Neural Network

COMP9444

Assignment 1

T3,2022

Zeyu Zhou

z5382403

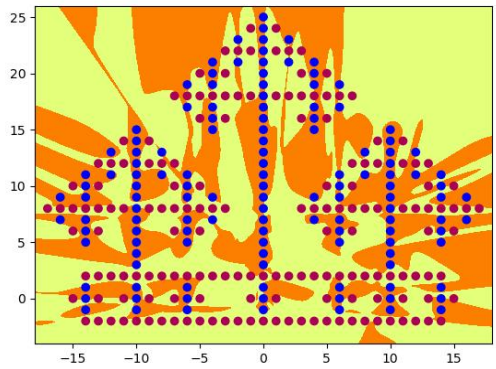
Z5382403@ad.unsw.edu.au

Part 1

Question 2:
The initial weight size is default 0.25.
The learning rate used as 0.001.
The first model which was successfully trained using these parameters was the one with 20 hidden nodes. The results from the tested models are shown below.

Hidden Nodes	Accuracy	Epochs
10	~84.93%	72500
15	~93.49%	76000
20	100%	95600

The output plot:

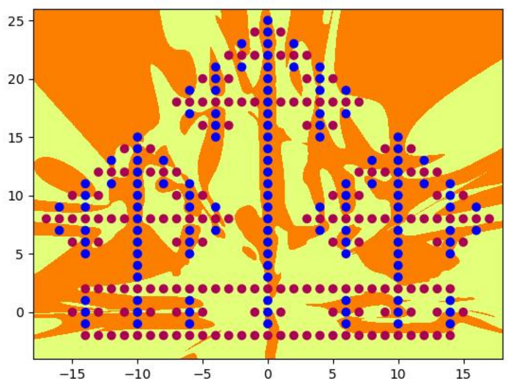


Total independent parameters: (20*2+20)+(20*20+20)+(1*20+1)=501

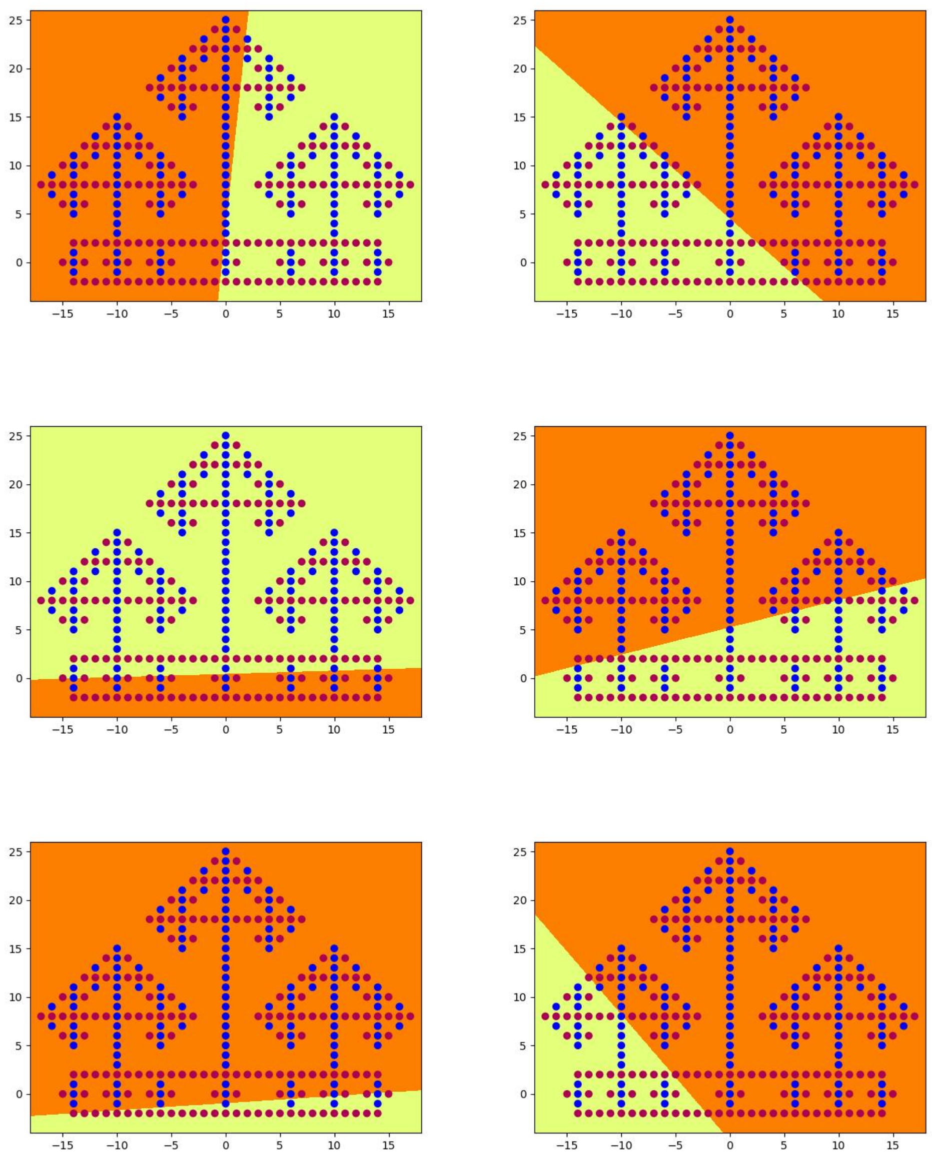
Question 4:
The initial weight size is default 0.25 .
The learning rate used as 0.001.
The first model which was successfully trained using these parameters was the one with 20 hidden nodes. The results from the tested models are shown below.

