Assignment 3
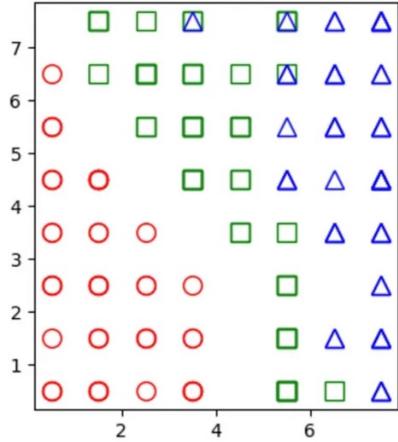
For this assignment, you're required to <u>implement from scratch</u> the **self-organizing maps**.

1- Please use the provided <u>breast cancer dataset</u>, which classifies the cancer as Benign or Malignant based on a set of features. (Please refer to https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic for a better understanding of the dataset).

   - Review the dataset and remove any irrelevant columns.
   - Normalize the data before training.

2- Follow the steps described in tutorial to implement a self-organizing map. Use the same described decay function f(t) to decay the learning rate:

$$\eta(t) = \eta_0 * f(t) \text{ , where } \eta_0 \text{ is the initial learning rate}$$

3- Train your map on the data. Set a suitable initial learning rate and grid size for your map. Describe why you chose the selected grid size.

4- <u>Visualize</u> your map (see hints below) and <u>comment</u> on the result:

   - After training, you should have a map of neurons, each neuron having a weight vector.
   - For each data point in your dataset, find its best matching neuron and assign the class of this data point to its best matching neuron. Plot the result (see figure below as an example on a different dataset. The 3 symbols correspond to 3 classes. Notice that some neurons can have no class assigned to them as they were not chosen as a BMU for any data point).

Instructions

1- The assignment is **<u>individual</u>**. Sharing code is not accepted. Cheating will result in a zero grade.

2- <u>Add comments</u> that describe your code.

3- Your conclusions and comments on the results can be included in markdown cells within your code.

4- **Late submissions** will be penalized.

5- **Submission** will be through providing me with a <u>link to your code on Google Colab</u>. The link should be <u>submitted and turned in on Google classroom</u>.
Please remember to provide me with access to the code as an <u>Editor.</u>

6- Editing your code after the submission deadline is not accepted.