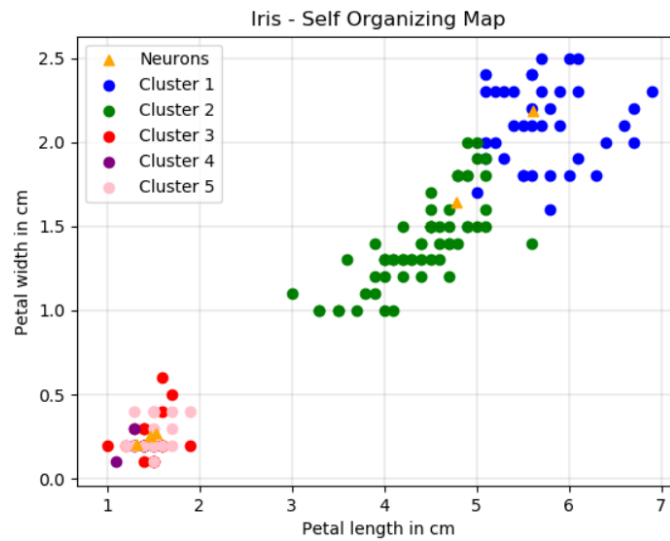# Lesson 5 Report – Iris Data Set

Thi Thu Trang Do 17774

1. K = 5 (5 clusters)

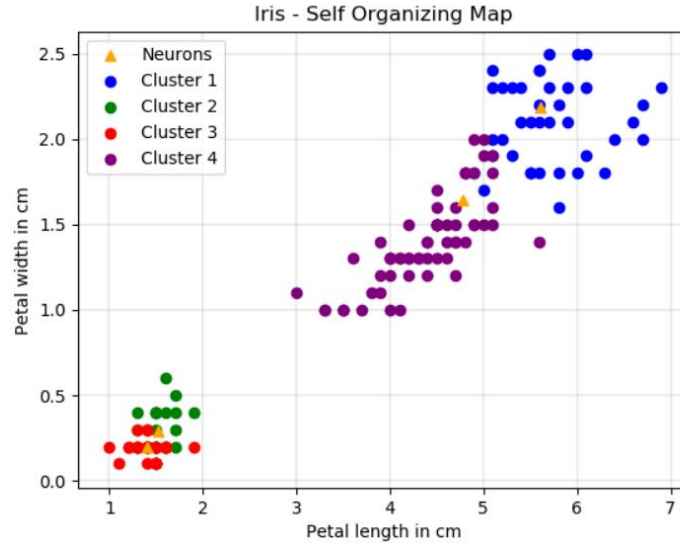Cluster 1: Dominated class: Iris-virginica - fraction: 0.9736842105263158

Cluster 2: Dominated class: Iris-versicolor - fraction: 0.7903225806451613

Cluster 3: Dominated class: Iris-setosa - fraction: 1.0

Cluster 4: Dominated class: Iris-setosa - fraction: 1.0

Cluster 5: Dominated class: Iris-setosa - fraction: 1.0
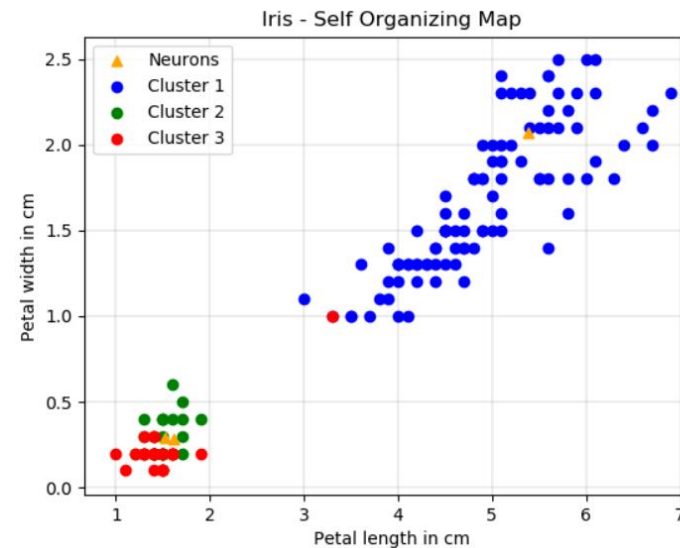
2. K = 4

Iris - Self Organizing Map

Cluster 1: Dominated class: Iris-virginica - fraction: 0.9736842105263158

Cluster 2: Dominated class: Iris-setosa - fraction: 1.0

Cluster 3: Dominated class: Iris-setosa - fraction: 1.0

Cluster 4: Dominated class: Iris-versicolor - fraction: 0.7903225806451613

3. K = 3

Iris - Self Organizing Map

Cluster 1: Dominated class: Iris-virginica - fraction: 0.5050505050505051

Cluster 2: Dominated class: Iris-setosa - fraction: 1.0

Cluster 3: Dominated class: Iris-setosa - fraction: 0.9583333333333334

Impact of K on result:

For k = 5, the cluster 3,4,5 are full of Iris-setora, they should be grouped into 1 cluster. Cluster 1 is good (0.97) and cluster 2 is not too bad. In overall, the algorithm works well with k = 5.

For k = 4, the clusters look quite the same as k = 5, but now Iris-setora is dominated in only 2 clusters (cluster 2 and 3). It is better than k = 5.

For k = 3, cluster 1 looks quite bad while clusters 2 and 3 are good.

The algorithm works the best when k = 4 for Iris data set.