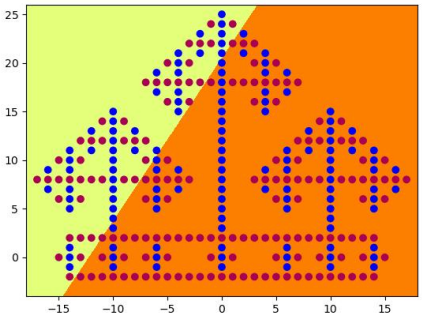
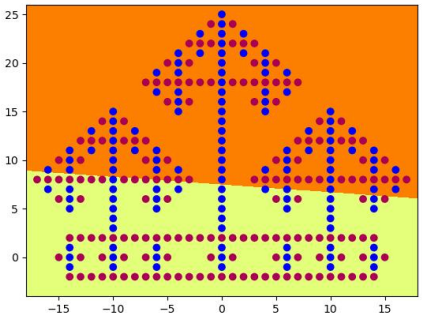
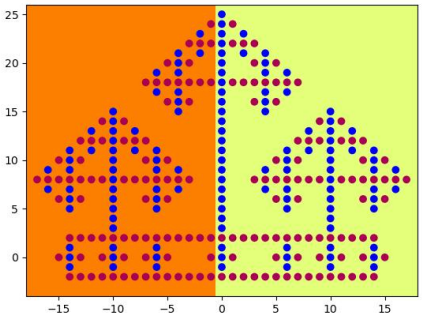
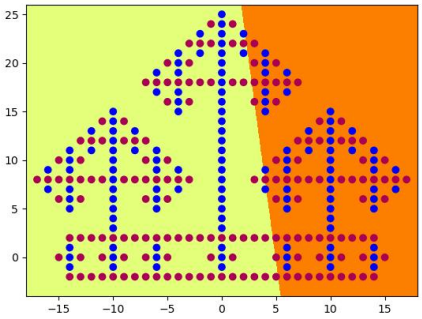
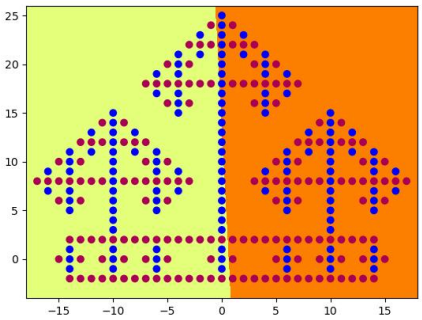
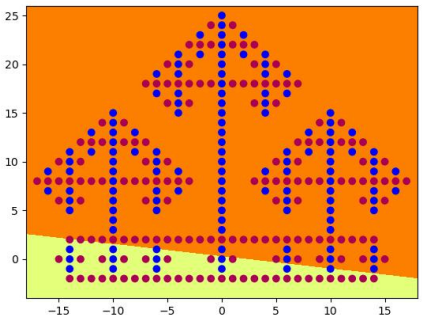
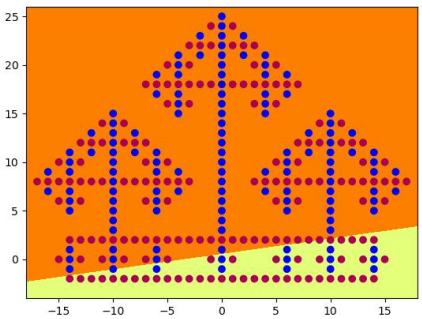
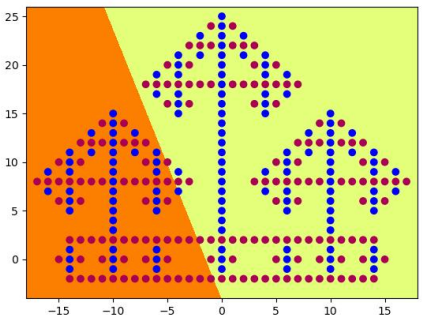
Hidden Nodes	Accuracy	Epochs
10	87.33%	200000
15	98.97%	200000
20	100%	32200

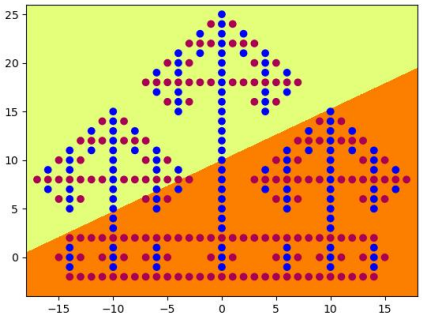
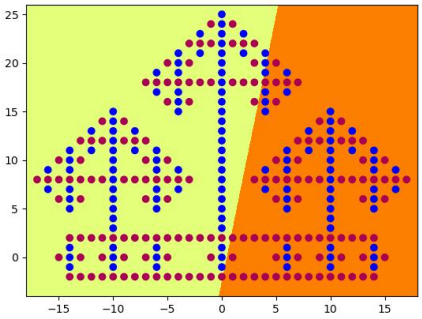
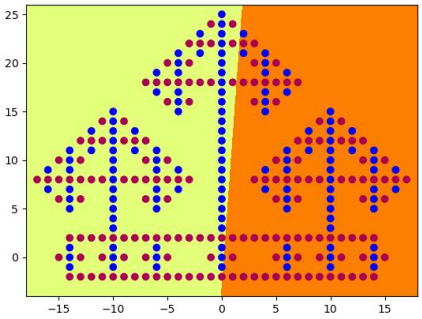
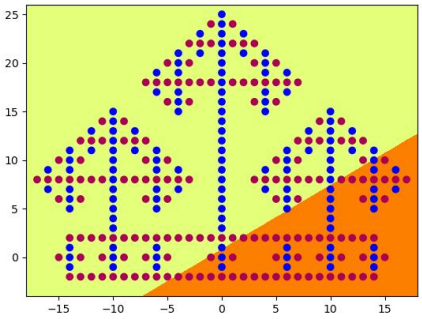
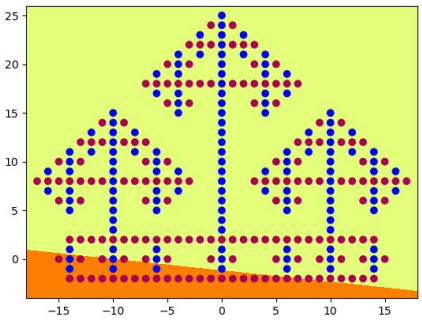
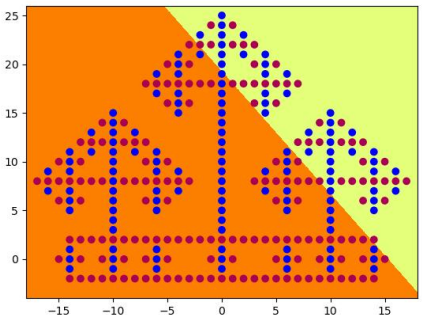
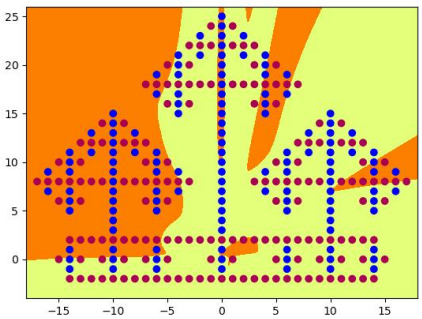
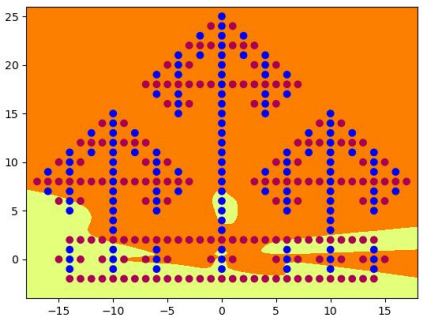
The output plot:

