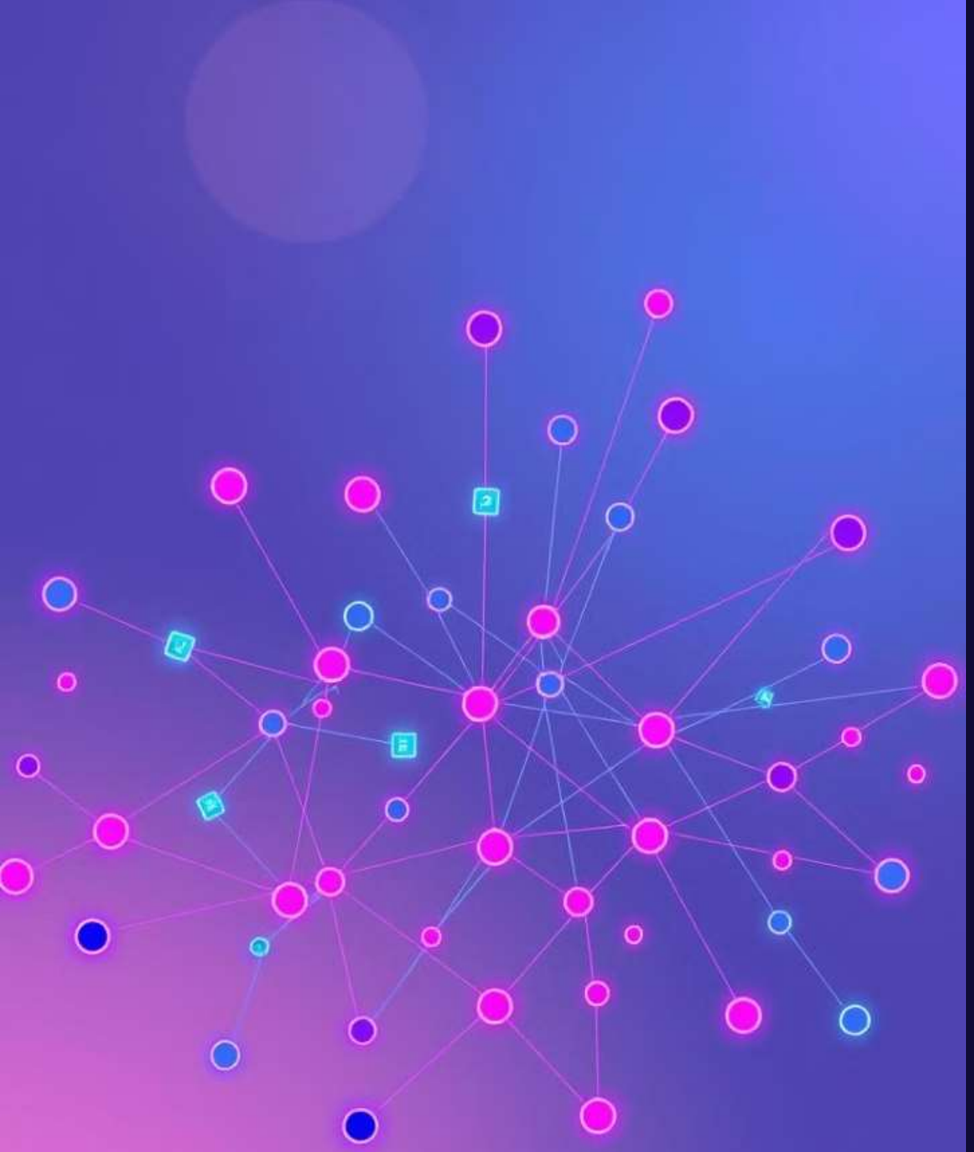
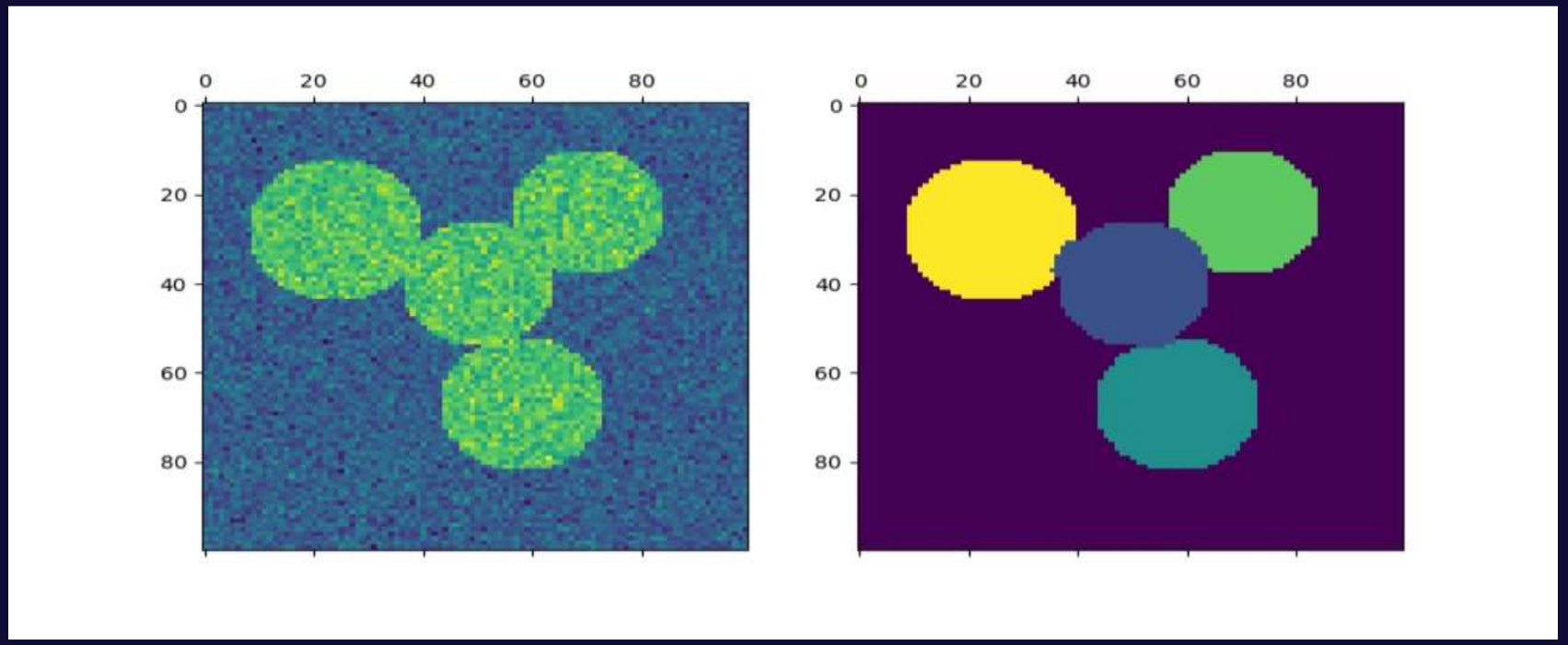


