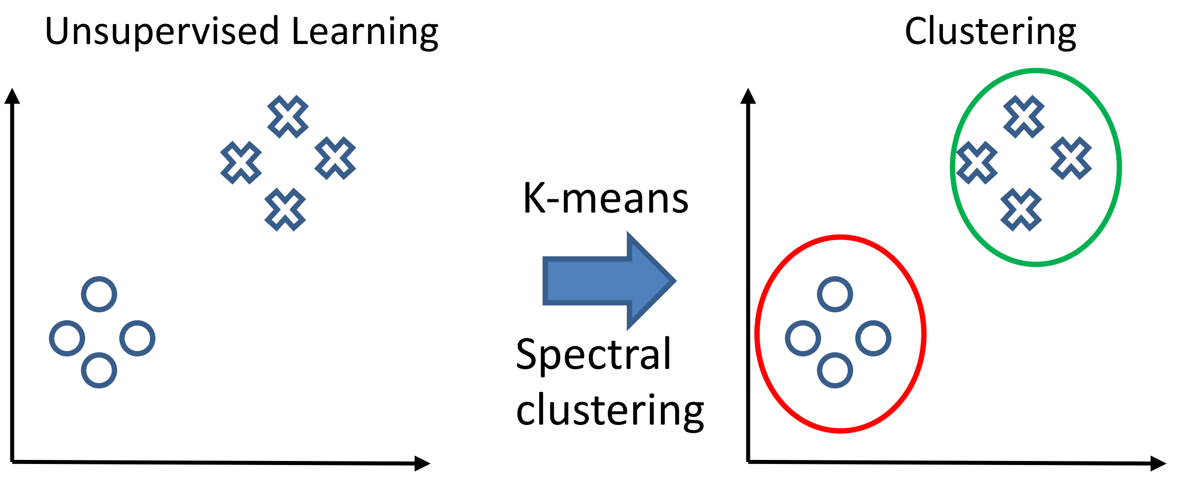
Spectral Clustering

Hongchang Gao

Spring 2024

What is clustering?

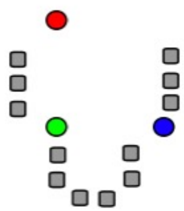
- Unsupervised Learning
  - Given: only samples, no labels
  - Clustering: find meaningful groups of samples s.t.
    - Samples in the same group are "similar"
    - Samples in different groups are "dissimilar"



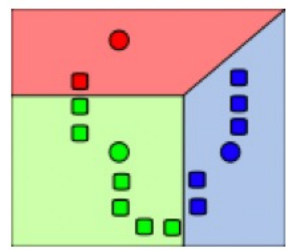
Know NOTHING about labels

Existing methods: K-Means

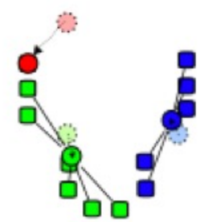
- Standard Algorithm



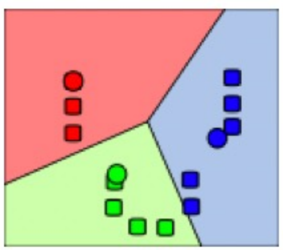
1) k initial "means" (in this case k=3) are randomly selected from the data set.



2) k clusters are created by associating every observation with the nearest mean.