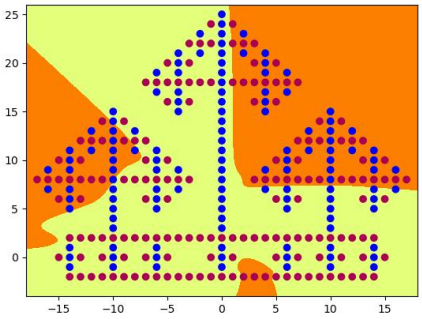
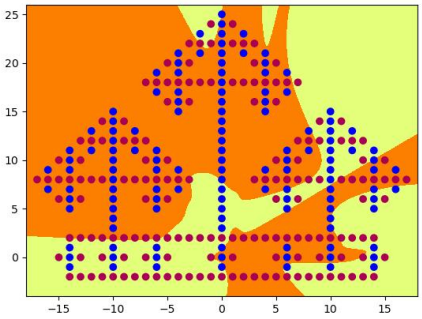
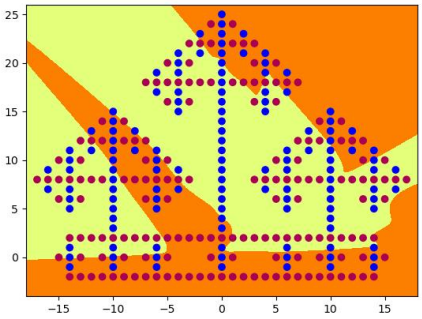
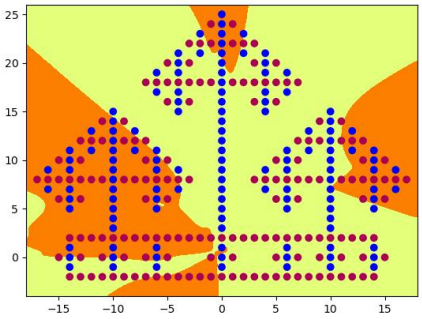
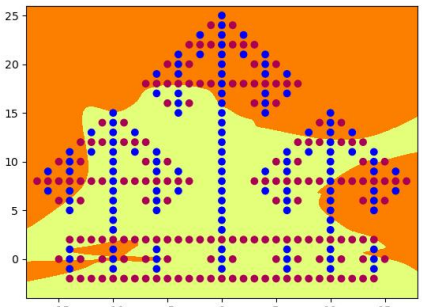
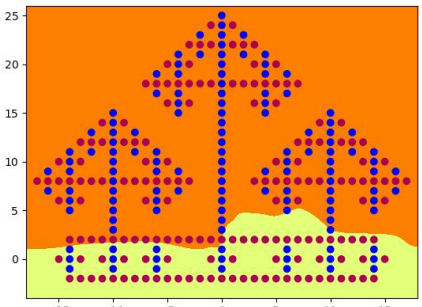
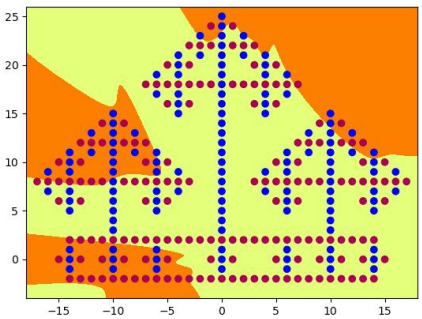
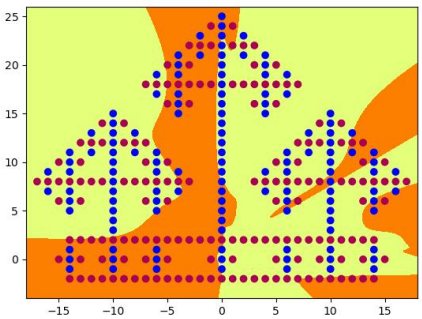


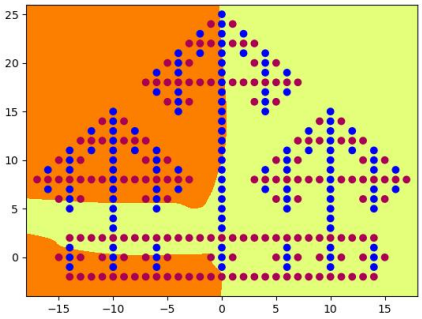
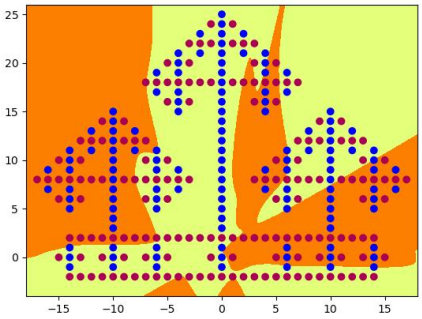
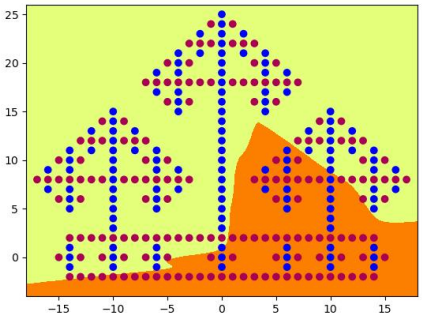
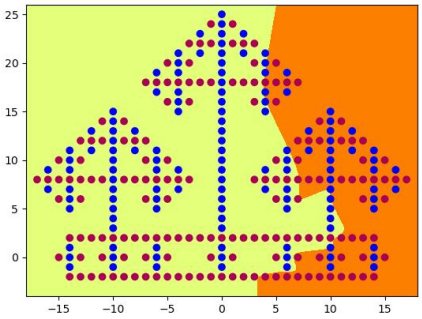
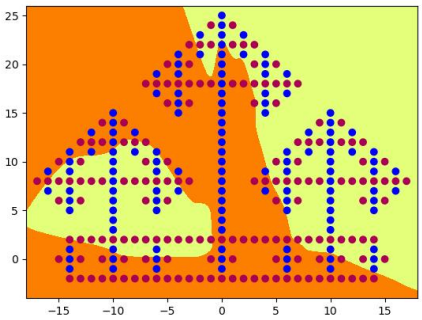
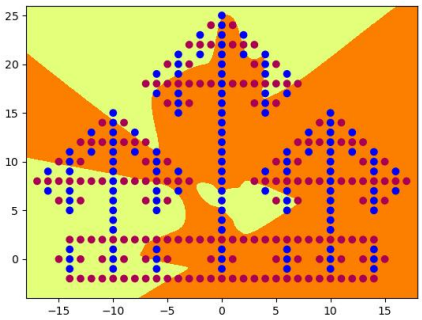
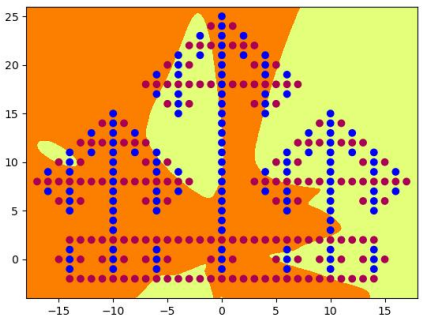
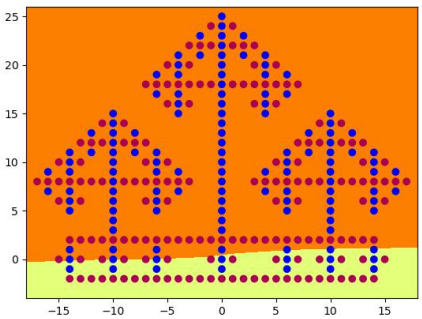
All hidden units:

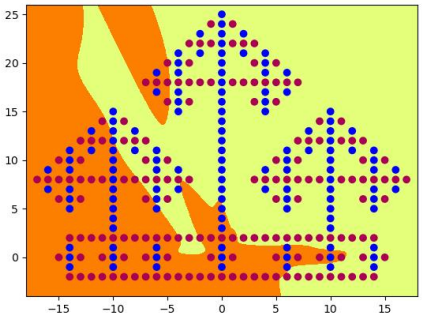
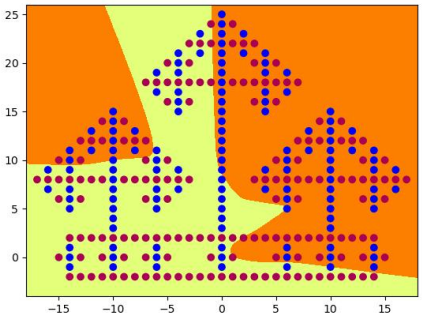
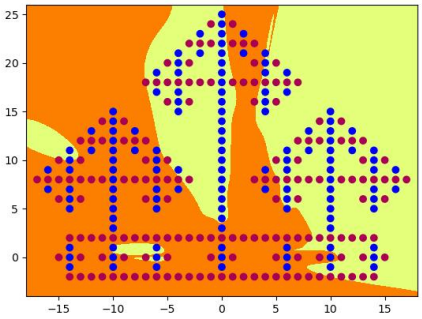
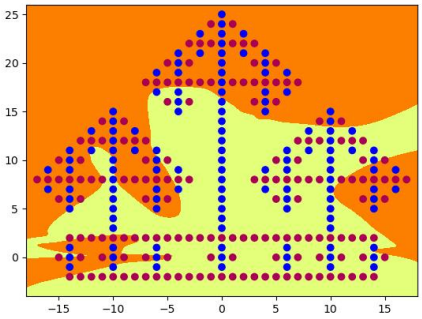
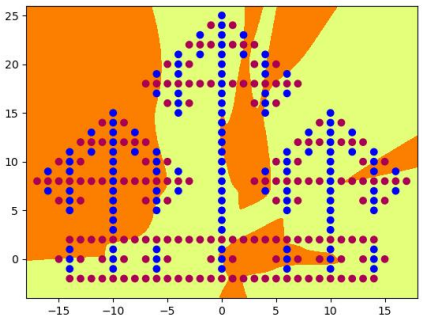
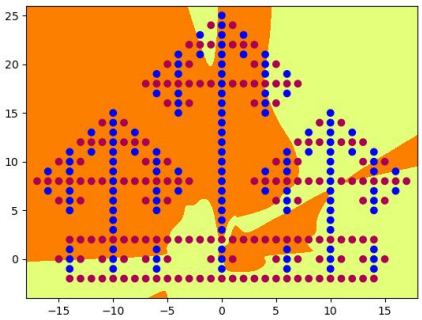
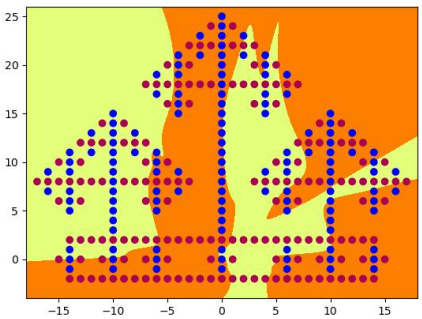
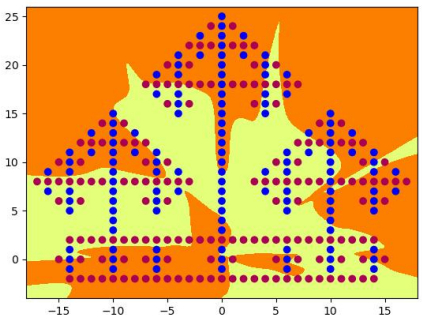


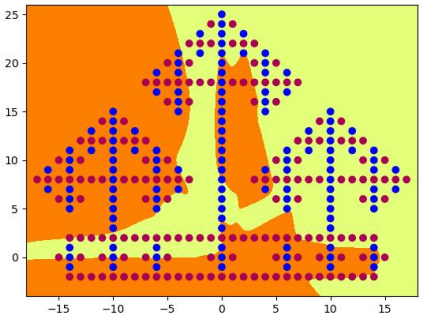
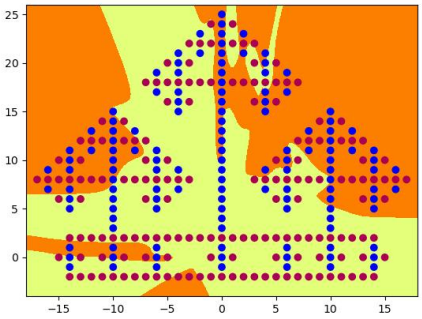
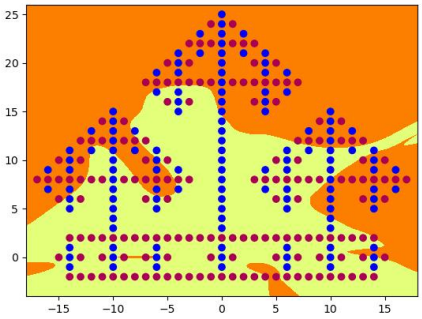
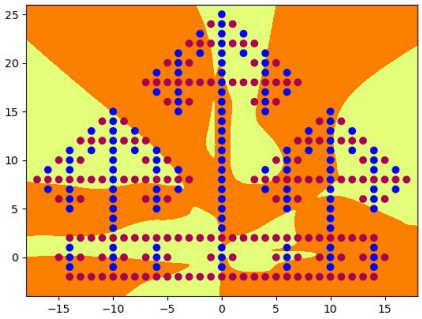
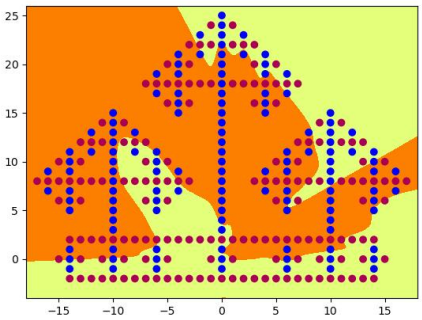
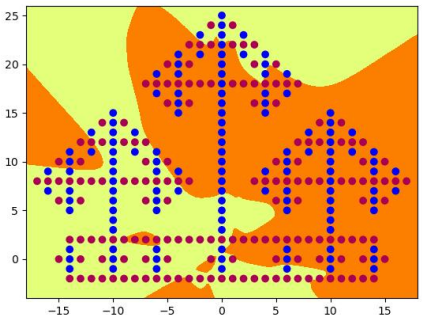
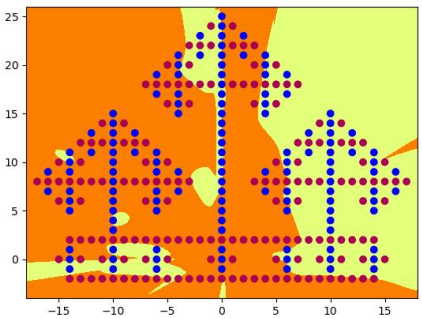
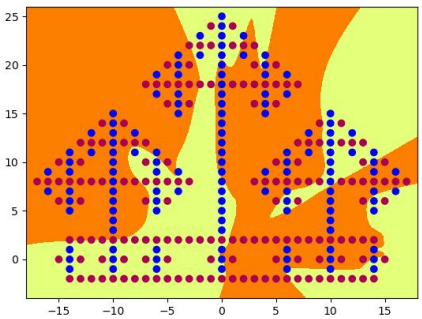









