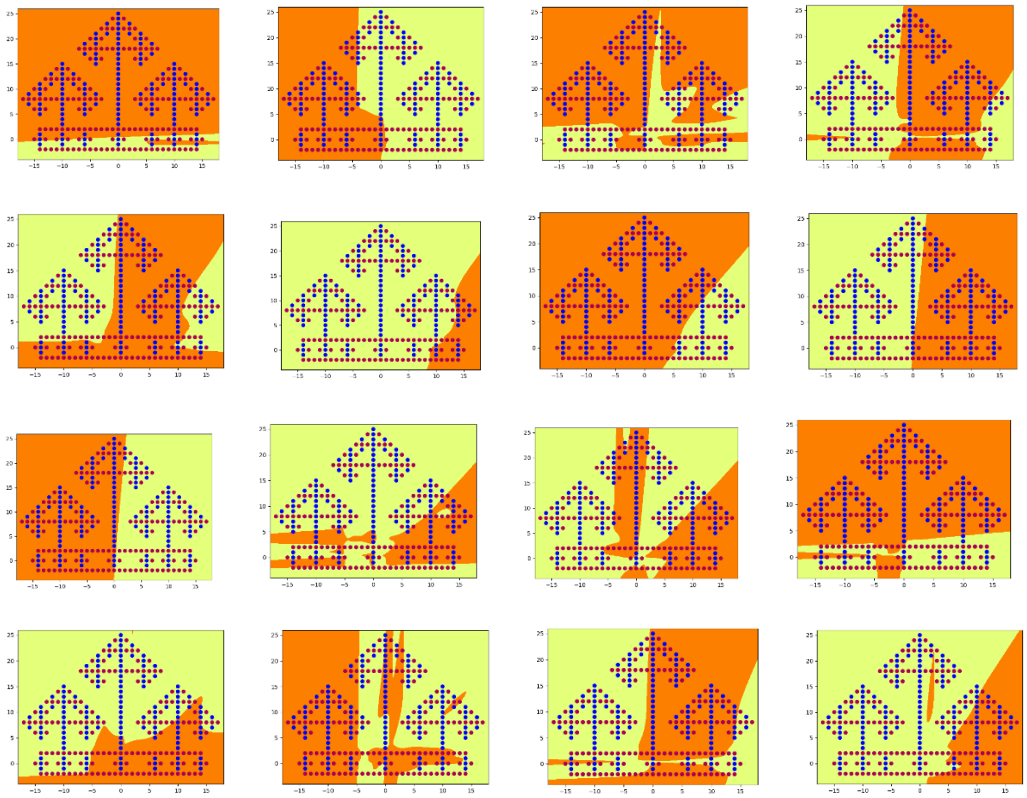
Full4 Net Hidden3



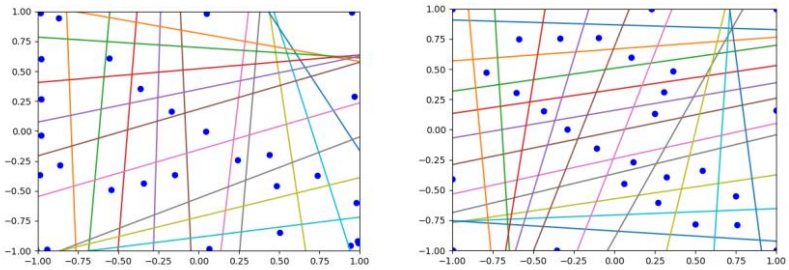
DenseNet Hidden 1



DenseNet Hidden 2



Part 2

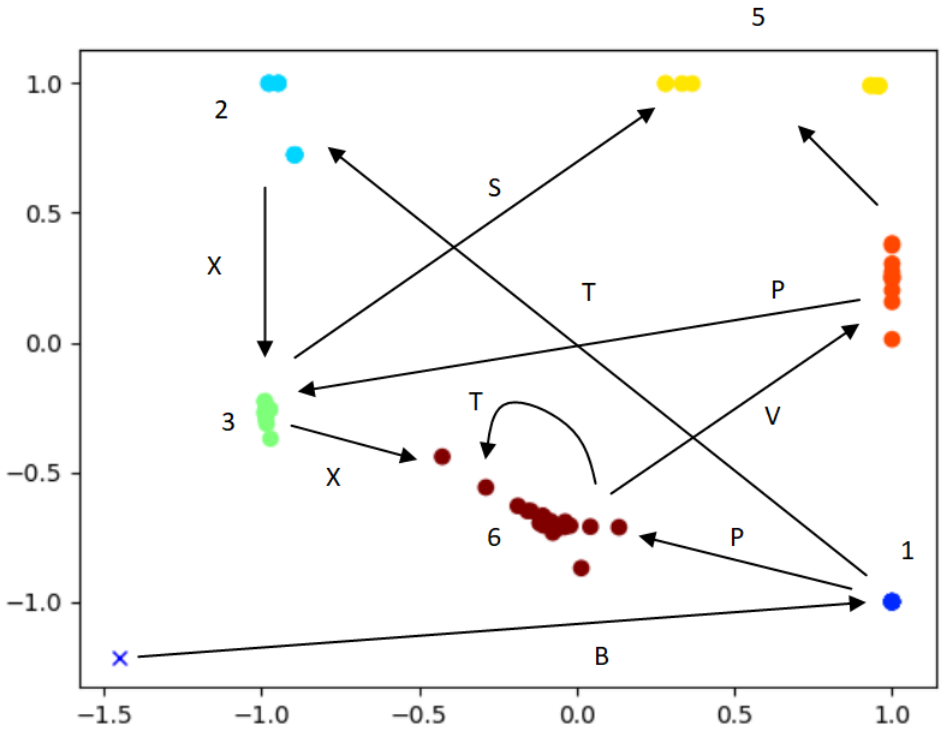


Left: 1000 times training process diagram Right: final result diagram

Part 3

1. Results of the training process:

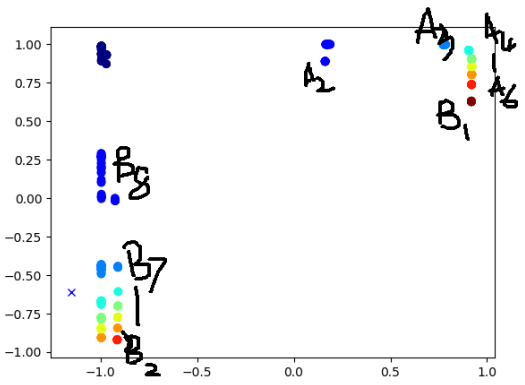
```
hidden activations and output probabilities [BTSXPVE]:
1 [ 1.   -0.98 -0.95  0.17] [ 0.    0.47  0.    0.    0.53  0.    0.  ]
2 [-0.99  0.23  0.93 -0.75] [ 0.    0.    0.43  0.56  0.    0.    0.  ]
3 [-0.99  0.92  0.51 -1.  ] [ 0.    0.    0.66  0.34  0.    0.    0.  ]
6 [-0.99 -0.66 -0.93 -0.98] [ 0.    0.45  0.    0.    0.    0.55  0.  ]
5 [-1.   -1.   -0.99  0.99] [ 0.    0.    0.    0.    0.4   0.59  0.  ]
3 [-1.    0.92  0.51 -0.99] [ 0.    0.    0.66  0.34  0.    0.    0.  ]
6 [-0.99 -0.65 -0.93 -0.98] [ 0.    0.45  0.    0.    0.    0.55  0.  ]
6 [-1.   -0.95 -0.95 -0.95] [ 0.    0.44  0.    0.    0.    0.56  0.  ]
5 [-1.   -1.   -0.97  0.99] [ 0.    0.    0.    0.    0.4   0.59  0.  ]
4 [-1.   -0.71  0.99  0.99] [ 0.  0.  0.  0.  0.  0.  1.]
epoch: 50000
error: 0.0024
```