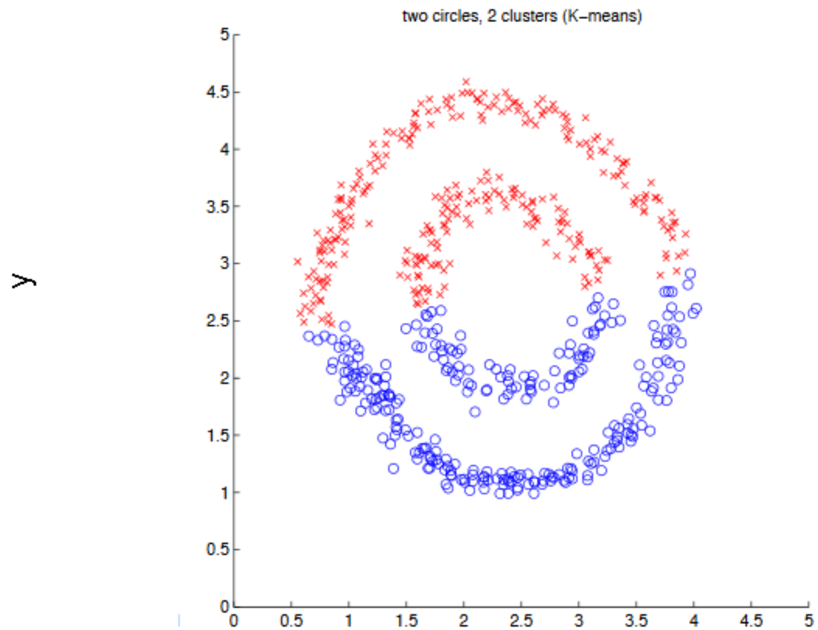
3) The centroid of each of the k clusters becomes the new means.



4) Steps 2 and 3 are repeated until convergence has been reached.

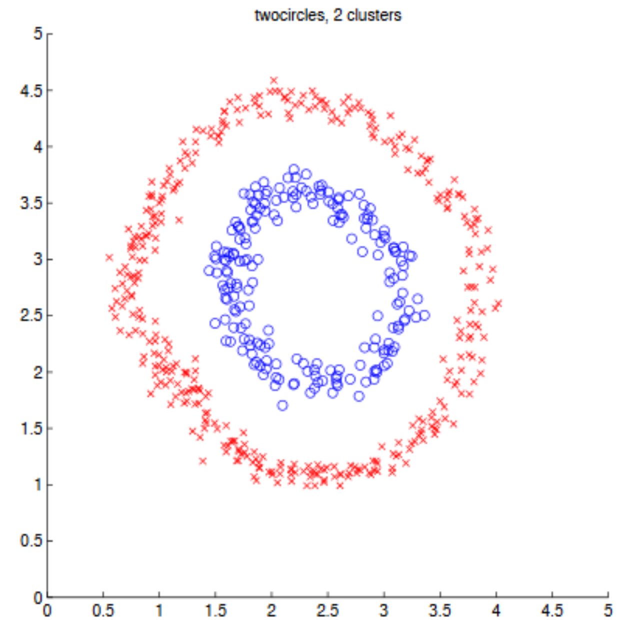
Existing problems

K-means



(i)

Spectral clustering



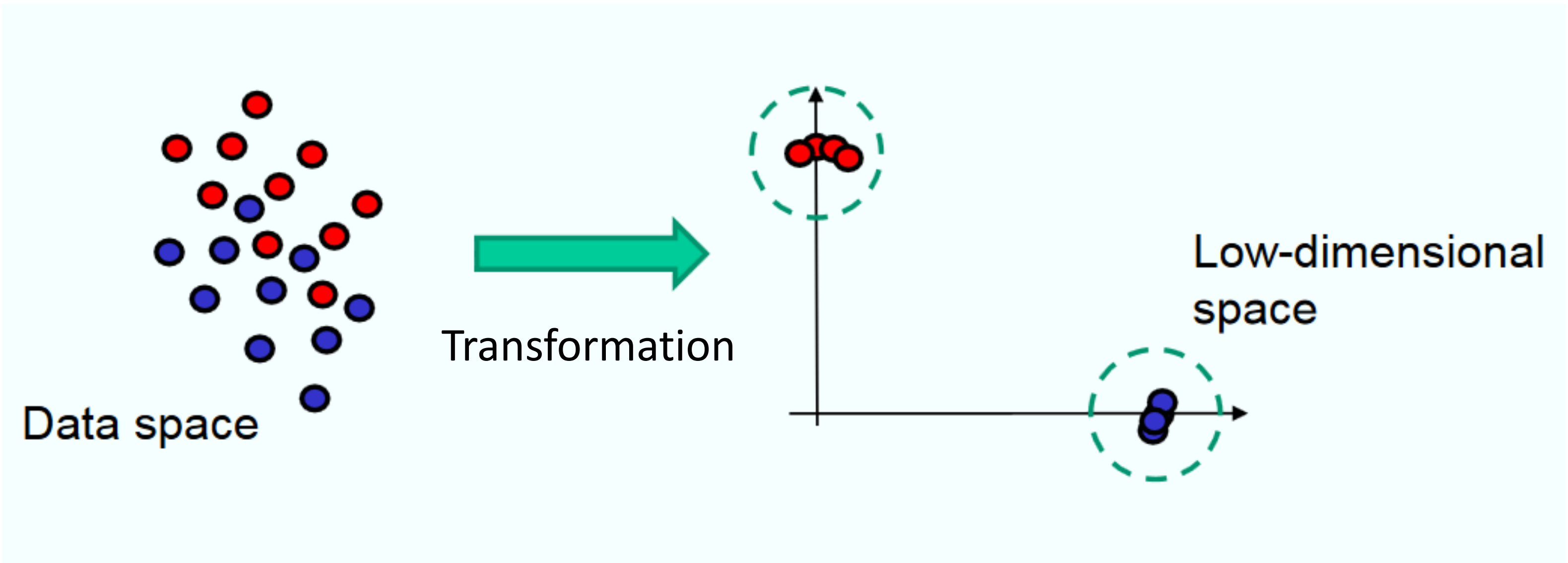
(e)