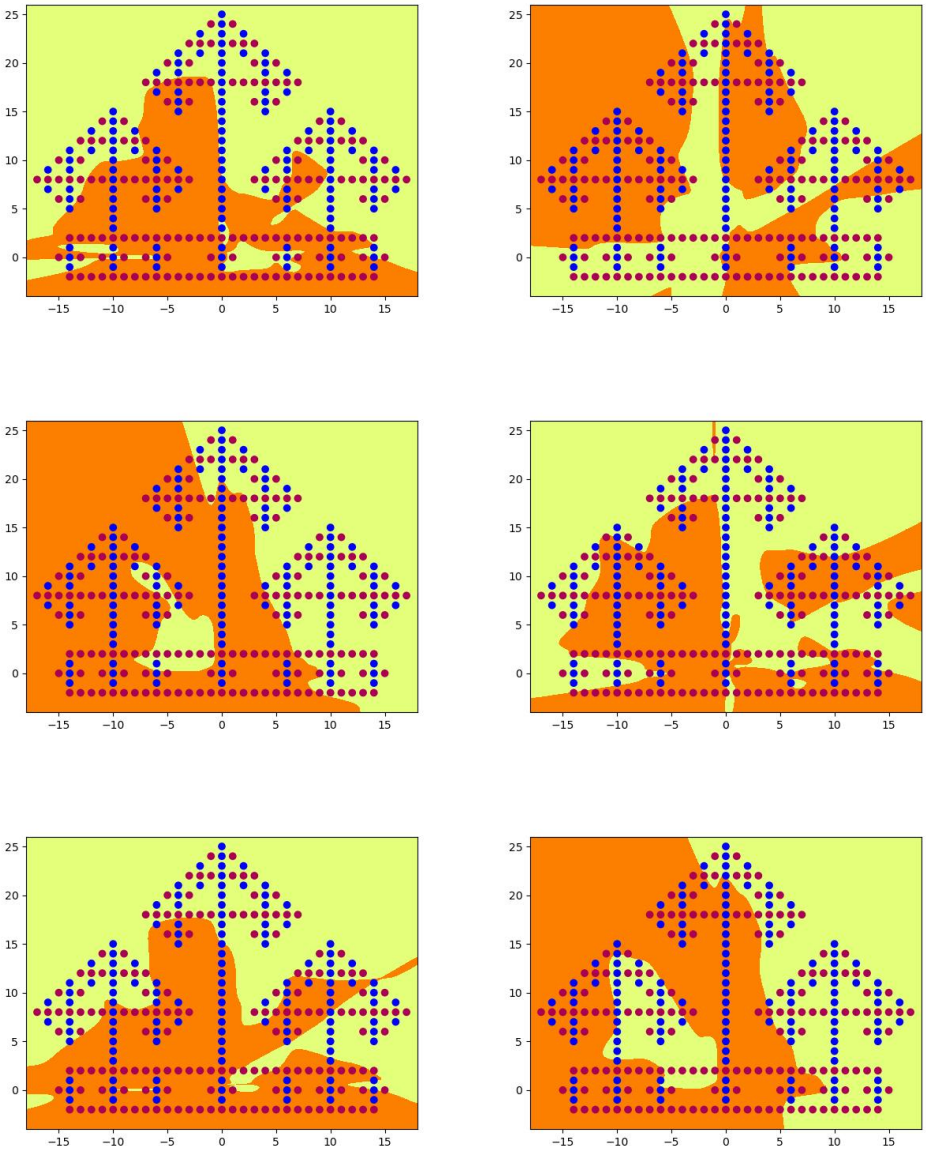











Total independent parameters:(20*2+20)+(20*20+20)*2+(1*20+1)=921

Question 6

DenseNet

The initial weight size is default 0.25 .

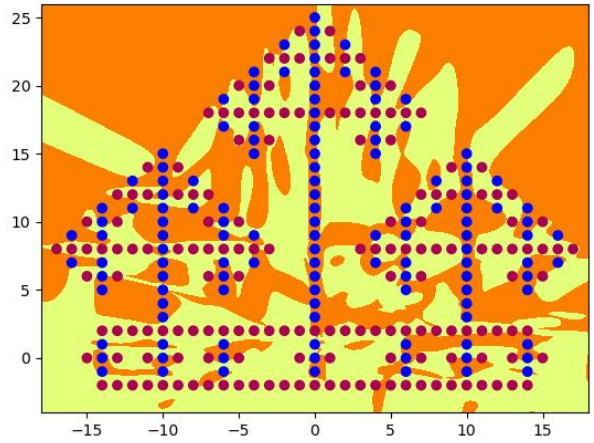
The learning rate used as 0.001.

The first model which was successfully trained using these parameters was the one with 15 hidden nodes. The results from the tested models are shown below.

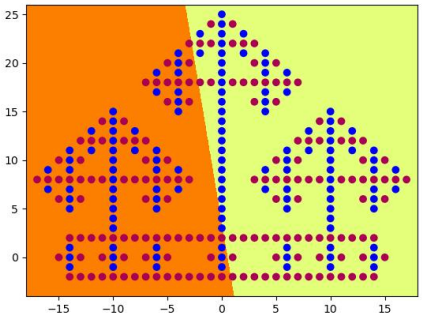
Hidden Nodes	Accuracy	Epochs
10	91.44%	61500
13	95.55%	33200

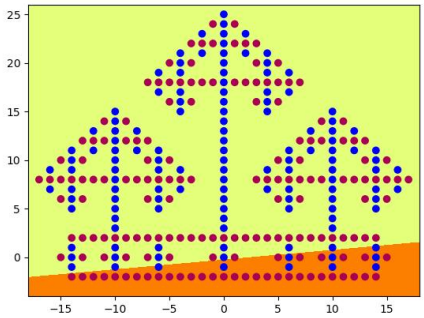
15	100%	72900

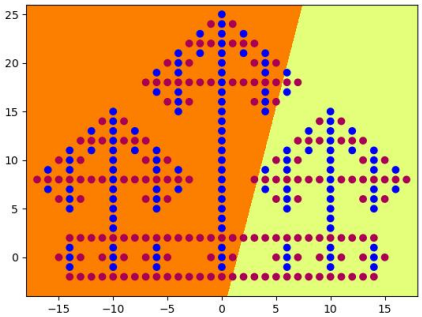
The output plot:

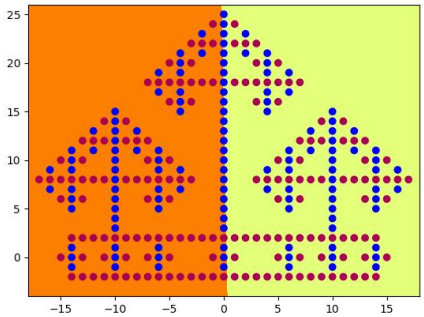


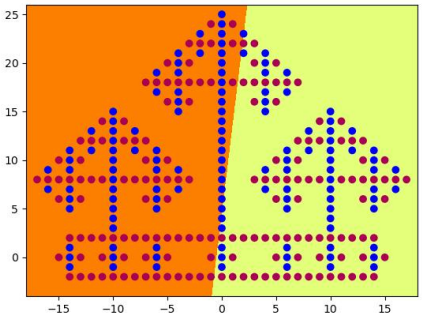
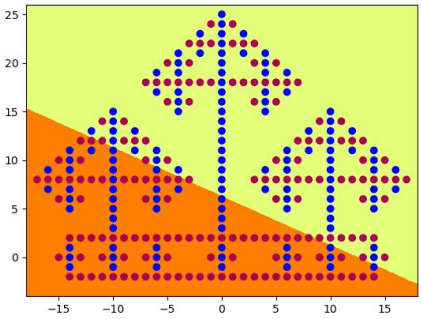
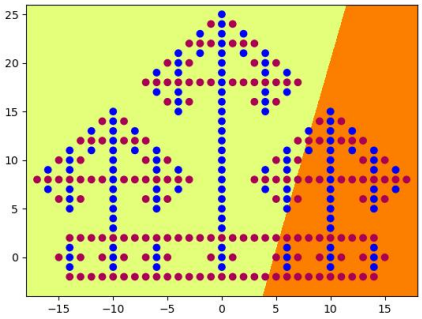
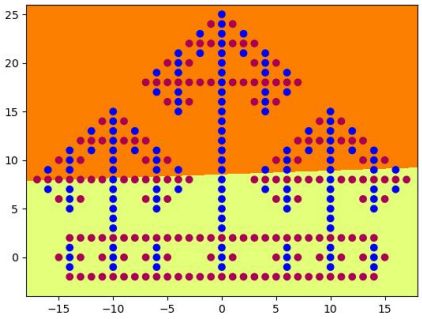
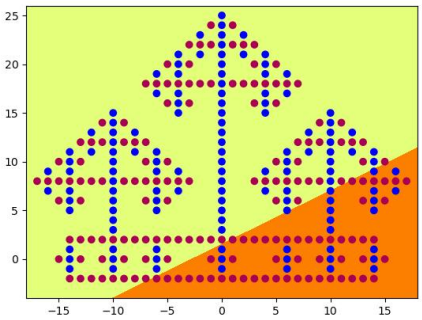
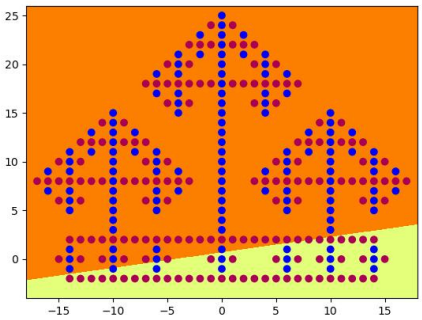
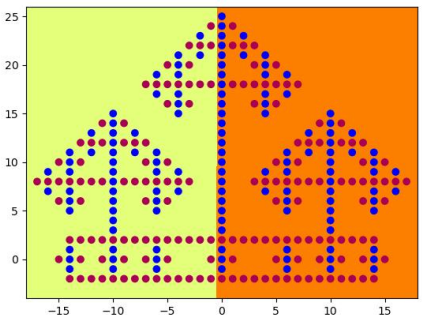
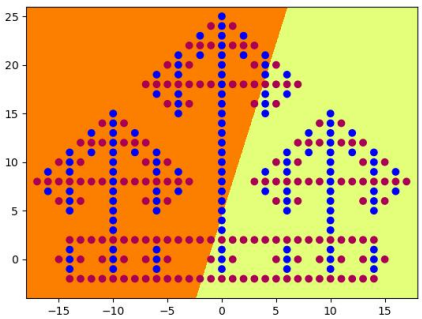
All hidden units:

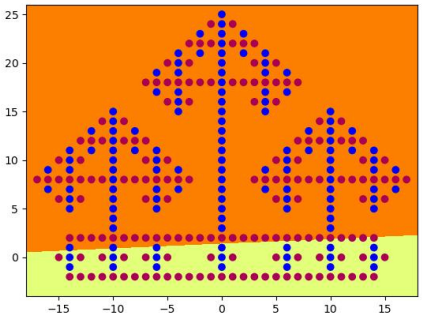
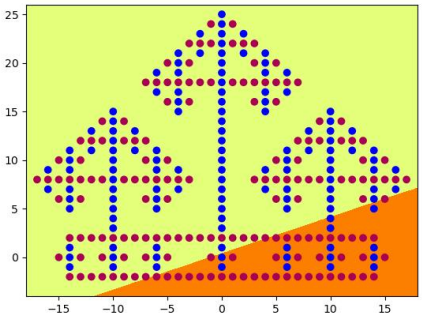
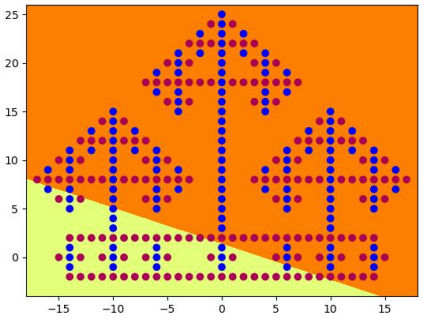
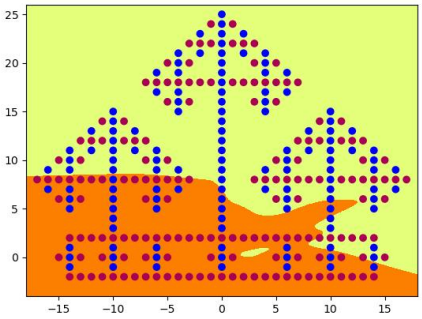
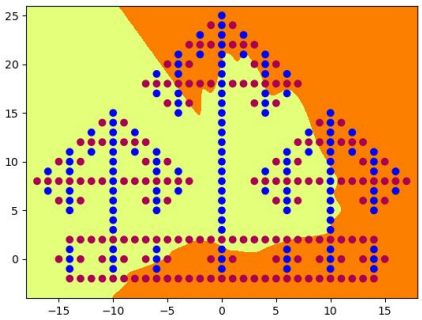
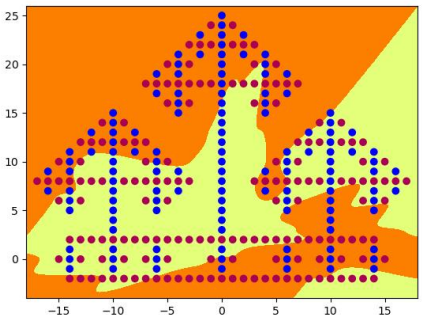
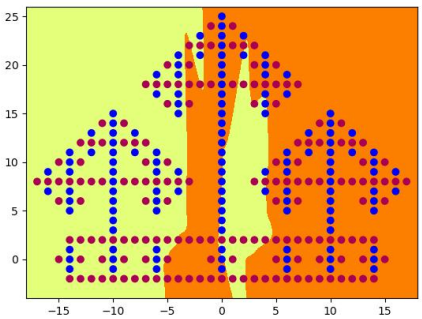
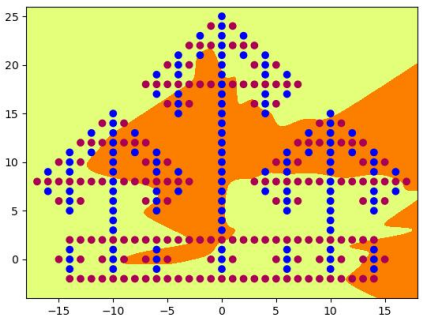