2、Results of the training process:

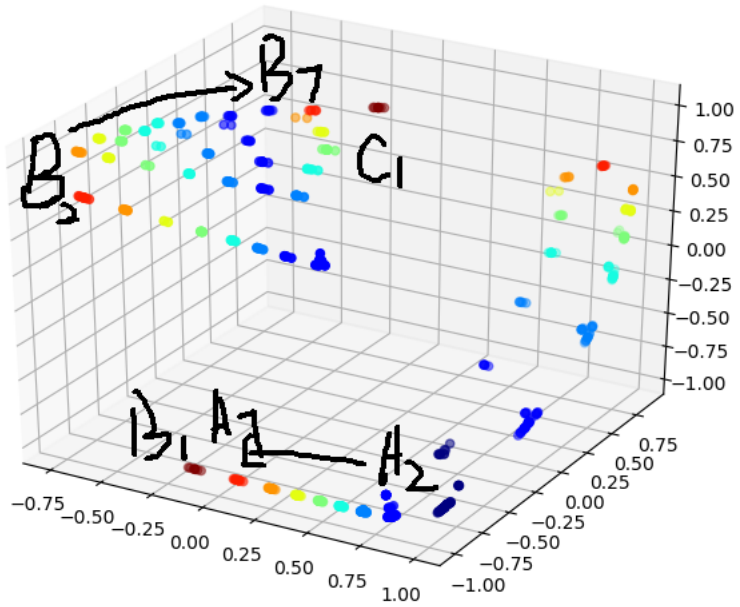
```
-----
color = 0123456543210123456765432101234565432101234565432101234567654321O
symbol= AAAAAABBBBBBAAAAAAABBBBBBAAAAAABBBBBBAAAAAABBBBBBAAAAAABBBBBBBA
label = 0000001111110000000111111100000011111100000011111100000001111111O
hidden activations and output probabilities:
A [-0.59  0.52  1.   -1.  ] [ 0.86  0.14]
A [-0.79 -0.31  0.99 -1.  ] [ 0.84  0.16]
A [-0.82 -0.63  0.92 -1.  ] [ 0.82  0.18]
A [-0.83 -0.71  0.8  -0.98] [ 0.78  0.22]
A [-0.85 -0.75  0.68 -0.96] [ 0.72  0.28]
B [-0.86 -0.78  0.53 -0.91] [ 0.6  0.4]
B [ 0.4   0.23 -0.48  0.99] [ 0.  1.]
B [ 0.54  0.73 -0.68  1.  ] [ 0.  1.]
B [ 0.54  0.8  -0.43  0.99] [ 0.  1.]
B [ 0.59  0.84 -0.05  0.95] [ 0.  1.]
B [ 0.66  0.89  0.51  0.56] [ 0.01  0.99]
A [ 0.75  0.94  0.92 -0.88] [ 0.98  0.02]
A [-0.59  0.56  1.   -1.  ] [ 0.85  0.15]
A [-0.79 -0.29  0.99 -1.  ] [ 0.84  0.16]
A [-0.82 -0.62  0.92 -1.  ] [ 0.82  0.18]
A [-0.83 -0.71  0.8  -0.98] [ 0.78  0.22]
A [-0.85 -0.75  0.68 -0.96] [ 0.72  0.28]
A [-0.86 -0.78  0.54 -0.91] [ 0.61  0.39]
B [-0.87 -0.81  0.34 -0.78] [ 0.37  0.63]
B [ 0.35  0.14 -0.71  1.  ] [ 0.  1.]
B [ 0.48  0.65 -0.86  1.  ] [ 0.  1.]
```

```
A [-0.83 -0.71  0.8  -0.98] [ 0.78  0.22]
A [-0.85 -0.75  0.68 -0.96] [ 0.72  0.28]
A [-0.86 -0.78  0.54 -0.91] [ 0.61  0.39]
B [-0.87 -0.81  0.34 -0.79] [ 0.37  0.63]
B [ 0.36  0.14 -0.71  1.  ] [ 0.  1.]
B [ 0.48  0.65 -0.85  1.  ] [ 0.  1.]
B [ 0.49  0.75 -0.67  1.  ] [ 0.  1.]
B [ 0.53  0.8  -0.43  0.99] [ 0.  1.]
B [ 0.59  0.84 -0.06  0.95] [ 0.  1.]
B [ 0.66  0.89  0.5   0.58] [ 0.01  0.99]
A [ 0.75  0.94  0.92 -0.87] [ 0.97  0.03]
epoch: 16000
error: 0.0097
```



