

Question 1)

Matlab code for visual representation.

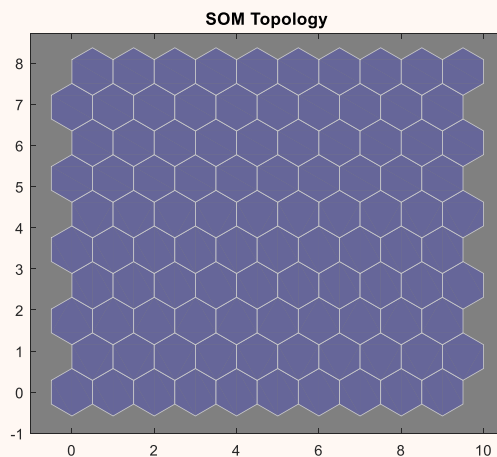
For data visualization in 4D, we use `plotMatrix`, which is a function that creates a 4x4 matrix with 2D data. The figure is arranged in a way that the position (m,n) of the matrix is the representation of the m and n dimensions. If m=n, then it represents a graph with the distribution of that axis on our data.

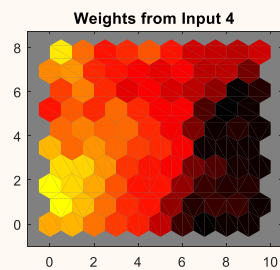
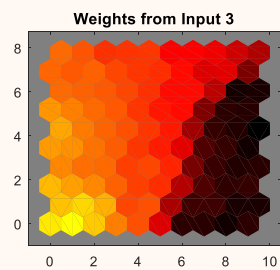
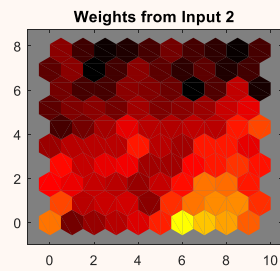
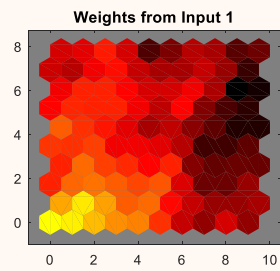
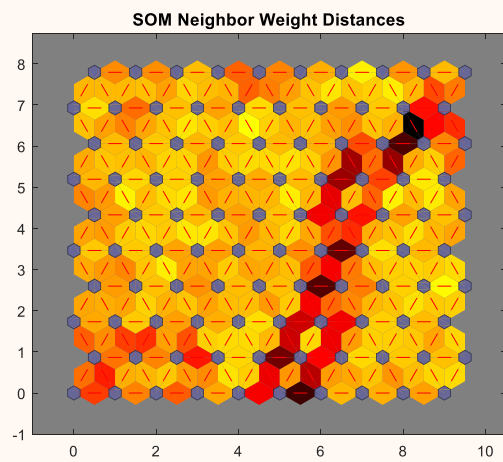
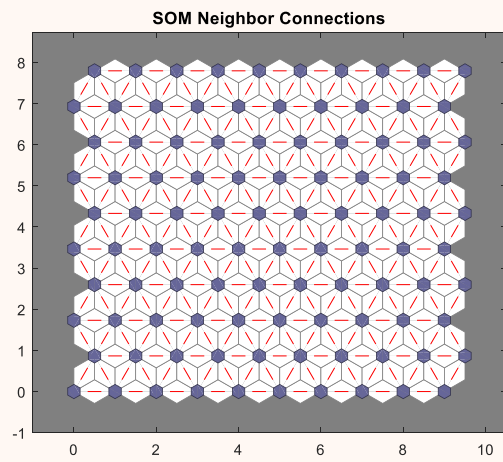
In the second figure we have a representation of what the results are depending on the number of clusters used, in this case we can say that the results with 2 and 3 clusters are similar and really good ones (because their values approach a lot to 1). The Silhouette Value represents the probability that the value belongs to the correct cluster. In some cases, we can find negative values that mean that those values probably don't belong to the correct cluster. Finally, we can say that the result is not the same result always, if we repeat the process, we can see that the values change between iterations of the script.

