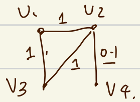
Review of Previous Lecture

① Graph Laplacian example



$$\underbrace{L = \overbrace{D}^{\text{degree}} - \overbrace{W}^{\text{adjacency}}}_{} = \begin{bmatrix} 2 & & & \\ & 2.1 & & \\ & & 2 & \\ & & & 0.1 \end{bmatrix} - \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0.1 \\ 1 & 1 & 0 & 0 \\ 0 & 0.1 & 0 & 0 \end{bmatrix}$$

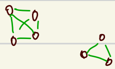
$$= \begin{bmatrix} 2 & -1 & -1 & 0 \\ -1 & 2.1 & -1 & -0.1 \\ -1 & -1 & 2 & 0 \\ 0 & -0.1 & 0 & 0.1 \end{bmatrix}.$$

$-w_{ij}$

d_i

② Spectral Clustering:

- Model data using a similarity graph



  - ε-neighborhood
  - k-nearest neighbor graph
    - symmetric nearest neighbor, $\{(u,v) \mid u \in N_k(v) \text{ or } v \in N_k(u)\}$
    - mutual nearest neighbor $\{(u,v) \mid u \in N_k(v) \text{ and } v \in N_k(u)\}$

- Find partitions $A_1, \dots, A_k$ that minimizes

$$\min_{A_1,\dots A_k \subseteq V} \sum_{i=1}^{k} \frac{cut(A_i, \bar{A}_i)}{|A_i|} \quad (\text{Normalized Cut}) \qquad (1)$$

$\forall i \neq j$, $A_i \cap A_j = \phi$,
$\bigcup A_i = V$

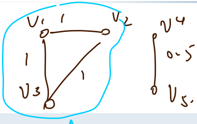
- Transform (1) into an optimization problem of Laplacian $L = D - W$.

$$\min_{f_1, \dots f_k \in \mathbb{R}^n} \sum_{i=1}^{k} \frac{f_i^T L f_i}{f_i^T f_i} \qquad (2).$$

$$\text{s.t. } f_i^T f_j = 0 \ \forall\, i \neq j$$

Properties of $\underset{(n\times n)}{L}$:
- smallest eigenvalue $\lambda_1 = 0$
- $\underset{\underset{\mathbb{1}.}{\uparrow}}{0 =} \lambda_1 \leq \lambda_2 \cdots \leq \lambda_n.$ ←
-

Graph Laplacian



$$W = \begin{bmatrix} \boxed{\begin{matrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{matrix}}_{W_1} & 0 \\ 0 & \boxed{\begin{matrix} 0 & 0.5 \\ 0.5 & 0 \end{matrix}}_{W_2} \end{bmatrix} \quad D = \begin{bmatrix} D_1 & \\ & D_2 \end{bmatrix} \quad \mathbb{1}_{A_1} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \Big\} \mathbb{1}$$

Proposition 1

$L = D - W \quad \begin{bmatrix} L_1 & \\ & L_2 \end{bmatrix}$ $\qquad \mathbb{1}_{A_2} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$

Let G be an undirected graph with non-negative weights W, the multiplicity k of eigenvalue 0 of L is the number of connected components $A_1, \ldots, A_k$ in G.
The eigenspace of eigenvalue 0 is spanned by vectors $\mathbf{1}_{A_1}, \ldots, \mathbf{1}_{A_k}$

1). When $k=1$, G is connected.

Suppose f is an eigenvector of L, with the eigenvalue 0

then $f^T L f = f^T (0 \cdot f) = 0.$

$$\sum_{i,j=1}^{n} w_{ij}(f_i - f_j)^2 = 0.$$

$$\sum_{(i,j)\in E}^{n} w_{ij}(f_i - f_j)^2 = 0,$$ For all $(i,j) \in E$, $f_i = f_j$. since G is connected.

