Lab 3 - Natural Computational Methods

Alejandro Villaron and David Håkansson

March 2021

Question 1: For som_P10 there is no overlap, For som_P20 theres is some overlap. For som_P30 there is even more overlap, we estimated there to be bout 15-20 overlapping nodes for som_P30.

Plot 1:

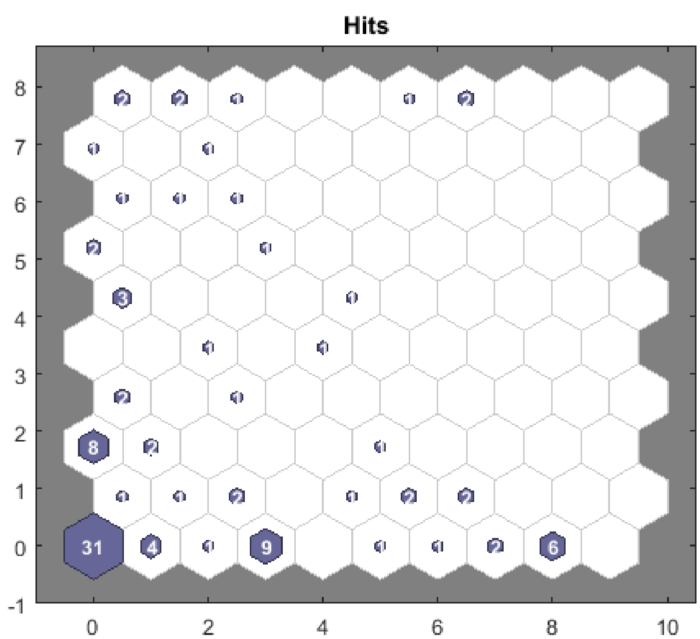


Figure 1: *Plot of cluster F1 for the SOM10 with the P30 dataset*

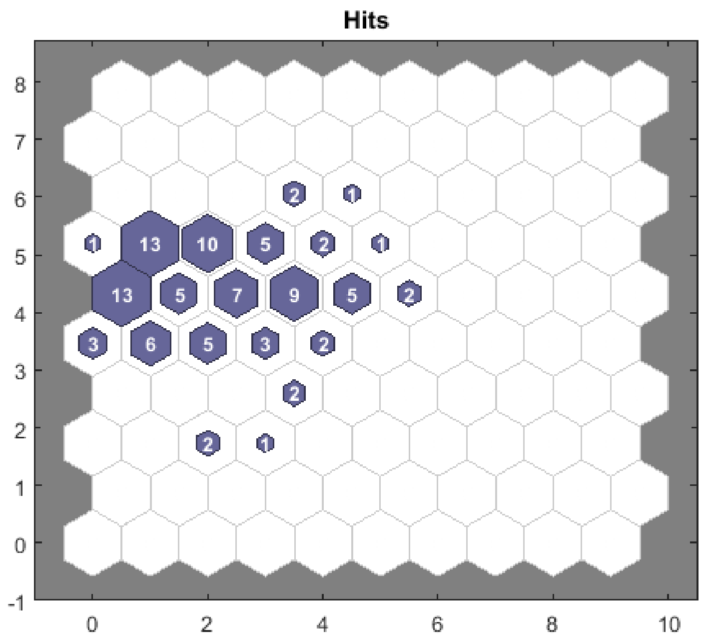


Figure 2: *Plot of cluster F1 for the SOM30 with the P10 dataset*

Question 2: som_P30 with P10 has a lot of hits for one node now. The reason for this is because the som_P30 nodes are are more spread out since P30 is more spread out compared to P10. We will have a network where the average node have a large voronoi region and then we will input data that only cover a small area, having the effect of a lot of data being covered by very few nodes.

In som_P10 with P30 we have a very compact network with very spread data. This will have the effect of the center most nodes in the node cluster not covering much if any data while the nodes at the outskirts will cover a lot of data. This is because what was considered to be in the outskirts have changed. Where when we trained the data the outskirts where much closer to the center. This means that the voronoi regions that originally covered the points at the outskirts will now cover points that are central and points at the outskirts.

