

This task is to implement an unsupervised learning algorithm, Self Organizing Map (SOM), on the 4-dimensional Iris dataset. This dataset consists of 4 features for each of the 3 species: Setosa, Versicolor, and Virginica. The 4 features likely represent sepal length, sepal width, petal length, and petal width.

The algorithm was initialized with a 40x40 grid of random weights. The weights were then adjusted iteratively using a neighborhood function, with the goal of having similar data samples map to adjacent grid points. Parameters such as learning rate ( $\eta$ ) and neighborhood width ( $\sigma$ ) were decreased over epochs to stabilize the learning process.

Figure 1 below presents the results. On the left, the initial random weights are displayed, with samples appearing scattered without clear distinction. The right side shows the final weights after training. Here, the data samples are organized into three distinct clusters, each corresponding to one of the Iris species.

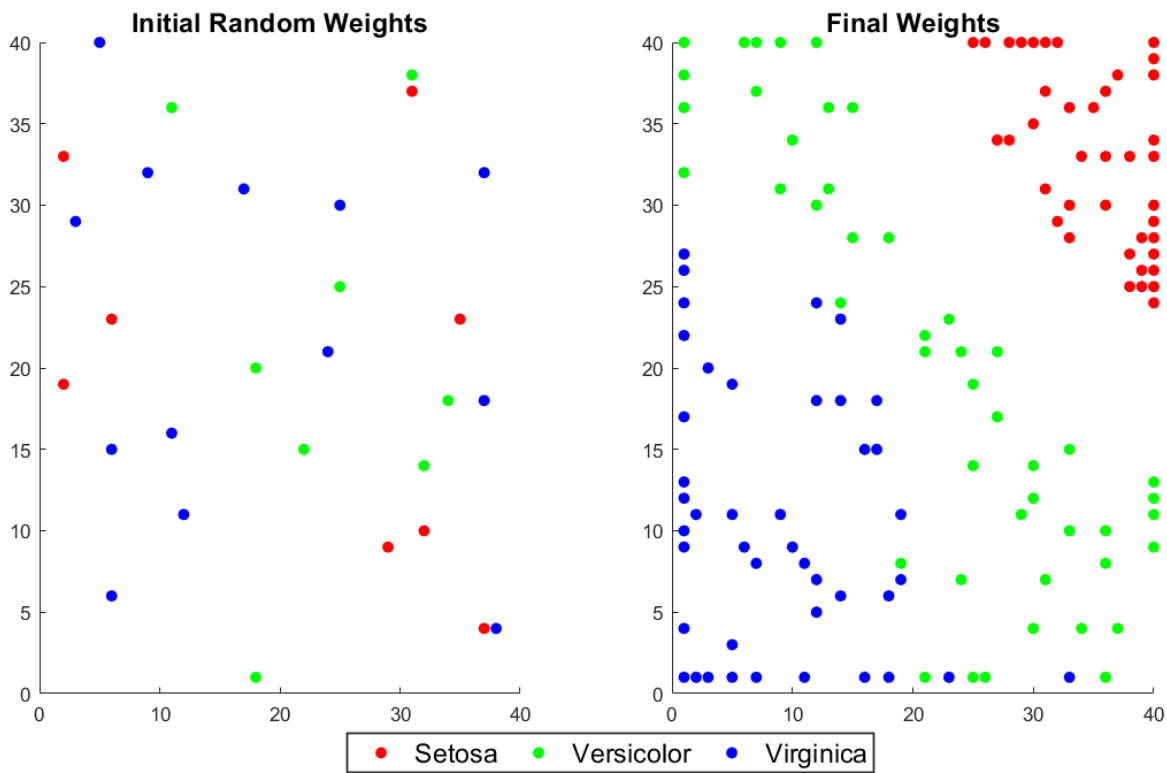


Figure 1: Initial random weights (left) compared to the organized final weights (right). On the right, the samples are grouped into three clusters, each corresponding to a flower species.

The SOM successfully mapped the Iris dataset onto a 2D grid, creating three distinct clusters. However, the Setosa (red) is distinctly separated from the other two species. While Versicolor and Virginica have different clusters, some of their points overlap. This overlap could be attributed to the inherent similarities between the features of Versicolor and Virginica. There's potential for fine-tuning parameters, such as the learning rate, neighborhood width, and their decay rates, to achieve better separation.

After some research on multiple sources that have analyzed the Iris dataset, it's clear that Setosa is more distinguishable. In contrast, Versicolor and Virginica exhibit closer feature similarities, leading them to be slightly less distinct from each other.