3. For the sequence with only two letters of A and B, we found that other Bs are predictable, and the first occurrence of B is mispredicted, because the previous predictions are all A, suddenly changed to B, hidden The layer state change is not obvious enough, resulting in a prediction error in the final calculation.

4、



For the AnBnCn task, the first B in the sequence is unpredictable, but the next B, C, and the first A in the next sequence are predictable. The model counts the amount of A and gradually moves to the B area, which is painted from dark blue to orange. Meanwhile, the probabilities of B are all 1s because all Bs are predictable. After that, the network should predict that the next symbol is C, and the color immediately changes from blue to red. Finally, if C's count reaches 0, the network will predict that A will be next and start over until it's done. So the probability of the last row of A in the next sequence is 1.

The output of the network is:

```
color = 012121012345654321654321012345432154321012343214321012343214321O
symbol= AABBCCAAAAAABBBBBBCCCCCCAAAAABBBBBCCCCCAAAABBBBCCCCAAAABBBBCCCCA
label = 00112200000011111122222200000111112222200001111222200001111222200
hidden activations and output probabilities:
A [-0.98 -0.98 -0.95  0.79] [ 0.86  0.14  0.  ]
B [-1.   -0.53 -0.95  1.  ] [ 0.81  0.19  0.  ]
B [ 0.68 -0.14 -0.94  0.35] [ 0.    0.99  0.01]
C [ 1.   -0.72 -0.83 -0.35] [ 0.  0.  1.]
C [ 0.24 -0.97 -0.67 -0.98] [ 0.  0.  1.]
A [-0.87 -0.99 -0.81 -1.  ] [ 0.99  0.    0.  ]
A [-1.   -0.98 -0.96  0.78] [ 0.88  0.12  0.  ]
A [-1.   -0.53 -0.95  1.  ] [ 0.81  0.19  0.  ]
A [-0.99 -0.02 -0.9   1.  ] [ 0.81  0.19  0.  ]
A [-0.99  0.38 -0.81  1.  ] [ 0.8  0.2  0. ]
A [-0.98  0.64 -0.66  1.  ] [ 0.76  0.24  0.  ]
B [-0.96  0.76 -0.45  1.  ] [ 0.65  0.35  0.  ]
B [ 0.98  0.77 -0.41  0.97] [ 0.  1.  0.]
B [ 1.    0.42  0.1   0.9 ] [ 0.  1.  0.]
B [ 1.    0.23  0.4   0.76] [ 0.  1.  0.]
B [ 1.    0.01  0.53  0.54] [ 0.  1.  0.]
B [ 1.   -0.33  0.5   0.07] [ 0.    0.98  0.02]
```

```
