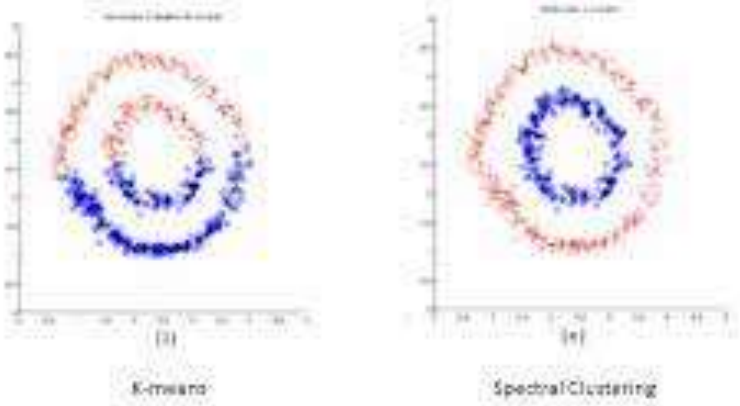
Spectral Clustering is a growing clustering algorithm which has performed better than many traditional clustering algorithms in many cases.

It treats each data point as a graph-node and thus transforms the clustering problem into a graph-partitioning problem.

k-means vs. Spectral Clustering



K-means

Spectral Clustering

1.Building the Similarity Graph:

This step builds the Similarity Graph in the form of an adjacency matrix which is represented by A. The adjacency matrix can be built in the following manners:-

- **Epsilon-neighbourhood Graph:** A parameter epsilon is fixed beforehand. Then, each point is connected to all the points which lie in it’s epsilon-radius. If all the distances between any two points are similar in scale then typically the weights of the edges ie the distance between the two points are not stored since they do not provide any additional information. Thus, in this case, the graph built is an undirected and unweighted graph.

K-Nearest NeighbourS : A parameter k is fixed beforehand. Then, for two vertices u and v, an edge is directed from u to v only if v is among the k-nearest neighbours of u.

Note that this leads to the formation of a weighted and directed graph because it is not always the case that for each u having v as one of the k-nearest neighbours, it will be the same case for v having u among its k-nearest neighbours.

To make this graph undirected, one of the following approaches are followed:-

1. Direct an edge from u to v and from v to u if either v is among the k-nearest neighbours of u **OR** u is among the k-nearest neighbours of v.
2. Direct an edge from u to v and from v to u if v is among the k-nearest neighbours of u **AND** u is among the k-nearest neighbours of v.

- **Fully-Connected Graph:** To build this graph, each point is connected with an undirected edge-weighted by the distance between the two points to every other point.
- Since this approach is used to model the local neighbourhood relationships thus typically the Gaussian similarity metric is used to calculate the distance.

2.Projecting the data onto a lower Dimensional Space:

This step is done to account for the possibility that members of the same cluster may be far away in the given dimensional space.

Thus the dimensional space is reduced so that those points are closer in the reduced dimensional space and thus can be clustered together by a traditional clustering algorithm.

It is done by computing the **Graph Laplacian Matrix** .

3.Clustering the Data:

This process mainly involves clustering the reduced data by using any traditional clustering technique – typically K-Means Clustering.

First, each node is assigned a row of the normalized of the Graph Laplacian Matrix.

Then this data is clustered using any traditional technique.

To transform the clustering result, the node identifier is retained.

Properties:

1. **Assumption-Less:** This clustering technique, unlike other traditional techniques do not assume the data to follow some property. Thus this makes this technique to answer a more-generic class of clustering problems.
2. **Ease of implementation and Speed:** This algorithm is easier to implement than other clustering algorithms and is also very fast as it mainly consists of mathematical computations.
3. **Not-Scalable:** Since it involves the building of matrices and computation of eigenvalues and eigenvectors it is time-consuming for dense datasets.