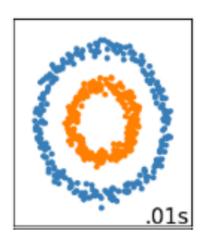
## Machine Learning Second Project

- 1. Consider the dataset `driver-data.csv`.
  - a. Implement the k-means algorithm with k=4 on this dataset and display the data and clusters (colors) on a page.
  - b. Implement the kernel k-means algorithm with k=4 and a Gaussian kernel on this dataset and display the clusters.
  - c. Cluster the data using the EM method with k=4. The initial assumptions for the clusters are as follows:

$$P(C1) = P(C2) = P(C3) = P(C4) = 0.25$$

- d. Generate the means and variances of the clusters randomly.
- e. Determine which clustering method has the best performance using the BetaCV method.
- 2. From the scikit-learn website, download the circle dataset shown below.
  - a. The clustering above shows the result of the DBSCAN algorithm. Implement the DBSCAN algorithm on this dataset. Can similar results be achieved?
  - b. Implement the k-means algorithm on this dataset and explain the reason for the superiority of the previous method.



## Machine Learning Second Project

3. Download the dataset `nerve`. Consider the following kernel functions for a given data point.

$$K(x) = 0.5I(x)$$

$$K(x)=rac{1}{\sqrt{2\pi}}e^{-rac{x^2}{2}}$$

$$K(x)=\frac{3}{4}(1-x^2)I(x)$$

Where:

$$I(x) = egin{cases} 1 & ext{if } |x| \leq 1 \ 0 & ext{if } |x| > 1 \end{cases}$$

Notice that the data points in this dataset are one-dimensional. Now, with different steps (h), estimate the density for each of the kernels, and describe the results.

Dataset links:

https://github.com/moein-monemi/datasets.git

https://scikit-learn.org/stable/modules/clustering.html