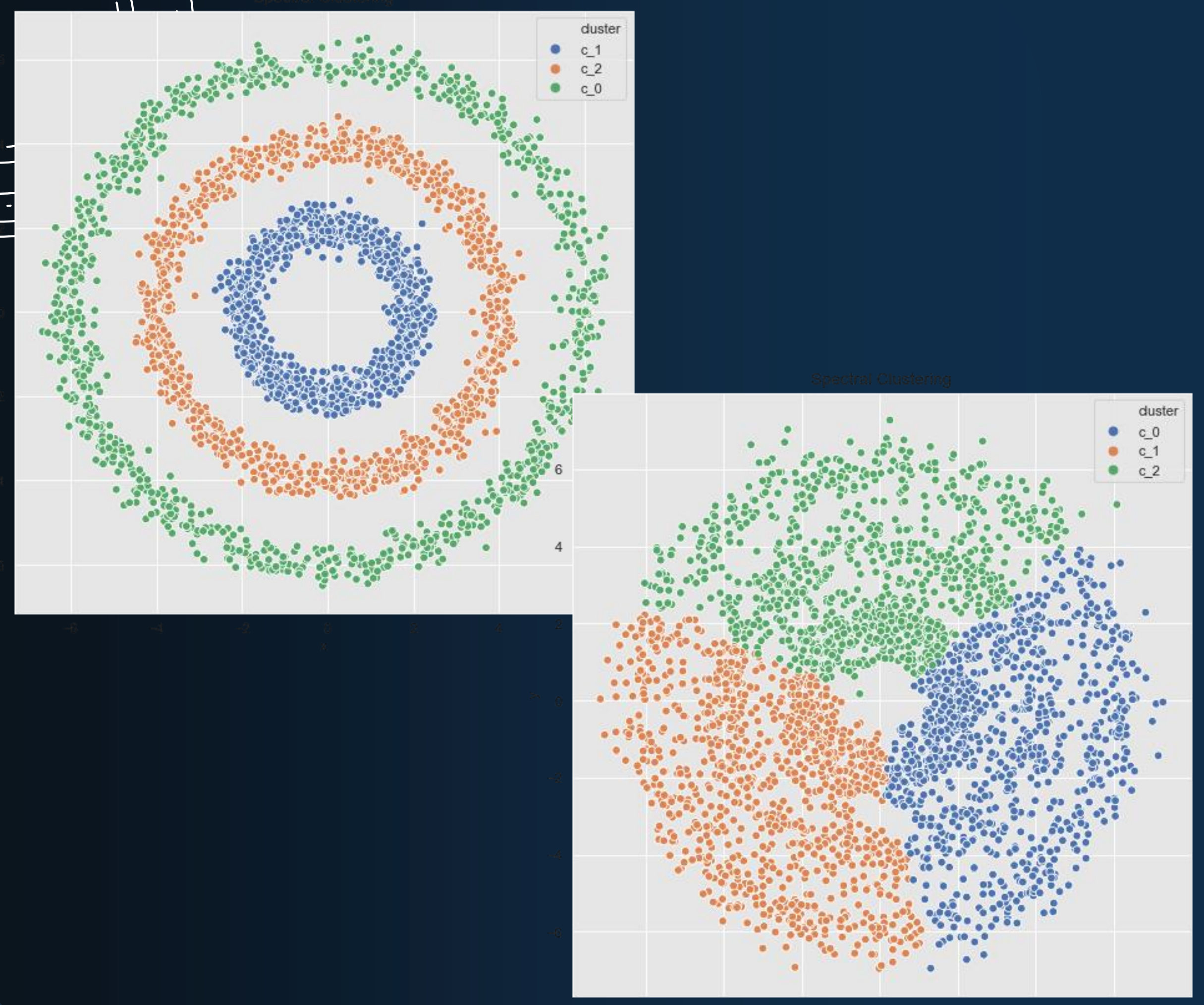
Spectral Clustering

MACHINE LEARNING CLUSTERING

Spectral Clustering

Introduction to Spectral Clustering

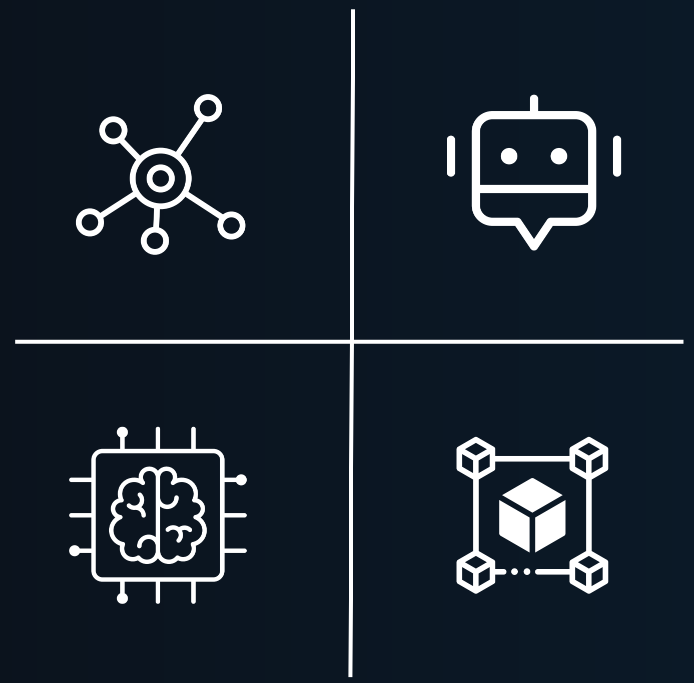
A clustering technique that uses the eigenvalues of a similarity matrix to perform dimensionality reduction before clustering. Captures the structure of data by considering the connectivity and proximity of points.



Key Concepts:

- **Similarity Matrix (W):** Represents the pairwise similarities between data points.
- **Laplacian Matrix (L):** Derived from the similarity matrix; crucial for spectral clustering.
- **Eigenvectors:** Used to map data points to a lower-dimensional space.

Mechanics of Spectral Clustering



1. **Construct Similarity Matrix:**
   - Compute the similarity matrix W from the data points.
2. **Compute Laplacian Matrix:**
   - $L = D - W$
   - Here, D is the diagonal degree matrix where $D(i,i) = \sum j W(i,j)$.
3. **Compute Eigenvectors:**