$w_{ij} > 0.$

Therefore f is a constant vector: $f = c \cdot \mathbb{1} = c\mathbb{1}_A$ $\quad f = c \cdot \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \Big\} |V|$

2) When $k>1$, G has k connected components $A_1, \dots, A_k$, w.l.o.g. reorder vertices

$$L = \begin{bmatrix} \boxed{L_1}^{\nearrow A_1} & & & 0 \\ & \boxed{L_2}^{\nearrow A_2} & & \\ & & \ddots & \\ 0 & & & \boxed{L_k} \end{bmatrix}$$ where L_i is the Laplacian of A_i.

→ For a block diagonal matrix L, eigenvalues of L are unions of $L_1, \dots, L_k$, whose eigenvectors are eigenvectors of L_i with 0 filled in other blocks.

i.e. Let λ_{ij}, f_{ij} be the jth eigenvalue, eigenvector of L_i,

$$\underline{LV.} = \underbrace{\begin{bmatrix} \ddots & & \\ & \boxed{L_i} & \\ & & \ddots \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} 0 \\ f_{ij} \\ 0 \end{bmatrix}}_{v} = \begin{bmatrix} 0 \\ L_i f_{ij} \\ 0 \end{bmatrix} = \lambda_{ij} \underbrace{\begin{bmatrix} 0 \\ f_{ij} \\ 0 \end{bmatrix}}_{V_i}.$$ therefore $\begin{bmatrix} 0 \\ f_{ij} \\ 0 \end{bmatrix}$ is an eigenvector of L

Each $\underline{L_i}$ has eigenvalue 0 with multiplicity 1. L has total multiplicity k, its eigenspace is spanned by $\underline{\mathbb{1}_{A_i}}$

(Normalized) Graph Laplacian

$$\begin{bmatrix}\frac{1}{d_1} & & & \\ & \frac{1}{d_2} & & \\ & & \ddots & \\ & & & \frac{1}{d_n}\end{bmatrix}$$

Normalized graph laplacian (Chung 1997) [1]:

$$L_{rw} = D^{-1}L = I - D^{-1}W$$

Properties of L_{rw}

- λ is an eigenvalue of L_{rw} with eigenvector v if and only if λ, v solve the generalized eigenproblem $Lv = \lambda D v$
- 0 is an eigenvalue of L with eigenvector $\mathbf{1}$
- L_{rw} is positive semi-definite and has n non-negative eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$

[1] "rw" comes from its interpertation as "random walk". Another definition of normalized graph Laplacian is $D^{-\frac{1}{2}}LD^{-\frac{1}{2}}$

(Normalized) Graph Laplacian

Normalized graph laplacian (Chung 1997) [1]:

$$L_{rw} = D^{-1}L = I - D^{-1}W$$

Properties of L_{rw}

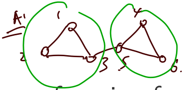
- λ is an eigenvalue of L_{rw} with eigenvector v if and only if λ, v solve the generalized eigenproblem $Lv = \lambda Dv$
- 0 is an eigenvalue of L with eigenvector $\mathbf{1}$
- L_{rw} is positive semi-definite and has n non-negative eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$

Proposition 2

Let G be an undirected graph with non-negative weights W, the multiplicity k of eigenvalue 0 of L_{rw} is the number of connected components $A_1, \ldots, A_k$ in G.
The eigenspace of eigenvalue 0 is spanned by vectors $\mathbf{1}_{A_1}, \ldots, \mathbf{1}_{A_k}$

[1] "rw" comes from its interpertation as "random walk". Another definition of normalized graph Laplacian is $D^{-\frac{1}{2}}LD^{-\frac{1}{2}}$

Solving graph cut



$|A| = 3.$ $\boxed{\rightarrow \frac{Cut(A,\bar{A})}{|A|}}$

$\frac{Cut(A,\bar{A})}{vol(A)}$

Define $f \in \{0,1\}^n$ to be the indicator function for partition $A \subset V$:

$n = |V|$ $\left\{ \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix} \right.$

$$f_i := \{\mathbf{1}_A\}_i = \begin{cases} 1 & v_i \in A \\ 0 & v_i \in \bar{A} \end{cases}$$

$$\begin{array}{ccc} & f_1 & f_2 \\ F = & \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \end{array}$$

