Spectral Clustering

Step-by-Step Calculation Example

Spectral Clustering Pipeline

1. Input Data
2D coordinate data

→

2. Similarity Graph
Build adjacency matrix
`A[i,j]`

→

3. Laplacian
Compute Laplacian
`L = D - A`

→

4. Eigen
Eigendecomposition
`Lv = λv`

→

5. Embedding
Low-dimensional space

→

6. K-means
Final clustering

Step-by-Step Calculation Example

1 Input Data

2 Adjacency Matrix

$p_1 = (1, 1)$	$p_2 = (1, 2)$
$p_3 = (2, 1)$	$p_4 = (5, 5)$
$p_5 = (5, 6)$	$p_6 = (6, 5)$

Data Description

6 points in 2D space

- p_1, p_2, p_3: Lower-left cluster
- p_4, p_5, p_6: Upper-right cluster

$A[i,j] = \exp(-||p_i - p_j||^2 / 2\sigma^2)$

Using σ = 1.0, higher weight for closer points

Matrix A (6×6)

1.00	0.61	0.61	0.00	0.00	0.00
0.61	1.00	0.37	0.00	0.00	0.00
0.61	0.37	1.00	0.00	0.00	0.00
0.00	0.00	0.00	1.00	0.61	0.61
0.00	0.00	0.00	0.61	1.00	0.37
0.00	0.00	0.00	0.61	0.37	1.00

Block diagonal structure represents two clusters

3 Degree Matrix

$D[i,i] = \Sigma_j A[i,j]$

Sum of edge weights for each node

Matrix D (6×6 diagonal)

4 Laplacian Matrix

$L = D - A$

Matrix representing graph structure

Matrix L (6×6)

2.22	0	0	0	0	0
0	1.98	0	0	0	0
0	0	1.98	0	0	0
0	0	0	2.22	0	0
0	0	0	0	1.98	0
0	0	0	0	0	1.98

1.22	-0.61	-0.61	0	0	0
-0.61	0.98	-0.37	0	0	0
-0.61	-0.37	0.98	0	0	0
0	0	0	1.22	-0.61	-0.61
0	0	0	-0.61	0.98	-0.37
0	0	0	-0.61	-0.37	0.98

Block diagonal structure = two separated clusters

5 Eigendecomposition

$Lv = \lambda v$

Select eigenvectors corresponding to k smallest eigenvalues

Eigenvalues

λ_1 = 0.00 (first, connected graph)

λ_2 = 0.00 (second, 2 clusters!)

λ_3 = 0.61

λ_4 = 1.59

6 Spectral Embedding

$X = [v_2, v_3, ..., v_{k+1}]$

For k=2 clusters, use v_2 and v_3

Embedding Coordinates (6×2)

$\lambda_5 = 1.59$

$\lambda_6 = 1.83$

⚠ Key Observation

2 eigenvalues near zero → 2 clusters exist!
Number of eigenvalues = Number of connected components (clusters)

Second Eigenvector (v_2)

$v_2 = [0.58, 0.58, 0.58, -0.58, -0.58, -0.58]^T$

Positive values → Cluster 1 (p_1, p_2, p_3)
Negative values → Cluster 2 (p_4, p_5, p_6)

0.58	-0.12
0.58	0.45
0.58	-0.33
-0.58	-0.12
-0.58	0.45
-0.58	-0.33

First column (v_2) clearly separates clusters
Positive (0.58) vs Negative (-0.58)

7 K-means Clustering

K-means(X, k=2)

Perform K-means in embedding space

✅ Final Cluster Assignment

Cluster 1: $\{p_1, p_2, p_3\}$
Cluster 2: $\{p_4, p_5, p_6\}$

🎯 Why Does This Work?

1. Original space requires non-linear separation
2. Laplacian eigenvectors reflect graph structure
3. Embedding space allows linear separation
4. K-means easily discovers clusters

💡 Key Insights

Advantages of Spectral Clustering

- ✅ Can find non-convex clusters
- ✅ Directly utilizes graph structure
- ✅ Can estimate number of clusters from eigenvalues
- ✅ Theoretical guarantees (minimizes normalized cut)

Complete Process Summary

Data → Similarity Graph → Laplacian → Eigenvectors → Low-dim Embedding → K-means