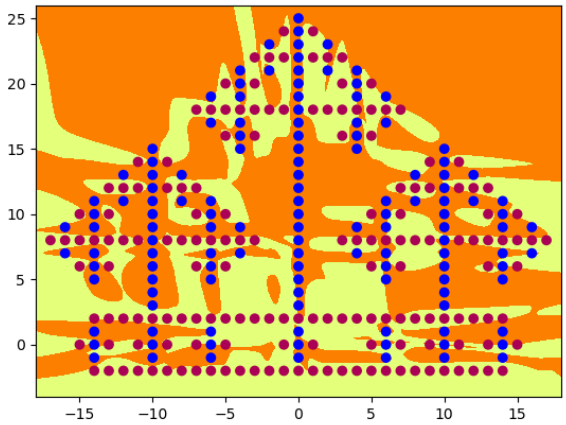
Report

Part1:

2. include output picture in your report, as well as a calculation of the total number of independent parameters in your network (based on the number of hidden nodes you have chosen)

I set the number of hidden nodes to 20 and it takes 200000 epochs to reach 100% accuracy.

Out_full3_20.png:



calculation of the total number of independent parameters:

I set hidden nodes number to 20 and successfully classify the data.

w1 : 2*20 = 40, b1 : 20

w2 : 20*20 = 400 , b2 : 20

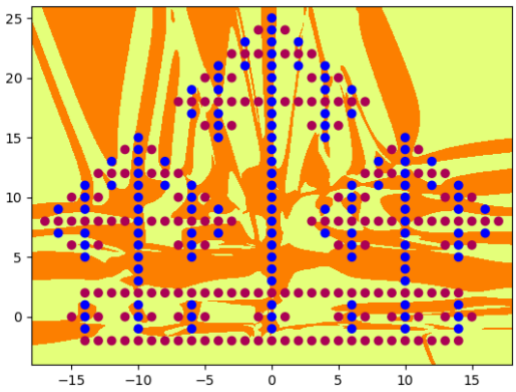
w3 : 20*1 = 20, b3 : 1

total number of parameters: (2*20+20) + (20*20+20) + (20*1+1) = 501

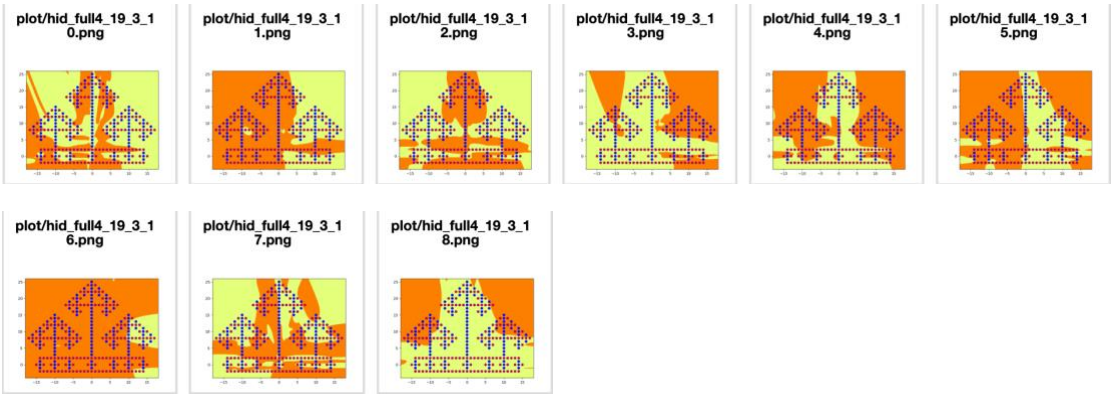
4. include the plot of the output and the plots of all the hidden units in all three layers in your report, as well as a calculation of the total number of independent parameters in your network.

I set the number of hidden nodes to 19 and it takes 78600 epochs to reach 100% accuracy.