Plot 2:

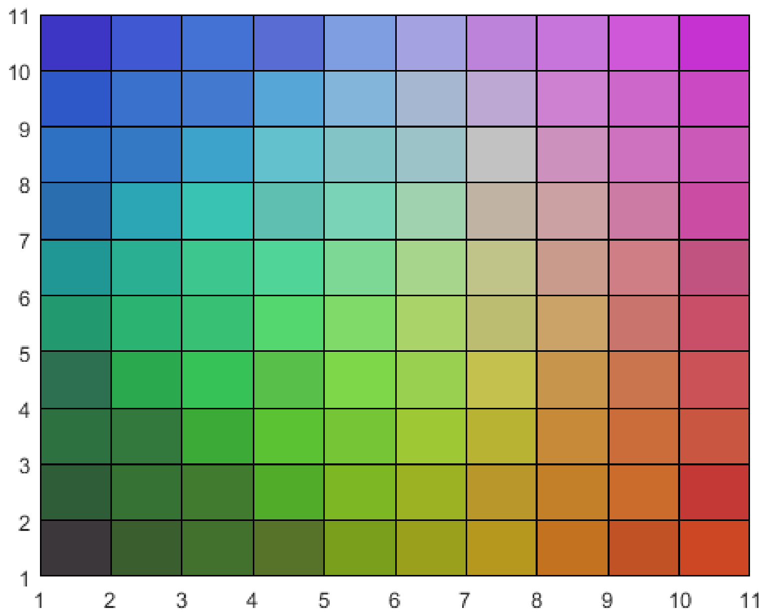


Figure 3: *Smoothest color map achieved with the configuration of 150 epochs in the ordering phase, neighborhood size of 8 and 90 epochs during the tuning phase.*

Question 3: We got an okay performance, where we could cluster most of it separately as seen in figure 4. The first two classes are easy to separate but the third one had some collisions with the second class' nodes. If we