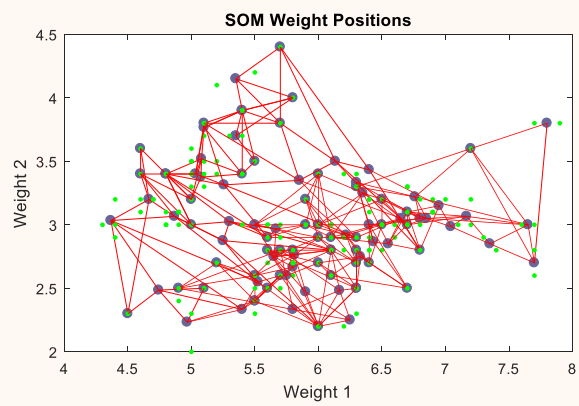
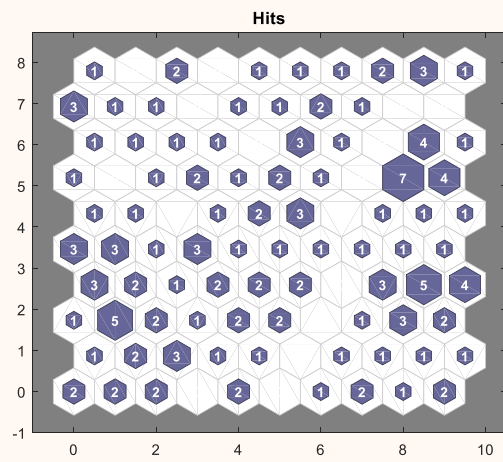
Question 2)

For this task, we are going to use the Matlab tool for neural network clustering called `nctool`. The outputs are different depending on the size of the 2D map:

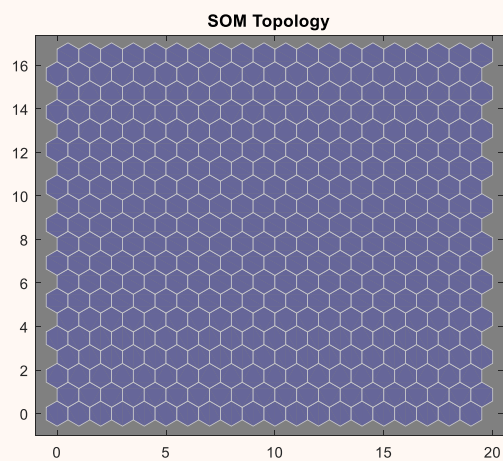
- Size = 10: we reached maximum epoch, so the training had to stop:

