Lab4 – Clustering

Artificial Intelligence and Pattern Recognition

In this exercise, we will be focusing on clustering using different methods. The first part is meant to teach how to implement k-means as a distance-based classifier. The second part will focus on a different method relying on density to perform the task and called DBScan.

Part 1

K-means

- Apply k-means on the iris data
- Evaluate the performance of the model using k=5
- What will be the optimal K value for the used dataset 'Iris' which has 4 different features?

DBScan

- Apply DBscan on the iris data
- Evaluate the performance of the model using epsilon=1 and Minpoint=10
- Which values will you recommend for this model to get the best performance? (try this one manually and find out the values)

Conclusion

- Which model (Kmeans, DBScan) is better for the r Iris dataset

Part 2

Let us generate our data with the help of RapidMiner and explore which clustering method is the best for this data

- Generate data using a target function named three ring clusters by setting
    - Number of samples to 1000
    - Attributes (features) to 2
    - Largest radius 20
- Propose the best clustering method and justify your answer