   - Find the eigenvectors corresponding to the smallest k eigenvalues of L.
4. **Form Lower-dimensional Representation:**
   - Use the eigenvectors to create a new representation of the data.
5. **Cluster in Lower-dimensional Space:**
   - Apply a clustering algorithm (e.g., K-means) on the lower-dimensional data to form clusters.

Application and Evaluation

Application:

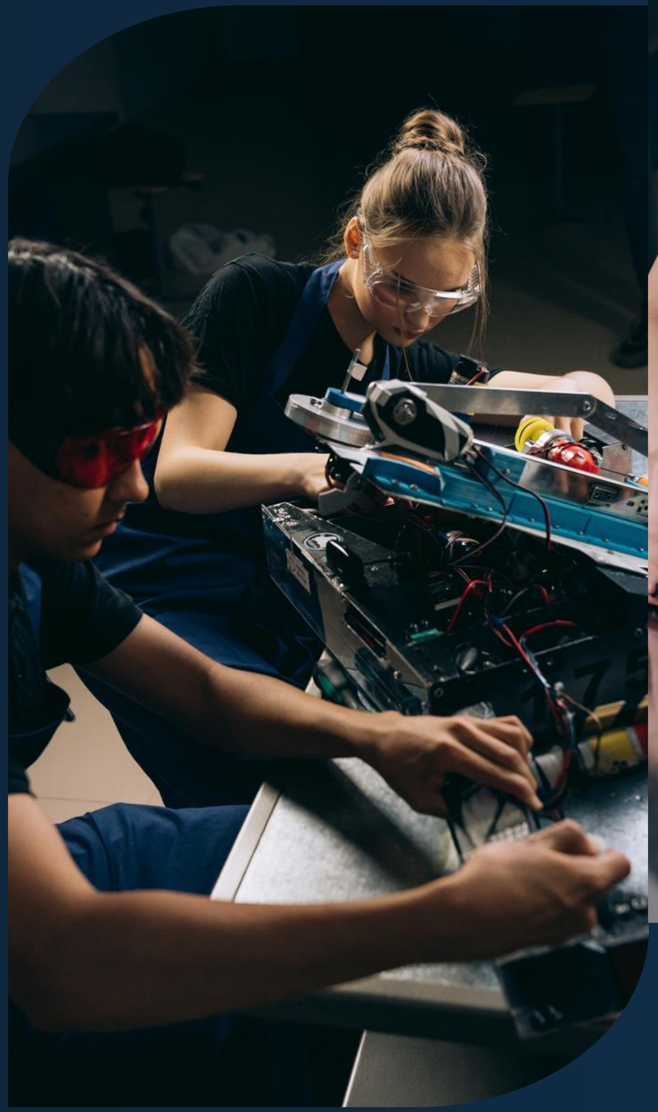
- Image segmentation, network analysis, document clustering.
- Steps:
  - ✓ Construct the similarity matrix.
  - ✓ Compute the Laplacian matrix and its eigenvectors.
  - ✓ Perform clustering in the lower-dimensional space.

Advantages:

- Can capture complex cluster structures.
- Works well with non-linearly separable data.
- No need to specify the number of clusters beforehand.

Disadvantages:

- Computationally intensive for large datasets due to eigenvector computation.
- Requires careful choice of the similarity function.
- Sensitive to the choice of the number of clusters (k).



Thank You