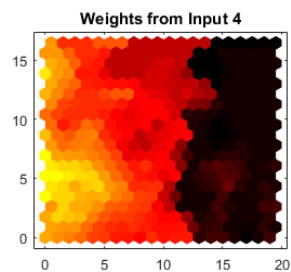
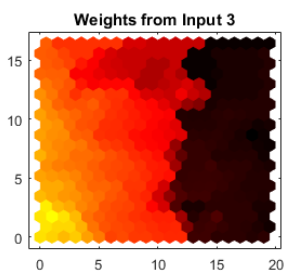
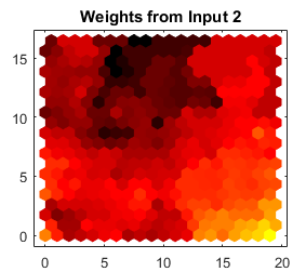
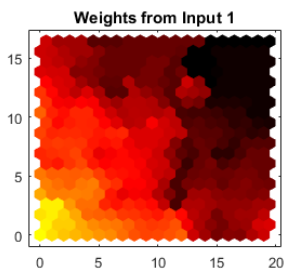
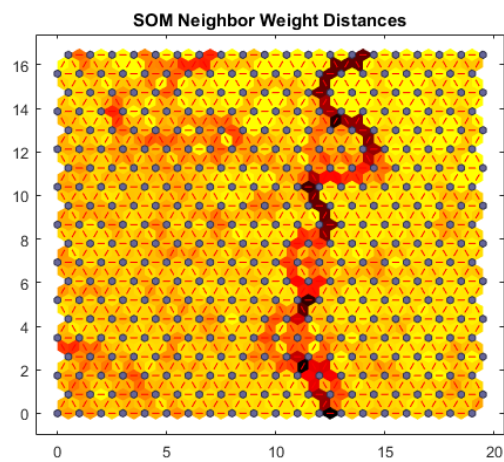
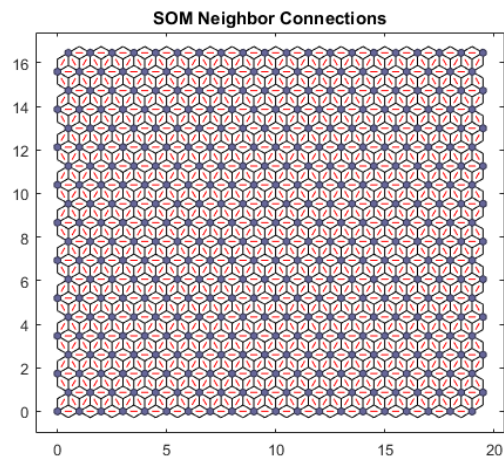


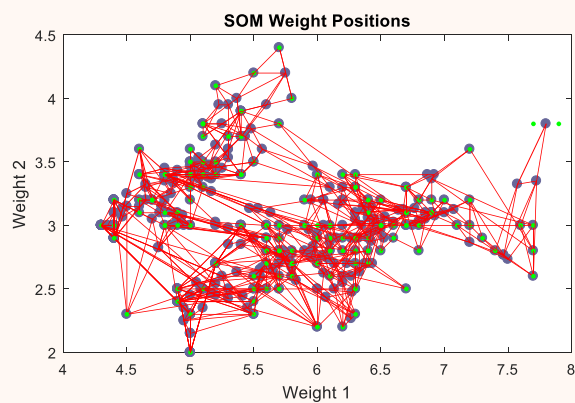
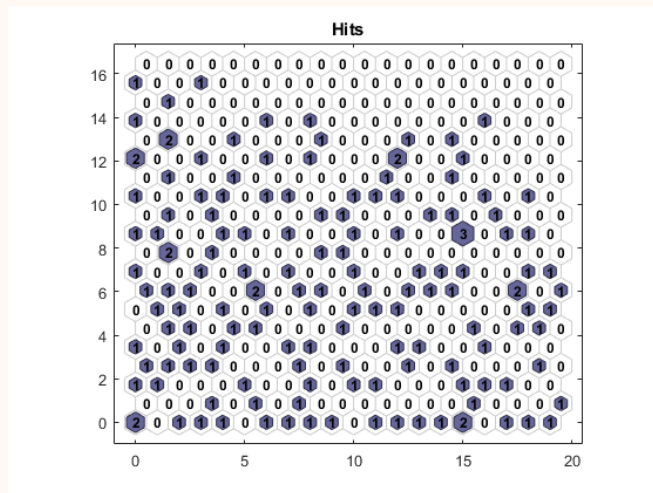




- Size = 20: the training had to stop again because we reached the maximum epoch:







In the end, the tool helps us to visualize and analyze the data in an easy way: each hexagon is a neuron and depending on the size of the map they will be bigger or smaller. On the hits figure we can see how many input vectors have gone through each one of the neurons and how they have been spread out in the map. The more size of the map, the longer time it takes for Matlab to train the network, so we should select reasonable values.

The samples are grouped as we expected, and we can say so by comparing the results to the 1st question of the task. Therefore, the results of the training are correct.