take a look at the data analyzed, it makes sense, since the values corresponding to the 3rd class are very close to the values corresponding to the 2nd class. This would make us consider the first class as the easiest to separate. We used 450 epoch during the ordering phase with a neighbourhood size of 6, and 100 number of epochs during the tuning phase.

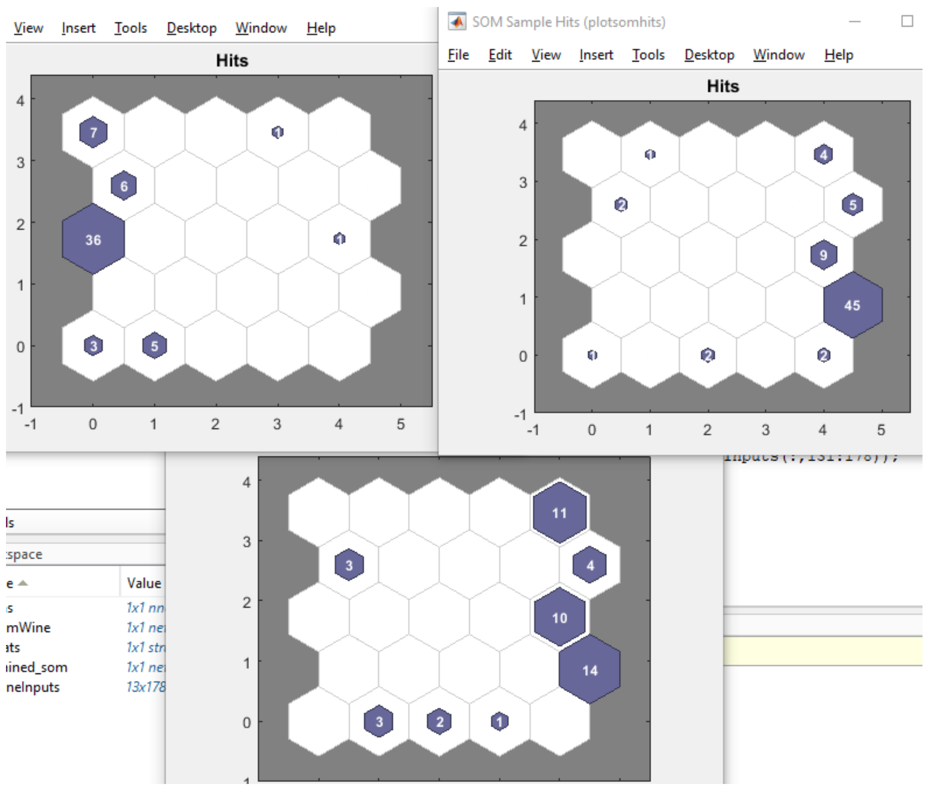
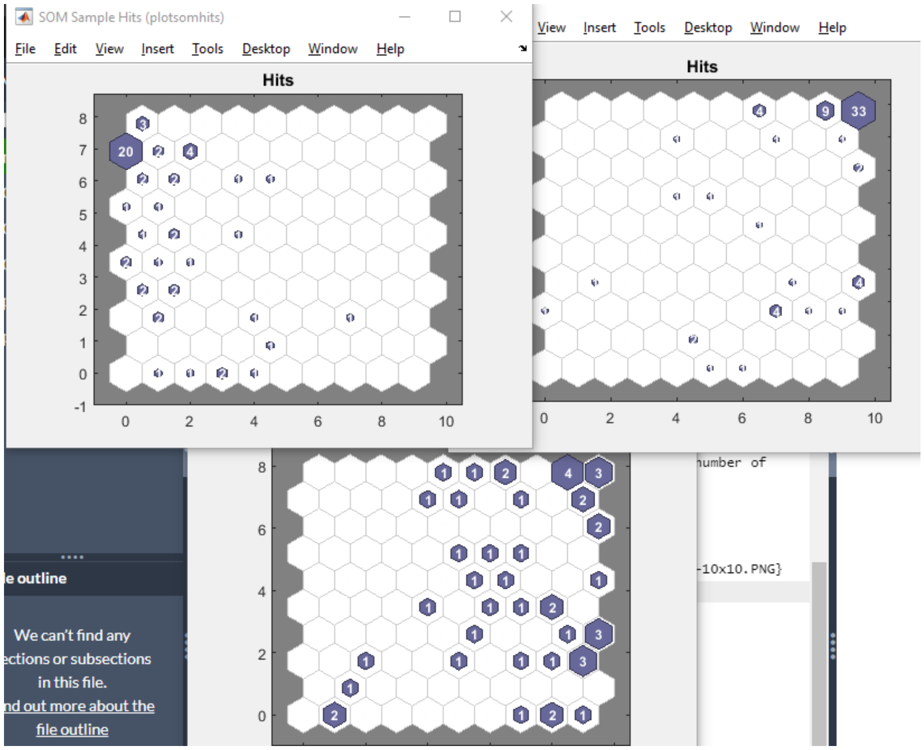