B [ 1.   -0.33  0.5   0.07] [ 0.    0.98  0.02]
C [ 1.   -0.79  0.22 -0.72] [ 0.  0.  1.]
C [ 0.99 -0.98  0.17 -0.99] [ 0.  0.  1.]
C [ 0.99 -0.99 -0.05 -1.  ] [ 0.  0.  1.]
C [ 0.95 -0.99 -0.28 -1.  ] [ 0.  0.  1.]
C [ 0.83 -0.99 -0.49 -1.  ] [ 0.  0.  1.]
C [ 0.34 -0.99 -0.66 -1.  ] [ 0.  0.  1.]
A [-0.81 -0.99 -0.8  -1.  ] [ 0.99  0.    0.01]
A [-1.   -0.98 -0.96  0.77] [ 0.88  0.12  0.  ]
A [-1.   -0.54 -0.95  1.  ] [ 0.81  0.19  0.  ]
A [-0.99 -0.03 -0.9   1.  ] [ 0.81  0.19  0.  ]
A [-0.99  0.38 -0.81  1.  ] [ 0.8  0.2  0. ]
B [-0.98  0.63 -0.66  1.  ] [ 0.76  0.24  0.  ]
B [ 0.94  0.7  -0.63  0.96] [ 0.  1.  0.]
B [ 1.    0.33 -0.18  0.88] [ 0.  1.  0.]
B [ 1.    0.06  0.07  0.68] [ 0.  1.  0.]
B [ 1.   -0.28  0.14  0.26] [ 0.    0.99  0.01]
C [ 1.   -0.74 -0.09 -0.58] [ 0.  0.  1.]
C [ 0.97 -0.98 -0.09 -0.99] [ 0.  0.  1.]
C [ 0.94 -0.99 -0.31 -1.  ] [ 0.  0.  1.]
C [ 0.79 -0.99 -0.52 -1.  ] [ 0.  0.  1.]
C [ 0.21 -0.99 -0.68 -1.  ] [ 0.  0.  1.]
A [-0.89 -0.99 -0.82 -1.  ] [ 0.99  0.    0.  ]
A [-1.   -0.98 -0.96  0.78] [ 0.88  0.12  0.  ]
A [-1.   -0.53 -0.95  1.  ] [ 0.81  0.19  0.  ]
A [-0.99 -0.02 -0.9   1.  ] [ 0.81  0.19  0.  ]
B [-0.99  0.38 -0.81  1.  ] [ 0.8  0.2  0. ]
B [ 0.86  0.56 -0.78  0.92] [ 0.  1.  0.]
B [ 1.    0.17 -0.42  0.82] [ 0.  1.  0.]
B [ 1.   -0.18 -0.28  0.48] [ 0.  1.  0.]
```

```
C [ 0.89 -0.96 -0.28 -0.97] [ 0.  0.  1.]
C [ 0.79 -0.99 -0.49 -1.  ] [ 0.  0.  1.]
C [ 0.27 -0.99 -0.66 -1.  ] [ 0.  0.  1.]
A [-0.85 -0.99 -0.81 -1.  ] [ 0.99  0.    0.01]
A [-1.   -0.98 -0.96  0.78] [ 0.88  0.12  0.  ]
A [-1.   -0.54 -0.95  1.  ] [ 0.81  0.19  0.  ]
A [-0.99 -0.03 -0.9   1.  ] [ 0.81  0.19  0.  ]
B [-0.99  0.38 -0.81  1.  ] [ 0.8  0.2  0. ]
B [ 0.86  0.56 -0.78  0.92] [ 0.  1.  0.]
B [ 1.    0.17 -0.42  0.82] [ 0.  1.  0.]
B [ 1.   -0.18 -0.28  0.48] [ 0.  1.  0.]
C [ 1.   -0.64 -0.4  -0.31] [ 0.    0.01  0.99]
C [ 0.89 -0.96 -0.29 -0.97] [ 0.  0.  1.]
C [ 0.78 -0.99 -0.49 -1.  ] [ 0.  0.  1.]
C [ 0.24 -0.99 -0.67 -1.  ] [ 0.  0.  1.]
A [-0.87 -0.99 -0.81 -1.  ] [ 0.99  0.    0.01]
epoch: 44000
error: 0.0006
```