Spectral Clustering: Unlocking Complex Data

Welcome! Today, we'll explore spectral clustering, a powerful graph-based technique for uncovering hidden structures within your data. This method goes beyond traditional approaches and tackles the complexities of non-linearly separable data.

Data as a Network



Nodes as Data

Data points are visualized as nodes within a graph.

Edges as Similarity

Edges connecting these nodes indicate the degree of similarity between data points, with stronger edges representing greater resemblance.

The Algorithm Unveiled

1

Similarity Matrix

A matrix is constructed, with each entry representing the similarity between two data points.

2

Laplacian Matrix

The Laplacian matrix is derived from the similarity matrix, capturing the relationships between nodes.

3

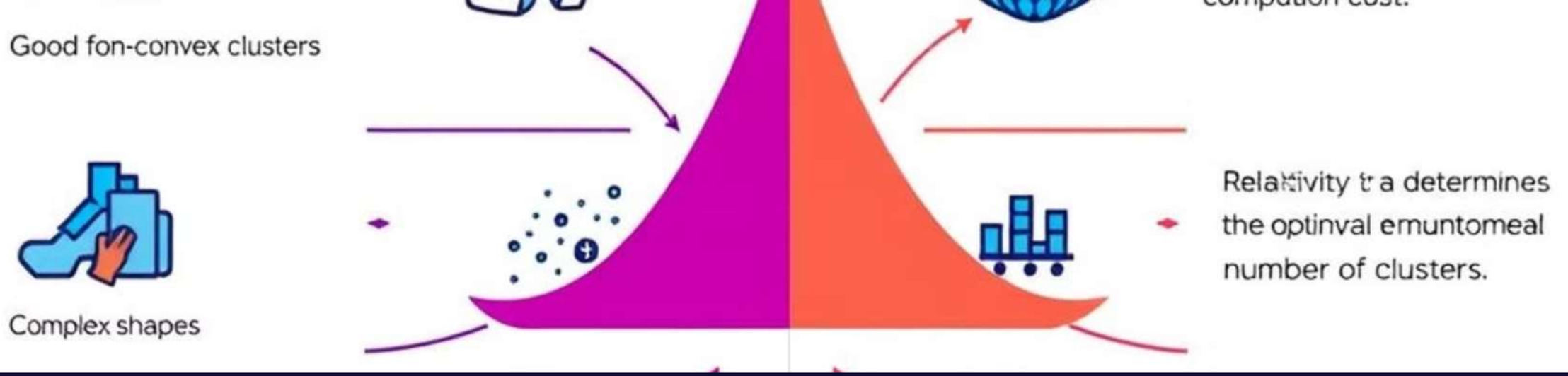
Eigenvector Extraction

The eigenvectors corresponding to the smallest eigenvalues of the Laplacian matrix are calculated.

4

Clustering

A clustering algorithm, such as K-means, is applied to group the eigenvectors, forming clusters.



Advantages and Disadvantages

Advantages

- Handles non-linear data
- Arbitrary cluster shapes
- Noise and outlier resilience

Disadvantages

- High computational complexity
- Choosing the number of clusters
- Sensitivity to similarity matrix