Output plot:



Hidden units plots:



calculation of the total number of independent parameters:

I set the hidden nodes number to 19 and successfully classify the data.

w1 : 2*19 = 40, b1 : 19

w2 : 19*19 = 361 , b2 : 19

w3 : 19*19 = 361 , b2 : 19

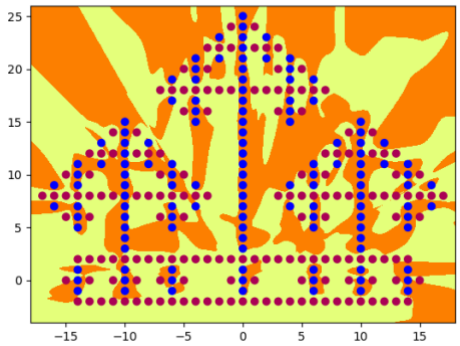
w4 : 19*1 = 20, b3 : 1

total: (2*19+19) + (19*19+19) + (19*19+19)+ (19*1+1) = 837

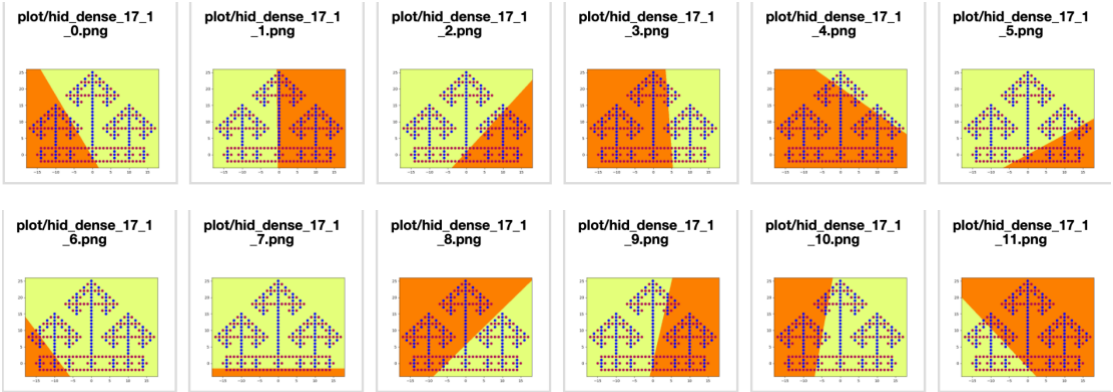
6. include the graphs of the output and all the hidden nodes in both layers in your report, as well as a calculation of the total number of independent parameters in your network.

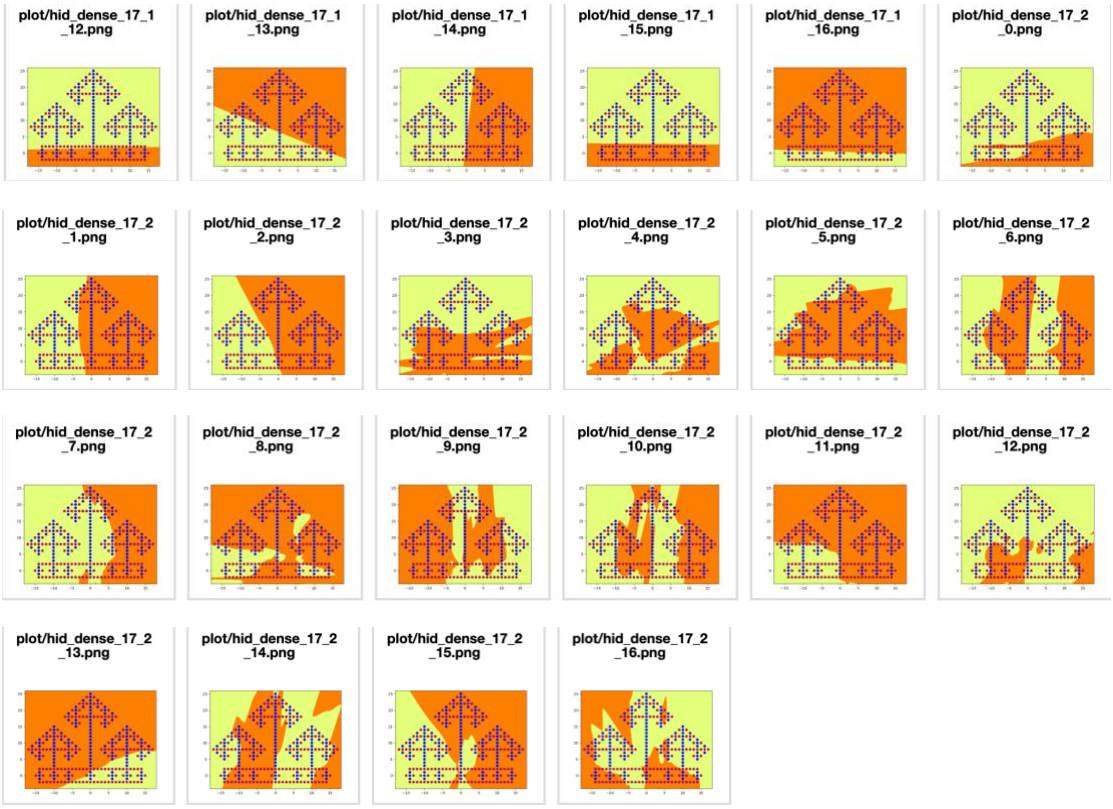
I set the number of hidden nodes to 17 and it takes 47400 epochs to reach 100% accuracy.