We have that $||\vec{f}||^2 = |A|$. $\|f\|^2 = \sum_{i=1}^{n} f_i^2 = \sum_{v_i \in A} 1^2 = |A|.$

$Cut(A, \bar{A})$ can be written as a function of f and graph Laplacian L:

$$f^T L f = \frac{1}{2} \sum_{i,j=1}^{n} w_{ij}(f_i - f_j)^2 = \frac{1}{2}.$$

$v_i \in A, v_j \in \bar{A} \Rightarrow (f_i - f_j)^2 = (1-0)^2 = 1$

$v_i \in \bar{A}, v_i \in A. \Rightarrow (f_i - f_j)^2 = (0-1)^2 = 1.$

$$= \frac{1}{2}\left(\sum_{v_i \in A, v_j \in \bar{A}} w_{ij} + \sum_{v_i \in \bar{A}, v_j \in A} w_{ij} \right) = \sum_{v_i \in A, v_j \in \bar{A}} w_{ij} = cut(A, \bar{A})$$

Let $f_{(1)}, \ldots, f_{(k)}$ be k indicator functions $\mathbf{1}_{A_i}, \ldots, \mathbf{1}_{A_k}$. They are mutually orthogonal (i.e. $f_{(i)}^T f_{(j)} = 0$ for all $i \neq j$).

Solving graph cut

Recall the definition of RatioCut:

$$\min_{A_1,\ldots,A_k} \sum_{i}^{k} \frac{cut(A_i, \bar{A}_i)}{|A_i|} \tag{3}$$

$$\Longrightarrow \min_{A_1,\ldots,A_k} \sum_{i}^{k} \frac{f_{(i)}^T L f_{(i)}}{f_{(i)}^T f_{(i)}} \leftarrow \|f_i\|^2 \tag{4}$$

Relax the $f_{(i)}$'s to be real vectors:

$$\min_{f_{(1)},\ldots,f_{(k)}\in\mathbb{R}^n} \sum_{i}^{k} \frac{f_{(i)}^T L f_{(i)}}{f_{(i)}^T f_{(i)}} \tag{5}$$

$$s.t.\ f_{(i)}^T f_{(j)} = 0, \text{ for all } i \neq j$$

Solving graph cut

Since rescaling $f_{(i)}$ by constants does not change the objective, (3) is equivalent to

$$\min_{f_{(1)},\ldots,f_{(k)}\in\mathbb{R}^n} \sum_{i}^{k} f_{(i)}^T L f_{(i)} \tag{6}$$

$$\text{s.t. } f_{(i)}^T f_{(j)} = 0, \text{ for all } i \neq j$$

$$f_{(i)}^T f_{(i)} = 1, \text{ for all } i = 1, \ldots, k$$

$$\begin{bmatrix} f_1 & \cdots & f_k \end{bmatrix}$$

Let $F = \left[f_{(1)} \ldots f_{(k)}\right]$, (5) can be written in matrix notation:

$$F = \begin{array}{c} f_1 \cdots f_k \\ \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & & \\ & & & & \end{bmatrix} \end{array} \begin{array}{c} v_1 \\ \vdots \\ v_n \end{array}$$

$$\min_{F\in\mathbb{R}^n} \operatorname{tr}(F^T L F)$$
$$\text{s.t. } F^T F = I$$

smallest.

- By Theorem 2 , optimal solution F^* is the first k eigenvectors of L.
- To get discrete cluster labels, we can apply k-means clustering on the rows of F^*.

Spectral Clustering Algorithm

Unormalized spectral clustering

Input: data points $x^{(1)}, \dots, x^{(n)}$ and cluster size k

- Build a graph connecting $x^{(1)}, \dots, x^{(n)}$ with weight W
- Compute first k smallest eigenvectors $V = [v_1, \dots, v_k]$ of L
- Define $y_i \in \mathbb{R}^k$ as the ith row of V, cluser $y_1, \dots, y_n$ into k clusters $C_1, \dots, C_k$ using k-means

$$V = \begin{bmatrix} -y_1- \\ -y_2- \\ \vdots \\ -y_n- \end{bmatrix}$$

Output: $A_1, \dots, A_k$ where $A_i = \{j | y_j = C_i\}$ cluster label

- Unormalized spectral clustering is relaxed solution to the RatioCut problem.