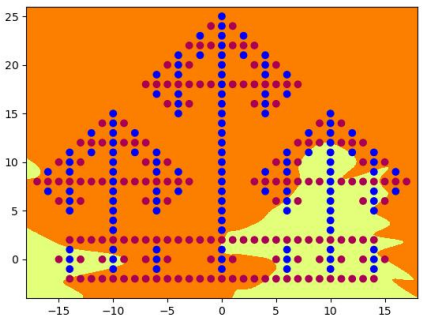
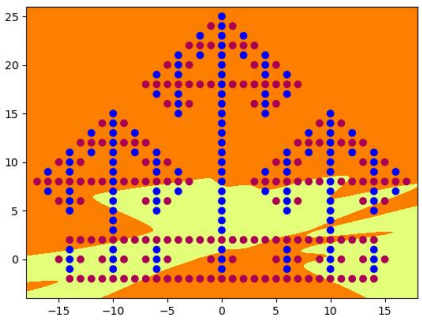
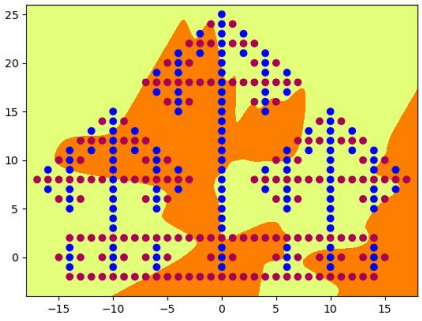
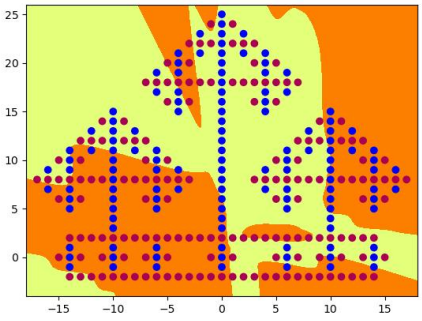
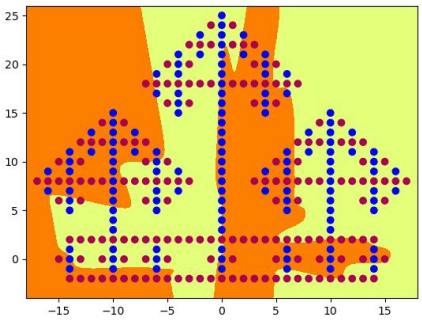
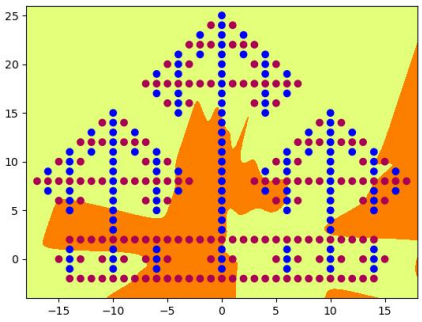
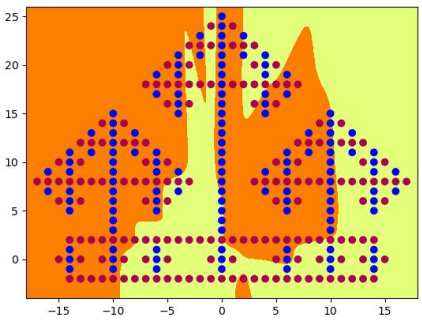
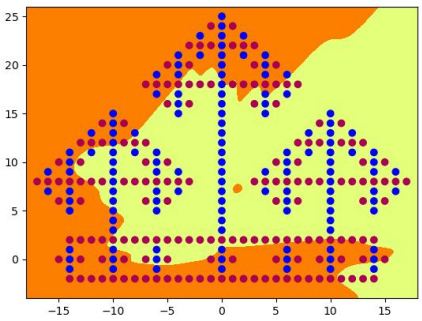


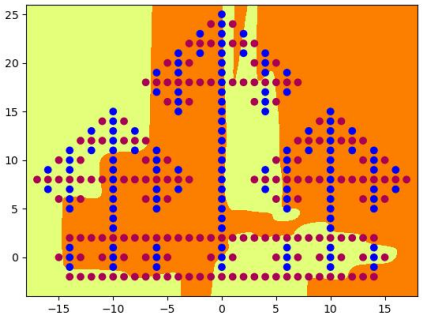


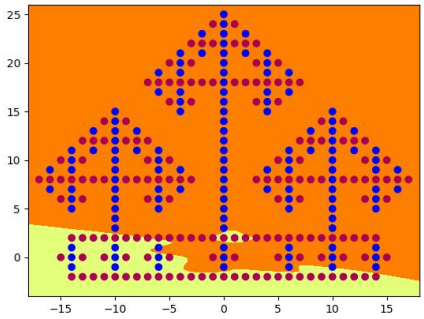






Total independent parameters:(15*1+1)+(15*2+15)+(15*15+15)+3=304

Question 7

(a).

Model	Hidden Nodes	Independent Parameters	Epochs
Full3Net	20	501	95600
Full4Net	20	921	32200
DenseNet	15	304	72900

We can easily find that The Full3Net required less hidden nodes than the Full3Net model. DenseNet was clearly the best model with the least number of hidden nodes and the least number of Independent Parameters .

(b).

The first layer of all three network models are linear lines.

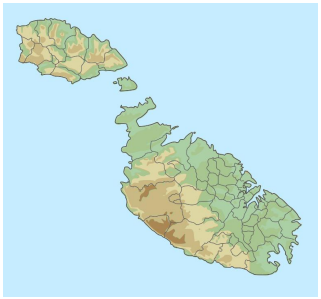
And then in the second layer , their complexity of the boundary increases,and this situation also further in dense,since it has more connections than the other two layers.

(c).

There is no significant difference between these three output maps. However, the dense map is able to identify the vertical features of the red points more effectively than the other two,and all three layers have diagonals on either side of the fractal shape that may indicate an alternating blue/red pattern, but the dense map instead showing a blue 'pocket' area in the bottom right.

Part 2

Solved this problem by making a matrix first and then assign each edge dot a numerical value(i.e. serial number)

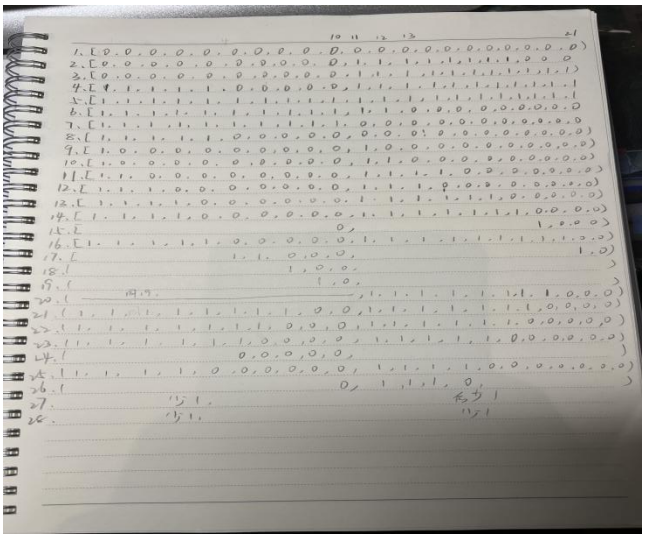


	A	B	C	D	E	F	G	H	I	J	K	L
1	1					8					7	
2		9	28									
3		10		27							6	
4					26							
5			11	12	25							
6						24	23					
7					13			22				
8					14				21			
9	2				15					20		
10						16						
11								17	18	19		
12	3					4					5	

And then ,we make two rules:

1. Dots on the top of a horizontal line is 0, bottom is 1.
2. Dots on the left of a vertical line is 0, right is 1.

Then we can have this following matrix of serial numbers of separating dots as rows and lines as columns.



Thus, we finally get the 28 dots' coordinates.