The graph of the output:



Hidden output plots:





calculation of the total number of independent parameters:

I set hidden nodes number to 17 and successfully classify the data.

w1 : 2*17 = 34, b1 : 17

w2 : 17*17+17*2 = 323 , b2 : 17

w3 : 17*1+17*1+2*1 = 36, b3 : 1

total number of parameters: （2*17+17） + [(17*17+17)+2*17] + [(17*1+1)+17*1+2*1] = 428

7. Briefly discuss the following points:

a. the total number of independent parameters in each of the three networks and the approximate number of epochs required to train each type of network.

I set the hidden nodes of the three networks as 20, 19 and 17. As calculated in q2, q4 and q6, the total number of independent parameters in each of the three networks is 501, 837 and 428 respectively. The approximate number of epochs required to train each type of network is 200000, 78600 and 47400.

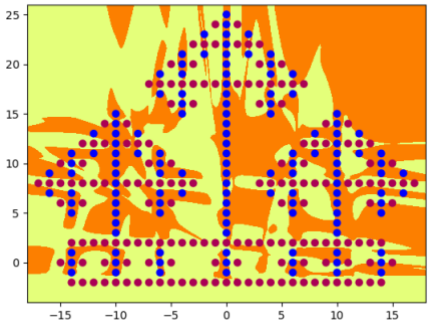
b. a qualitative description of the functions computed by the different layers of Full4Net and DenseNet.

As we can see in the hidden layer plots shown in the question4 and question6, the dividing line of the first hidden layer tends to very sharp and linear, and then the dividing line of the second hidden layer or the third turns less linear, more smooth and turns into irregular shape, which can better classify the data.

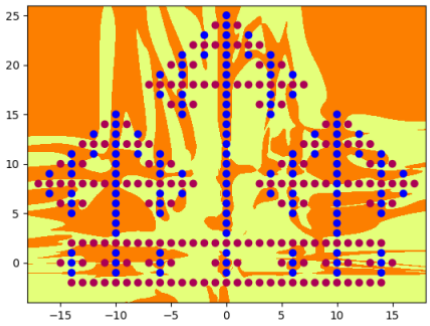
c. the qualitative difference, if any, between the overall function computed by the

three networks.

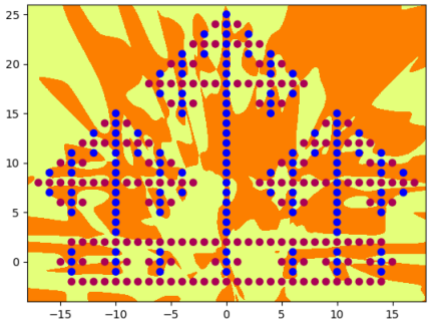
If we set the hidden nodes number of the three networks to the same, for example 20. The output plots are as followed.