Spectral Clustering Algorithm

Normalized spectral clustering (Ng, Shi and Malik 2000)

Input: data points $x^{(1)}, \ldots, x^{(n)}$ and cluster size k

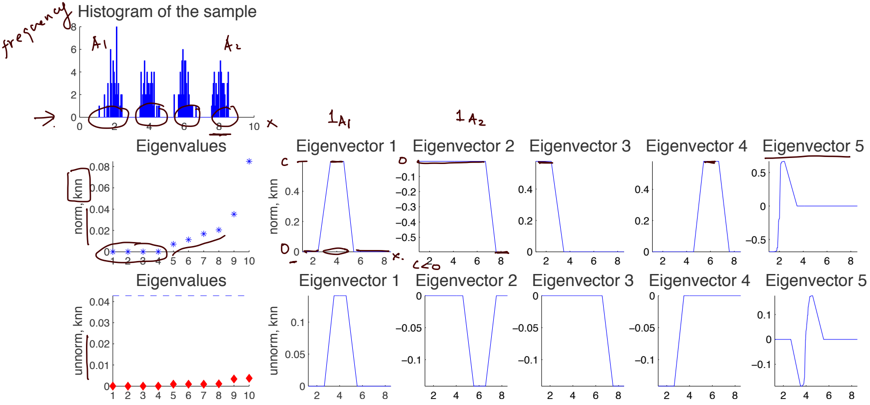
- Build a graph connecting $x^{(1)}, \ldots, x^{(n)}$ with weight W
- Compute first k eigenvectors $V = [v_1, \ldots, v_k]$ of generalized eigen problem $Lv = \lambda Dv$
- Define $y_i \in \mathbb{R}^k$ as the ith row of V, cluser $y_1, \ldots, y_n$ into k clusters $C_1, \ldots, C_k$ using k-means

Output: $A_1, \ldots, A_k$ where $A_i = \{j | y_j = C_i\}$

- Normalized spectral clustering (L_{rw}) is a relaxed solution to the NCut problem.

Toy Example

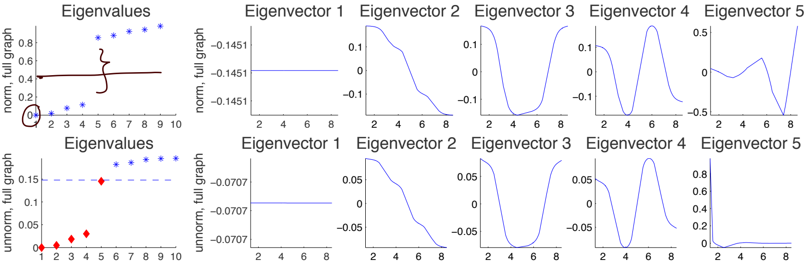
- 200 data points sampled from 4 Gaussian distributions
- KNN similarity graph ($k = 10$)



First 4 eigenvalues are 0 with eigenvectors 1_{A_i} , $i = 1, \ldots, 4$

Toy Example

- Fully connected graph with Gaussian similarity graph ($\sigma = 1$)



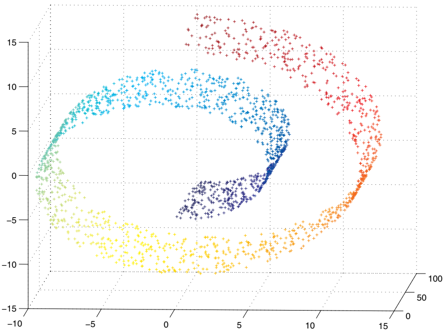
First eigenvector is **1** since the graph has only 1 connected component

Spectral Embedding

Also known as Laplacian Eigenmaps [Belkin et. al., 2003]: $X = \mathbb{R}^{n \times d}$. $d >> k$

- Learn a k-dimensional embedding $Y = \begin{bmatrix} -y_1- \\ \vdots \\ -y_m- \end{bmatrix} \in \mathbb{R}^{n \times k}$