[Shi & Malik ‘00; Ng, Jordan, Weiss NIPS ‘01]

- Distance based method (like k-means) fails to capture the complicated geometric structure

Spectral Clustering

- Basic Idea: Low-dimensional embedding point of view
  - Obtain data representation in the low-dimensional space that can be easily clustered



How to learn this transformation?

http://www.cvl.isy.liu.se:82/education/graduate/spectral-clustering/SC_course_part1.pdf

Similarity Graph

Similarity Graphs: Model local neighborhood relations between data points

- ε-neighborhood graph

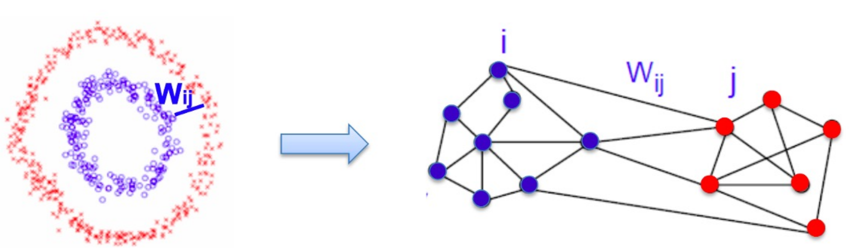
  Connect all vertices whose pairwise distances are smaller than ε

- k-nearest neighbor graph

  Connect vertex v_i with vertex v_j if v_j is among the k-nearest neighbors of v_i.