Out_full3_20



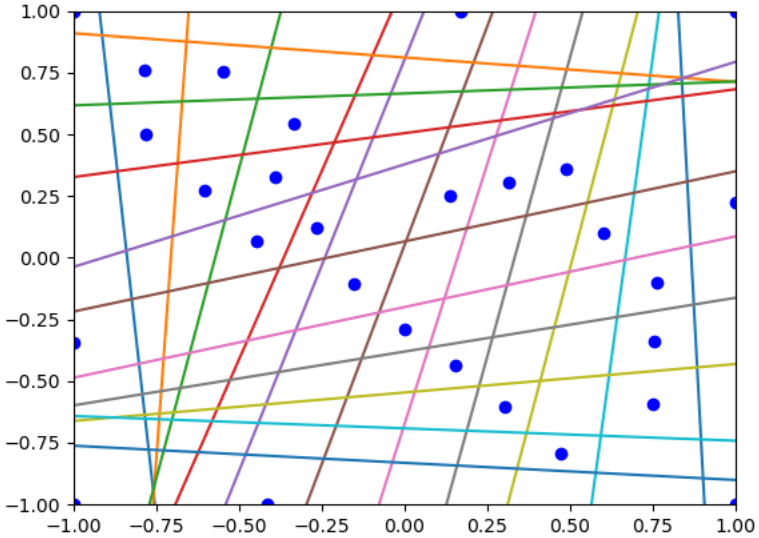
out_full4_20



out_dense_20

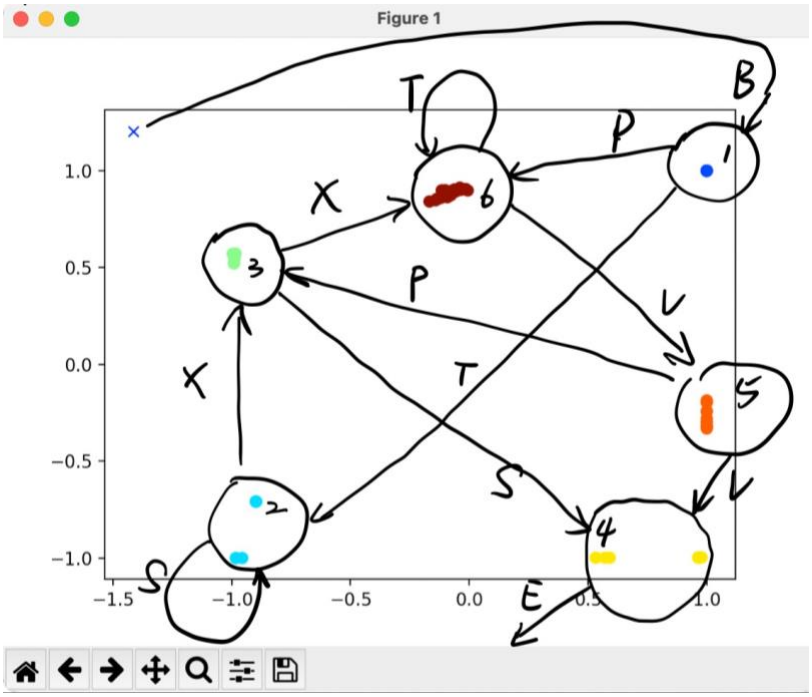
Part2:

Include the final image in your report.

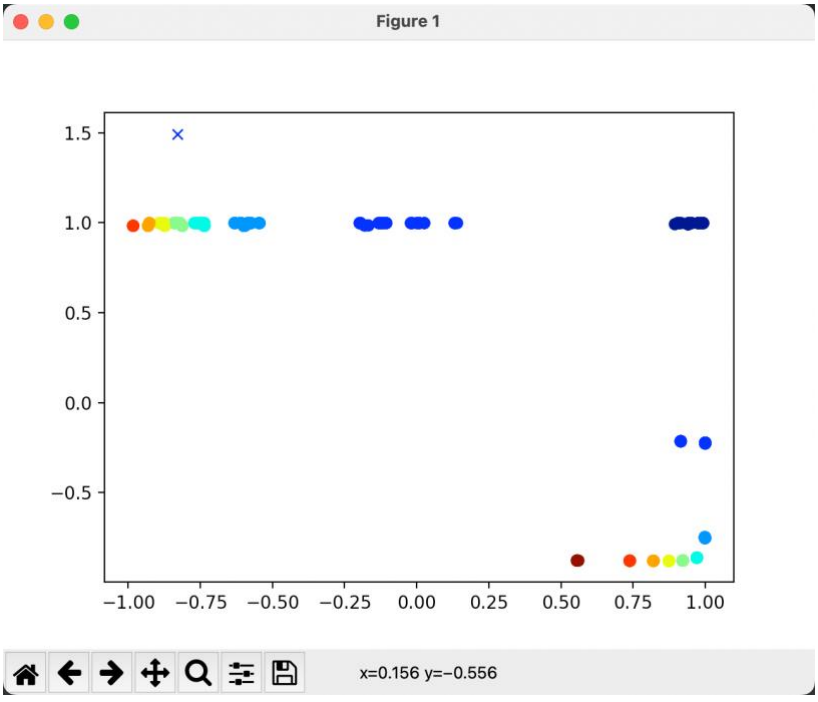


Part3:

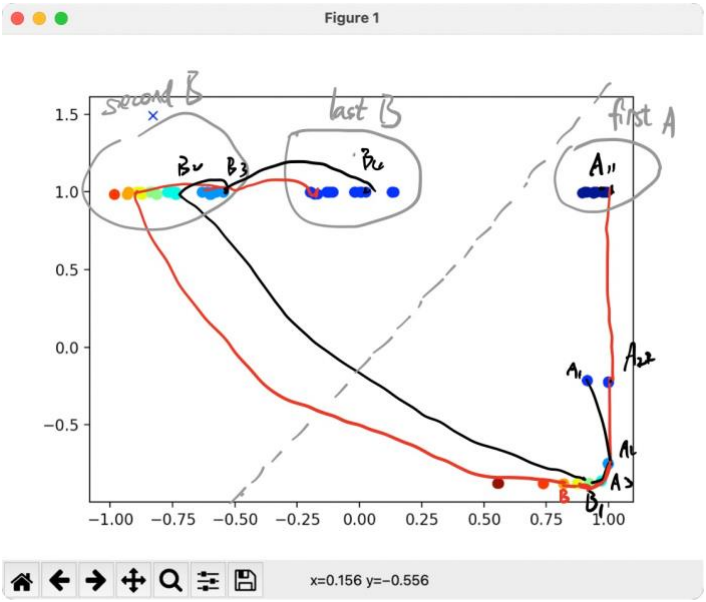
1. Include the annotated figure in your report.



2. Include the resulting figure in your report.



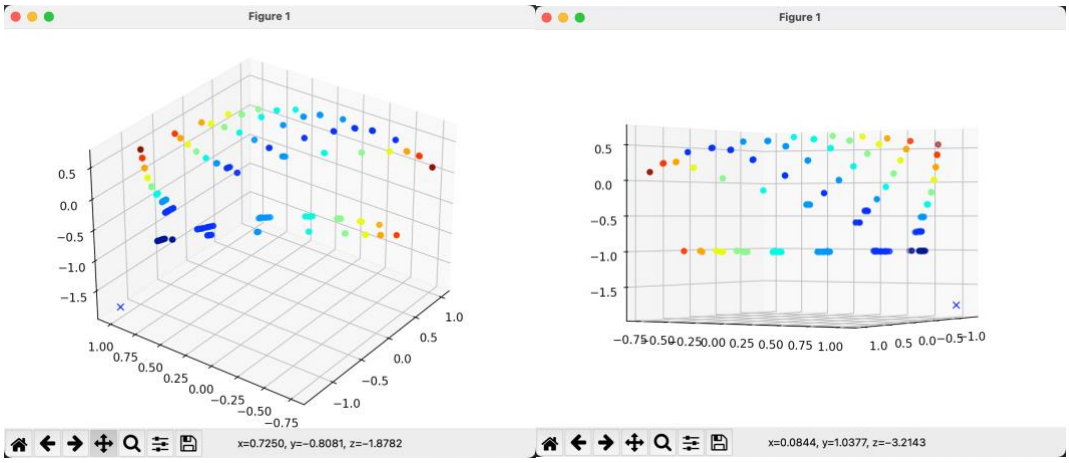
3. Briefly explain how the anbn prediction task is achieved by the network, based on the figure you generated in Question 2. Specifically, you should describe how the hidden unit activations change as the string is processed, and how it is able to correctly predict the last B in each sequence as well as the following A.



When the string is processed, the hidden unit activations change from one state to another, as we see in the figure, it is jumping from one color to another.

The last B, as we can see in the picture, is always in the blue group in the up middle. Similarly, the following A is always in the dark blue in the up right.

4. Rotate the figure in 3 dimensions to get one or more good view(s) of the points in hidden unit space.



5. Briefly explain how the anbncn prediction task is achieved by the network, based on the figure you generated in Question 4. Specifically, you should describe how the hidden unit activations change as the string is processed, and how it is able to correctly predict the last B in each sequence as well as all of the C's and the following A.

When the string is processed, the dot of hidden unit activation changes from one color to anther in the plot.

The last B are in the same color group, which is blue, the same color as the question3 illustrated. All the C tend to in the dot group whose z is -1. The following A is exactly in the

same plane with the C group, whose z is -1 as well.

6. try to analyze the behavior of the LSTM and explain how the task is accomplished.