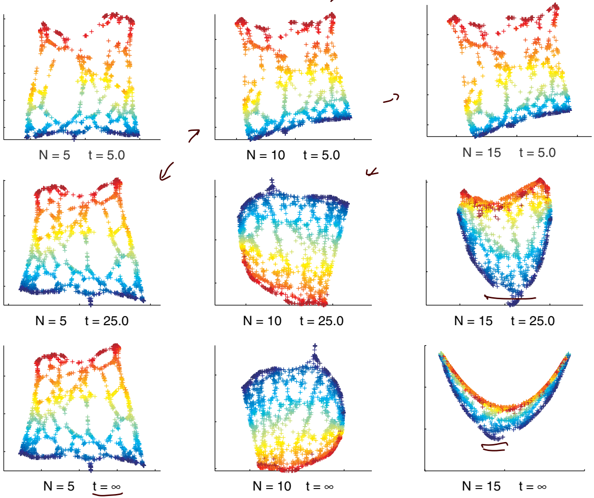
$$\min_{\substack{Y^T D Y = I \\ Y^T D \mathbf{1} = 0}} \frac{1}{2} \sum_{ij} w_{ij} ||y_i - y_j||^2$$



Spectral Embedding

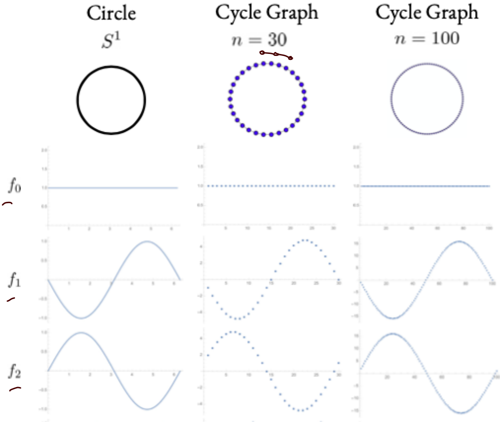
Example: 2D embedding results:

- N: number of neighbors in kNN graph
- t: hyperparameter in the similarity function $W_{i,j} = \exp(\frac{||x_i - x_j||^2}{t})$



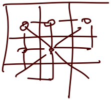
Spectral Embedding

Example: Embedding results on a 2D cycle graph



Also studied in graph signal processing and differential geometry

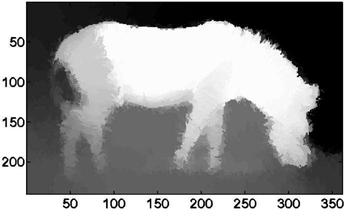
Additional topics of graph Laplacian methods



Graph spectra can be used as topological features for supervised and unsupervised learning

- Laplacian eigenmaps for dimension reduction and visualization
- Unsupervised segmentation
- Graph-based semi-supervised learning (manifold regularization)

f^TLf



Unsupervised segmentation using NCut [Shi & Malik, 2000]

Lazy Snapping (semi-supervised graph cut) [Li et. al. 2004]

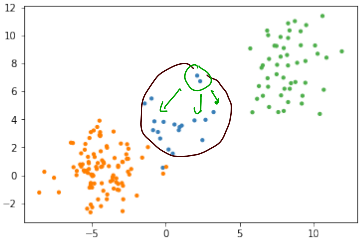
Summary

Representation learning

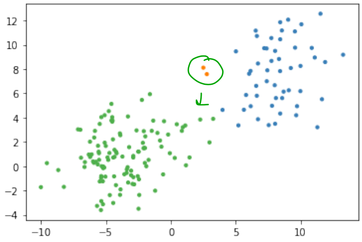
- Transform input features into "simpler" or "interpretable" representations.
- Used in feature extraction, dimension reduction, clustering etc

Unsupervised learning algorithms and their assumptions

- **K-Means**: assumes data are isotropic Gaussian, different clusters have the same prior probability
- **Spectral Methods**: manifold assumption, cluster labels of a node depends on its neighbors



ground truth cluster



spectral clustering (nCut)