- fully connected graph

  Connect all points with positive similarity with each other



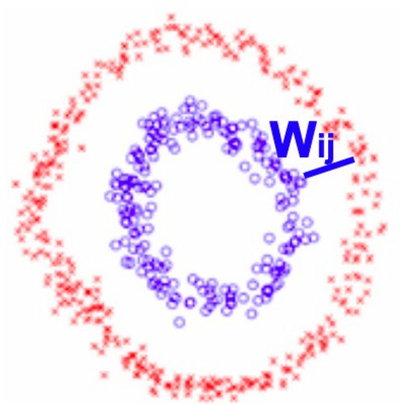
Similarity Matrix

E.g. epsilon-NN

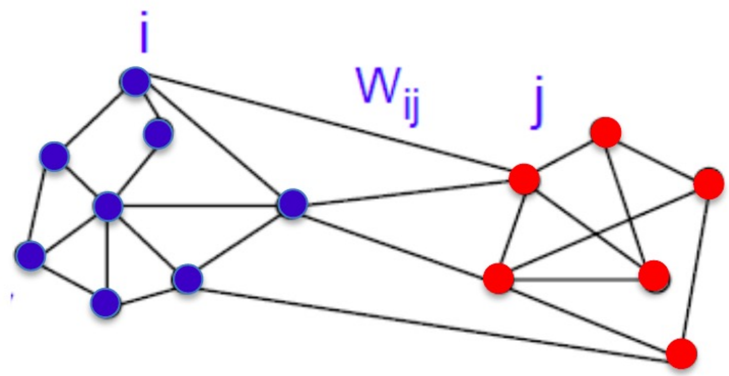
$$W_{ij} = \begin{cases} 1 & \|x_i - x_j\| \le \epsilon \\ 0 & \text{otherwise} \end{cases}$$

Controls size of neighborhood

or mutual k-NN graph (W_{ij} = 1 if x_i or x_j is k nearest neighbor of the other)



Data clustering



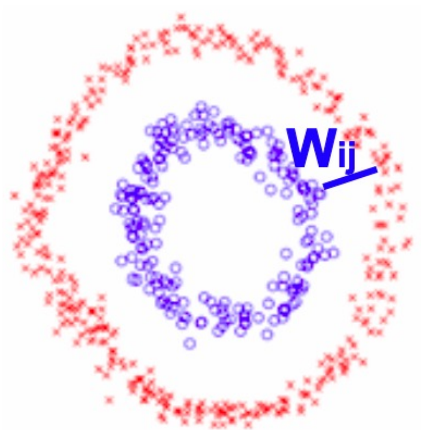
$G = \{V,E\}$

https://www.cs.cmu.edu/~epxing/Class/10701-12f/Lecture/lecture21-SpecClus.pdf

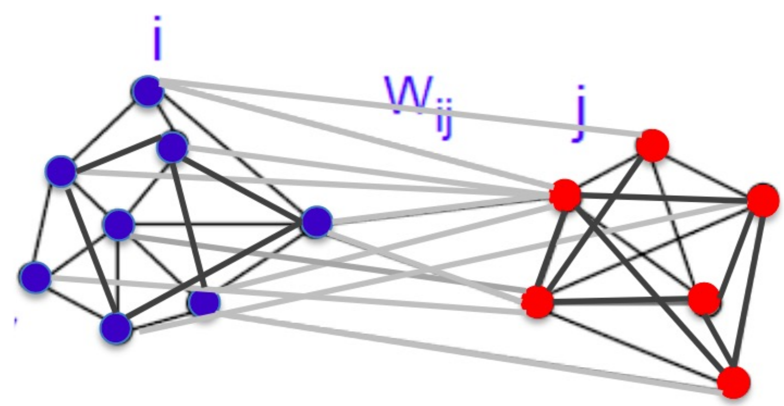
Similarity Matrix

E.g. Gaussian kernel similarity function

$$W_{ij} = e^{\frac{\|x_i - x_j\|^2}{2\sigma^2}}$$ → Controls size of neighborhood



