# 10   Final lab

You can use external libraries for linear algebra operations but you are expected to write your own algorithms.

## 10.1   Exercise 1

Implement your own versions of:

- Density peaks clustering (without the "halo" feature)

- Normalized Mutual Information (NMI)

- F-ratio

Using the `Aggregation` dataset available at `https://cs.joensuu.fi/sipu/datasets/` under Shape sets:

1. Use **your** version of density peaks at several values of the $\epsilon$ parameter and choose the one that maximises the NMI with the ground truth (third column).

2. Compute the NMI with the ground truth and the F-ratio test for this assignation.

3. Perform $k$-means with $k$ equal to the true number of clusters, compute the NMI with the ground truth and the F-ratio test.

4. Apply the `sklearn` implementation of DBSCAN and MeanShift to the dataset. Discuss the differences among the obtained results.

### Note

Hints for point 1. As a rule of thumb, $\epsilon$ should be chosen in such a way that, on average, between 1% and 2% of the data points are included in the neighbourhood. Use the exponential kernel.

### Side Note

I will upload the solutions to this exercise on `github` in a week's time. I'll notify you on Teams when the solution will be available online.