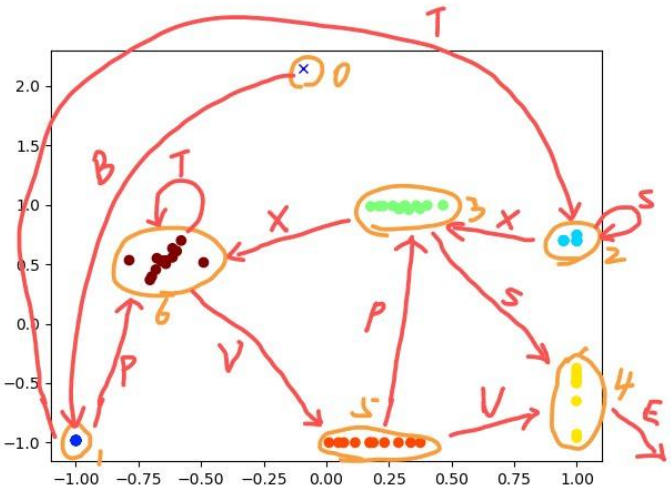
Part 3

Question 1 :

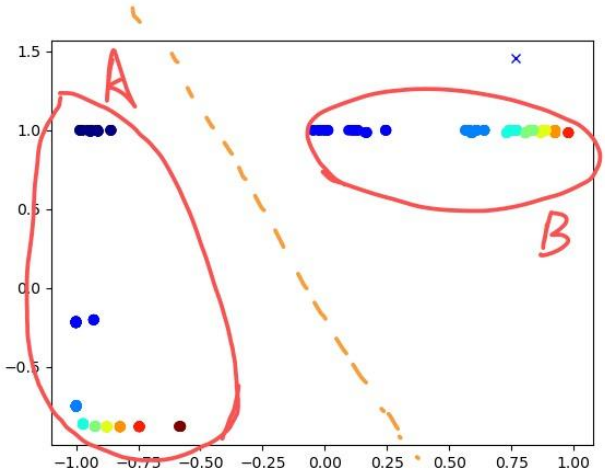
From the colormap , we can get the following diagram.

Orange circles and numbers = states

Red arrows and letters = symbols



Question 2 :

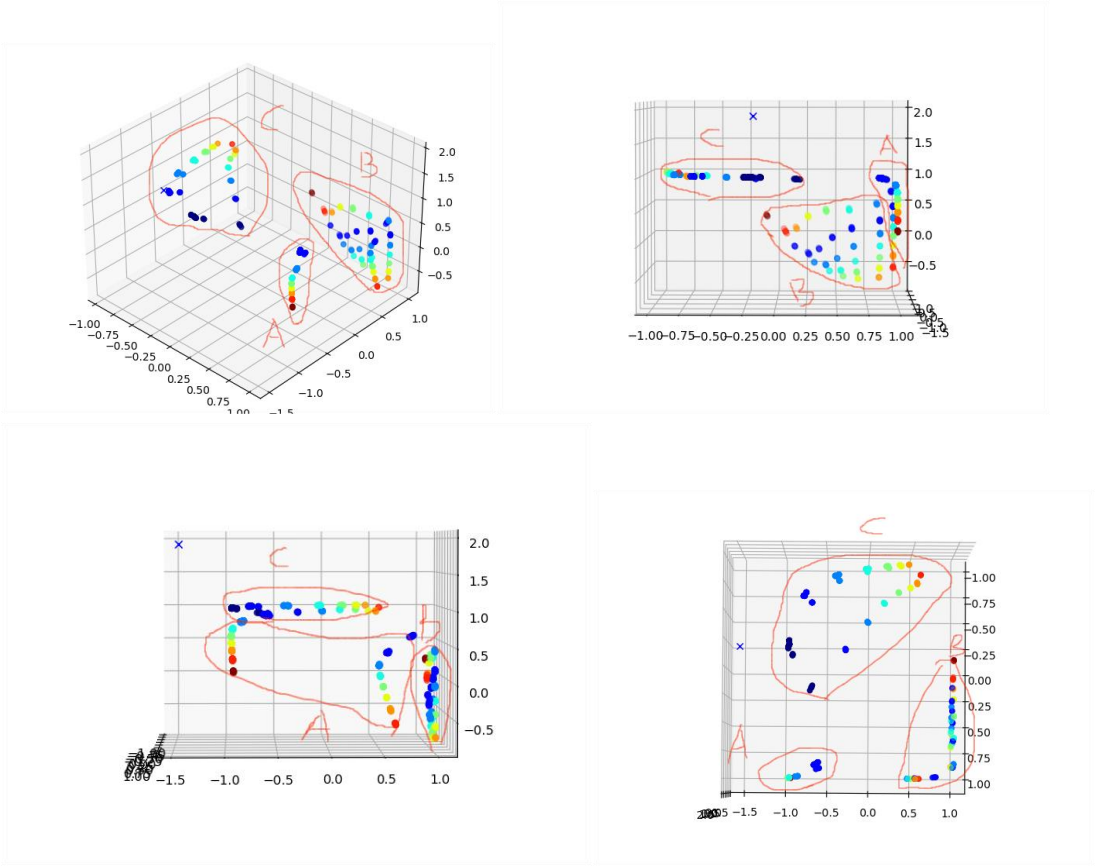


Question 3 :

Firstly, we can know quite clearly which clusters of A and B are some of them, and the simplest as well as the most straightforward way to tell is that if our training method is successful in predicting them, the next A corresponding to B will be indicated as the next rank of B's colour. There are some letters that the network can definitely predict, because it has a predictable pattern. These are represented by the "reddest" points in Q2. Thus, the first A in each sequence is the 'reddest' because it is definite and has a high activation on the Y-axis, as shown in the set in

the bottom left corner. All the B's that appear after the first B are deterministic and are in the upper right because they best activate the hidden units represented by the X-axis.
The darker blue dots represent 'uncertain' hypotheses. For instance, a sequence of 'A's cannot go on indefinitely, but there is no definite way to determine how many 'A's will appear. Thus, after each successive "A" hypothesis, the network reduces the probability that the next hypothesis is A and increases the probability that the next hypothesis is B.

Question 4 :



Question 5 :
The three hidden units' activation correspond to each letter(A,B,C) are shown by the graphs above.
By isolating three 2D perspective from the 3D image,we can easily get the following rules:
1.There is not much x's variation in A.
2.There is not much y's variation in B.
3.There is not much z's variation in C.
Among these 3 letters, the letter C has the least variation. C falls almost exactly in the plane of z = 1, while A and B have insignificant deviations. This is an expected result since C is the only one which has the deterministic category. There is a certain probability that A and B will fall because it is not known when the sequence A will stop and the sequence B will start.
Thus, the activation of the hidden node on the X-axis represents the probability of becoming A.
The activation of the hidden node on the Y-axis represents the probability of becoming B.
The activation of the hidden node on the Z-axis represents the probability of becoming C.

Question 6 :

LSTM is suitable for processing and predicting important events with very long intervals and delays in time series.

Here the LSTM model works:

1. initialize model and nodes' weights
2. For each number in the sequence do the following steps:
    a. Make gates based on the weights.
    b. Utilize the Sigmoid function to "ignore" some contexts if it is below some thresholds.
    c. Change inputs with Sigmoid and tanh.
    d. Use a Sigmoid function to output result.
    e. Update hidden units / gates/thresholds.
    f. Go back to step a.(recurrent)

3.Output

Epoch	Hid1	Hid2	Hid3	Hid4
0	-0.2	-0.07	0.06	0.12
5000	-0.97	0.73	0.47	-0.91
10000	-0.93	0.74	0.24	-0.95
15000	-0.97	0.72	0.93	-0.98
20000	-0.98	0.72	0.53	-0.97
25000	-1	0.75	0.68	-0.90
30000	-1	0.75	0.69	-0.91
35000	-1	0.75	0.74	-0.94
40000	-0.99	0.75	0.75	-0.95
45000	-1	0.76	0.73	-0.98
50000	-1	0.75	0.74	-0.96

We can easily know that the model is learning,over time the LSTM model can effectively have long time memory through these hid nodes as they approach a value nearly {-1,0.75,0.75,-1} which can help it lead to the highest accuracy in its predictions.