Figure 4: Question 3

Question 4: We get about the same results but everything is more spread. Also the third class is We used 500 epoch during the ordering phase with a neighbourhood size of 6, and 100 number of epochs during the tuning phase.



Question 5: We do get a lot better results now, the third class that previously was very chaotic now is as stable as the other classes.

Question 6: We think the classes get separated relatively well with one in the left corner, one to the right and one around the middle.

Question 7: Yes there is a clear dark line between two classes, this means that there is a large distance between the nodes in the two sections, in other words it means there are different clusters. If we know that there is at least two different clusters we can assume that there are different species or distributions the data is drawn from.

Question 8: Yes. First, third and fourth inputs (attributes) show the same pattern of clusters, having two more noticeable distinct areas in the bottom left and right corners, and one more hard to spot in the top. These results correspond to the distribution of the sample hits obtained, where the first group is located in the bottom left corner, the second group mostly on the top left part, and the last group on the right side of the diagram.

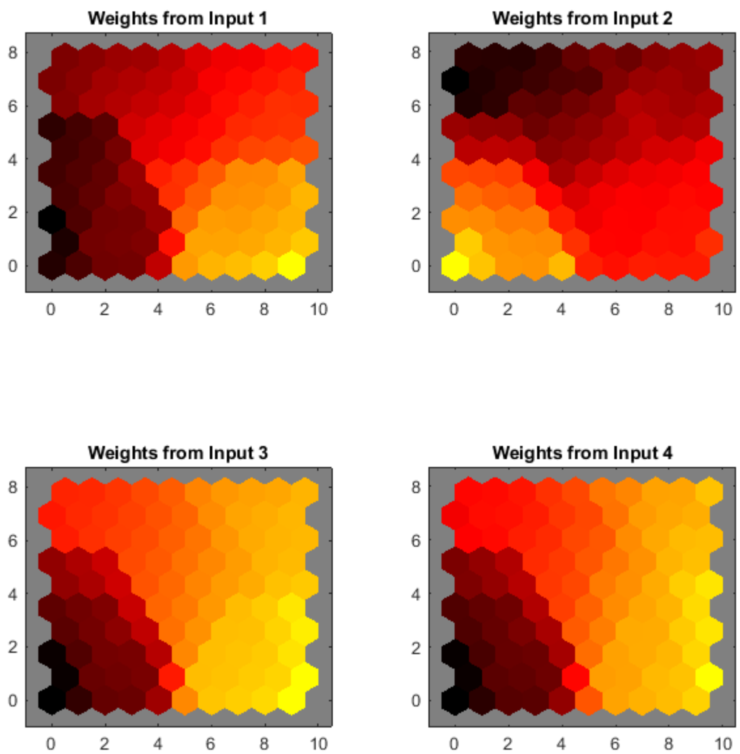
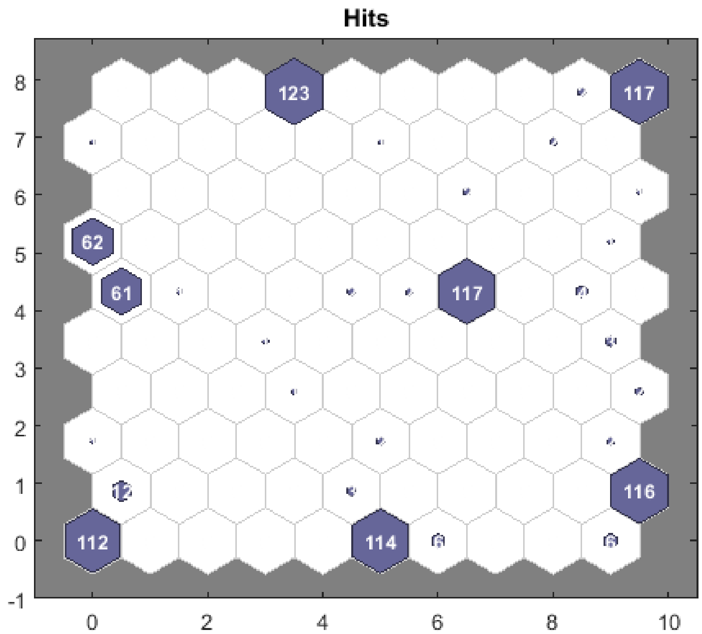


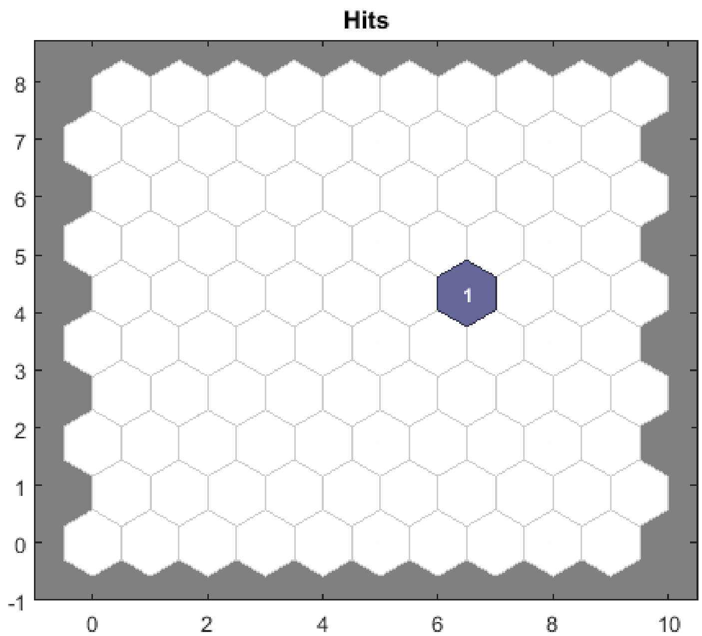
Figure 5: *Result of the Input Planes for the 4 different attributes*

Question 9: We know from the previous graph that the three dimensions of input correspond to the distributions of the three types of flowers. Also from that graph we know that the lighter the color the higher the value of the variable represented. From this information we can tell that in terms of petal size (third and fourth inputs) the values corresponding to the bottom left clusters (First group of flowers) are the smallest, and the values of the bottom -right clusters (third group of flowers) are the highest.

Question 10: As we can see in the image below there are 7 clusters.



Question 11: By plotting which node the different points are in, we see that point one and four belongs to the same cluster, the same is true for three and five. The figure below shows an example of how we plotted it.



Question 12: As we all know the world is three dimensional but for certain applications like google maps and such we need to work in a 2d space. Since SOFM is topologically preserving we should be able to turn a 3d map of the world into a 2d one(although either the sizes or the angles will be correct). So as input for this we would want a 3d world map.