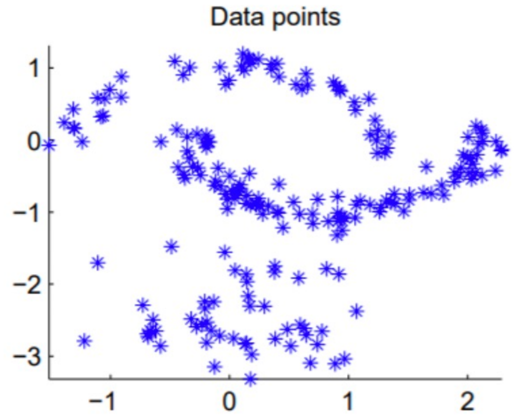
Data clustering

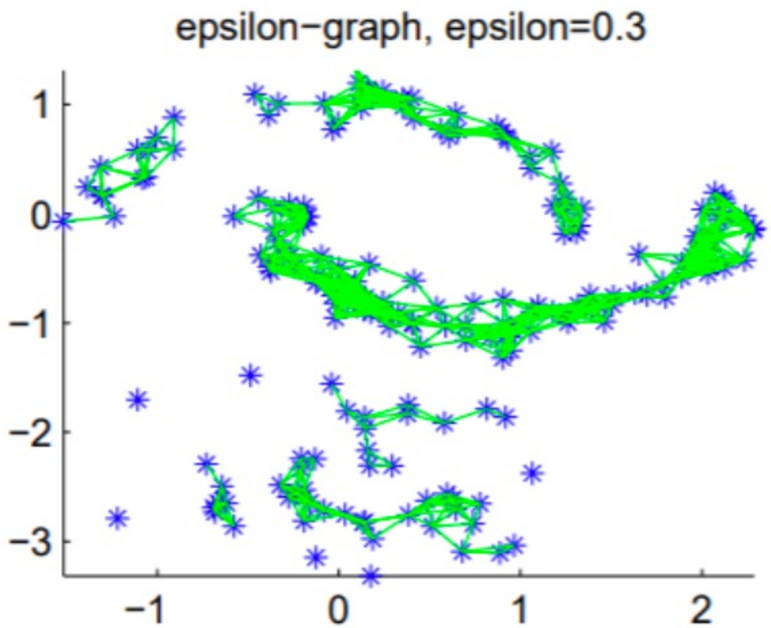


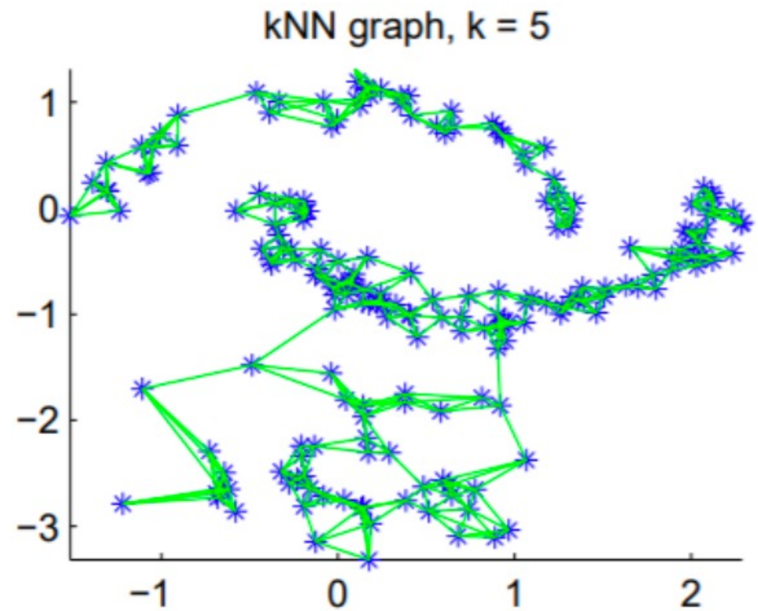
$G = \{V,E\}$

Similarity Graph

- Example

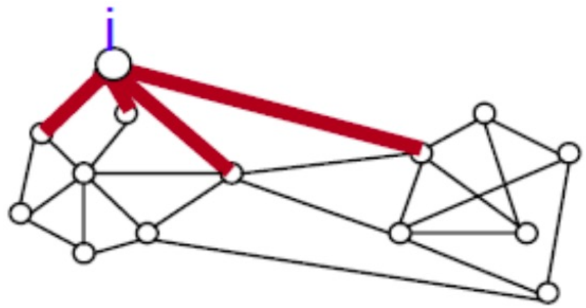






Degree

- $d_i = \sum_j w_{ij}$ degree of a vertex
- $D = diag(d_1, \ldots, d_n)$ degree matrix



	x_1	x_2	x_3	x_4	x_5	x_6
x_1	0	0.8	0.6	0	0.1	0
x_2	0.8	0	0.8	0	0	0
x_3	0.6	0.8	0	0.2	0	0
x_4	0	0	0.2	0	0.8	0.7
x_5	0.1	0	0	0.8	0	0.8
x_6	0	0	0	0.7	0.8	0

