Indian Institute of Technology Indore

Discipline of Computer Science & Engineering

CS 403/603 Machine Learning

Lab Assignment - K-Means Clustering Algorithm

Some general instructions:

- Plagiarism in any form will not be tolerated.
- You are not allowed to use in-built libraries related to the topic. Code everything from scratch.
- You are allowed to do only one submission before the deadline. However, in the case of multiple submissions, only the last submitted file will be used for evaluation.
- Submission of the assignment should be made using the Google Classroom platform only.
- Last date for submission of the assignment: **29th Oct 2021**

Problem Statement:

Use the "IRIS" dataset and create clusters using the K-Means Clustering Algorithm with a suitable value of k over 10 iterations.

Only the first 4 columns of the dataset are to be used as attributes. Consider Euclidean distance as the distance measure.

Use the ground truth cluster label present in the data set to compute and print the Jaccard distance of the obtained clusters with the ground truth clusters for each of the clusters.

What is the effect of changing the initial cluster centres?

Again compute the Jaccard distance between the clusters and analyze if there is any difference in the results.

Results:

Attach your code file and include a single write-up (pdf) file which includes a brief description explaining what you did. Include any observations and/or plots required by the question in this single write-up file.