Similarity matrix

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	1.5	0	0	0	0	0
x_2	0	1.6	0	0	0	0
x_3	0	0	1.6	0	0	0
x_4	0	0	0	1.7	0	0
x_5	0	0	0	0	1.7	0
x_6	0	0	0	0	0	1.5

Degree matrix

Graph Laplacian matrix

- Un-normalized Graph Laplacian

$$L = D - W$$

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	1.5	-0.8	-0.6	0	-0.1	0
x_2	-0.8	1.6	-0.8	0	0	0
x_3	-0.6	-0.8	1.6	-0.2	0	0
x_4	0	0	-0.2	1.7	-0.8	-0.7
x_5	-0.1	0	0	-0.8	1.7	-0.8
x_6	0	0	0	-0.7	-0.8	1.5

=

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	1.5	0	0	0	0	0
x_2	0	1.6	0	0	0	0
x_3	0	0	1.6	0	0	0
x_4	0	0	0	1.7	0	0
x_5	0	0	0	0	1.7	0
x_6	0	0	0	0	0	1.5

−

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	0	0.8	0.6	0	0.1	0
x_2	0.8	0	0.8	0	0	0
x_3	0.6	0.8	0	0.2	0	0
x_4	0	0	0.2	0	0.8	0.7
x_5	0.1	0	0	0.8	0	0.8
x_6	0	0	0	0.7	0.8	0

- Normalized Graph Laplacian

$$L_{\mathrm{rw}} := D^{-1}L = I - D^{-1}W.$$

Graph Laplacian matrix

- Important properties:
  - Symmetric, positive semi-definite

    A symmetric matrix M is positive semidefinite (PSD) *if* $\forall x \in \mathbb{R}^n$,

    $$x^T M x \geq 0.$$

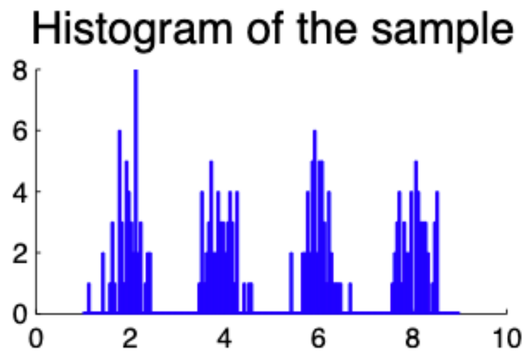
    - Eigenvalues are non-negative real numbers $0 = \lambda_i \leq \lambda_2 \leq \cdots \leq \lambda_n$
    - Eigenvectors are real and orthogonal
    - Smallest eigenvalue is 0, corresponding eigenvector is 1.

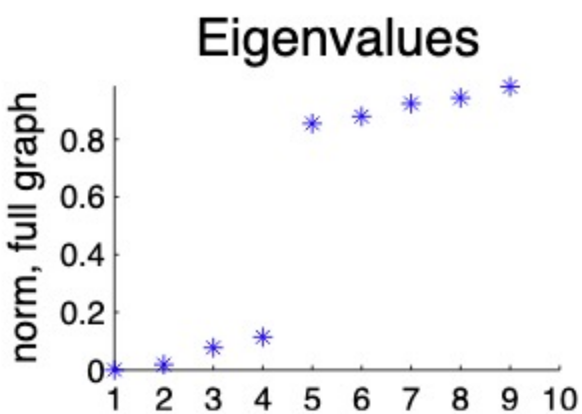
$$L\mathbf{1} = D\mathbf{1} - W\mathbf{1} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix} - \begin{bmatrix} \sum_j w_{1j} \\ \sum_j w_{2j} \\ \vdots \\ \sum_j w_{nj} \end{bmatrix} = 0$$

Graph Laplacian matrix

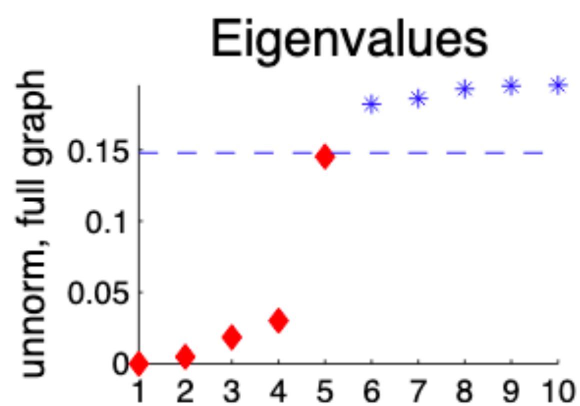
- Important properties:
  - Symmetric, positive semi-definite
  - Eigenvalues and eigenvectors provide an insight into the connectivity of the graph
    - The number of eigenvalues equal to zero for the Laplacian is equal to the number of connected components in the graph.
    - The rank of Laplacian of a graph is number of vertices minus number of connected components.

Eigenvalue





$$L = D - W.$$



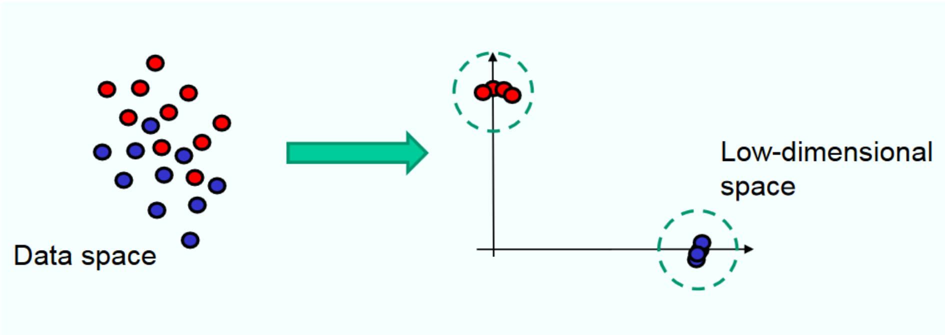
$$L_{\mathrm{rw}} := D^{-1}L = I - D^{-1}W.$$

Spectral Clustering

- Idea (1D): Given the similarity matrix *W*,

$$\min \sum_{i,j}^{n} w_{ij}^2 (f_i - f_j)^2$$

  - f_i *is the low-dimensional representation*
  - *Larger* w_{ij} *enforces* f_i *and* f_j *more similar*
  - *Two similar points in the original space will be similar in the low-dimensional space*



Spectral Clustering

- Reformulation:

$$f'Lf = f'Df - f'Wf = \sum_{i=1}^{n} d_i f_i^2 - \sum_{i,j=1}^{n} f_i f_j w_{ij}$$

$$= \frac{1}{2}\left(\sum_{i=1}^{n} d_i f_i^2 - 2\sum_{i,j=1}^{n} f_i f_j w_{ij} + \sum_{j=1}^{n} d_j f_j^2\right) = \frac{1}{2}\sum_{i,j=1}^{n} w_{ij}\left(f_i - f_j\right)^2$$

$$\min \sum_{i,j}^{n} w_{ij}^2\left(f_i - f_j\right)^2 \quad \Rightarrow \quad \min f'Lf$$

- Where un-normalized graph Laplacian $L = D - W$

Spectral Clustering

- How to obtain low-dimensional embedding?

$$L = D - W$$

Multiplicity of eigenvalue 0 = number k of connected components $A_1, \ldots, A_k$ of the graph.

Ky Fan's Theorem

$$\sum_{i=1}^{c} \sigma_i(L\ \ ) = \min_{F \in \mathbb{R}^{n \times c}, F^T F = I} Tr(F^T L\ \ F)$$

The optimal solution F to the problem is formed by the c eigenvectors of L corresponding to the c smallest eigenvalues.

Spectral Clustering

- 1. Construct the similarity matrix
- 2. Compute the Laplacian matrix L_{nxn}
- 3. Conduct the eigen-decomposition to get the first C eigenvectors, U_{nxc}
- 4. Apply k-means to U_{nxc}

Spectral Clustering

K-means

Spectral clustering