Learning from Data
Lecture 8: Unsupervised Learning I

Yang Li **yangli@sz.tsinghua.edu.cn**

TBSI

November 25, 2021

Today's Lecture

Unsupervised Learning (Part I)

- Overview: the representation learning problem
- K-means clustering
- Spectral clustering

Project Introduction

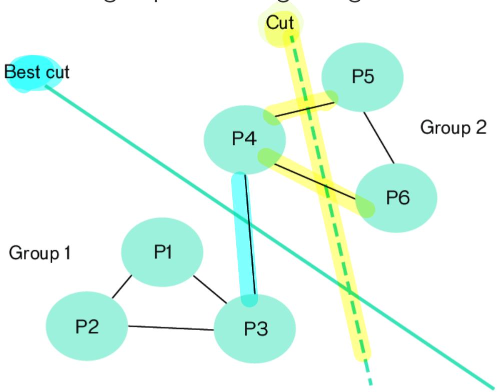
Clustering from a graph point of view

- Given data points $x^{(1)}, \ldots, x^{(n)}$ and **similarity measure** $s_{ij} \geq 0$ for all $x^{(i)}, x^{(j)}$
- A typical **similarity graph** $G = (V, E)$ is
  - $v_i \leftrightarrow x^{(i)}$
  - v_i and v_j are connected if $s_{ij} \geq \delta$ for some threshold δ
- **Clustering**: Divide data into groups such that points in the same group are similar and points in different groups are dissimilar
- **Spectral Clustering (informal)**: *Find a partition of G such that edges between the same group have high weight and edges between different groups have very low weight.*

Spectral Clustering as Graph Partitioning

Find a partition of the graph such that

- Edges between groups have a low weight
- Edges within each group have a high weight



Graph Cut Formulation

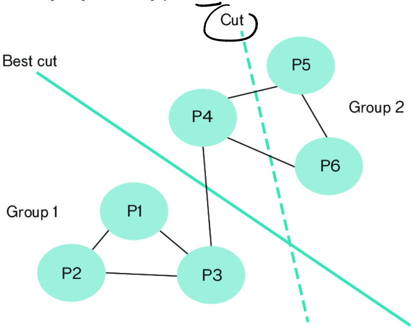
Case $k = 2$:

- Given partition $A, \bar{A}$, define a cut as the total weight of edges weights between groups:

$$cut(A, \bar{A}) := \sum_{i \in A, j \in \bar{A}} w_{ij}$$

$w_{ij} = 1$ if i, j connected by edge.

- Example: $cut(\{p_1, p_2, p_3\}, \{p_4, p_5, p_6\}) = 1$, $cut(\{p_1, p_2, p_3, p_4\}, \{p_5, p_6\}) = 2$



Graph Cut Formulations

Case $k > 2$:

- Given partition $A_1, \ldots, A_k$, define a cut as the total edges weights between groups:

$$cut(A_1, \ldots, A_k) := \frac{1}{2} \sum_{i=1}^{k} cut(A_i, \bar{A}_i)$$

Graph Cut Formulations

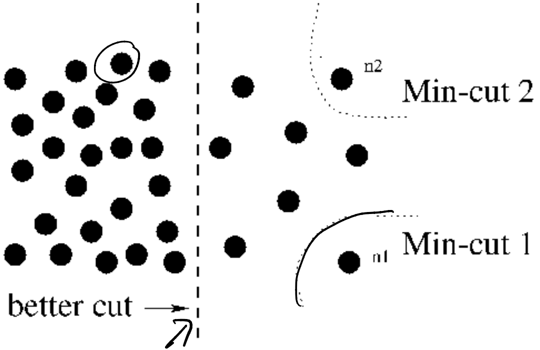
Case $k > 2$:

- Given partition $A_1, \dots, A_k$, define a cut as the total edges weights between groups:

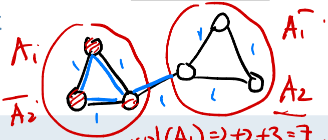
$$\min_{A_1,\cdots A_k} \underline{cut(A_1, \dots, A_k)} := \frac{1}{2}\sum_{i=1}^{k} cut(A_i, \bar{A}_i)$$

min-cut algorithm

Minimizing cut directly tends to favor small isolated clusters.



Balanced Graph Cut



$Vol(A_i) = \sum_{i \in A_i} \sum_{j \in V} w_{ij} = \sum_{i \in A} d_i$

$vol(A_1) = 2+2+3 = 7$, $vol(A_2) = 7$.

RatioCut and NCut

Find a k-way partition of graph G ($A_i \cup \ldots \cup A_k = V, A_i \cap A_j = \varnothing$) that minimizes:

$$RatioCut(A_1, \ldots, A_k) = \frac{1}{2} \sum_{i=1}^{k} \frac{cut(A_i, \bar{A}_i)}{|A_i|}$$ [Hagen & Kahng,1992]

$|A_1| = 3$. $|A_2| = 3$.

$$NCut(A_1, \ldots, A_k) = \frac{1}{2} \sum_{i=1}^{k} \frac{cut(A_i, \bar{A}_i)}{vol(A_i)},$$

$$vol(A_i) = \sum_{i \in A, j \in V} w_{ij}$$ [Shi & Malik ,2000]

$= \sum_{i \in A} d_i$

Balanced Graph Cut

RatioCut and NCut

Find a k-way partition of graph G ($A_i \cup \ldots \cup A_k = V, A_i \cap A_j = \varnothing$) that minimizes:

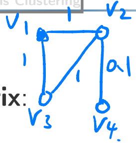
$$RatioCut(A_1, \ldots, A_k) = \frac{1}{2}\sum_{i=1}^{k} \frac{cut(A_i, \bar{A}_i)}{|A_i|}$$ [Hagen & Kahng,1992]

$$NCut(A_1, \ldots, A_k) = \frac{1}{2}\sum_{i=1}^{k} \frac{cut(A_i, \bar{A}_i)}{vol(A_i)},$$

$$vol(A_i) = \sum_{i \in A, j \in V} w_{ij}$$ [Shi & Malik ,2000]

Both RatioCut and NormalizeCut can be **approximated** *by spectral method.*

Graph Laplacian



$$L = D - W \qquad w_{ij} \geq 0.$$

Unnormalized graph laplacian matrix:

$$L = D - W$$

$$= \begin{bmatrix} 2 & & & \\ & 2.1 & & \\ & & 2 & \\ & & & 0.1 \end{bmatrix} - \underbrace{\begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0.1 \\ 1 & 1 & 0 & 0 \\ 0 & 0.1 & 0 & 0 \end{bmatrix}}_{W}$$

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

Properties of L

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

$$\frac{1}{2} \sum_i^n \sum_j^n w_{ij}(f_i^2 + f_j^2 - 2f_i f_j)$$

$$= \frac{1}{2}\Big(\underbrace{\sum_i^n \sum_j^n w_{ij}}_{d_i} f_i^2 + \underbrace{\sum_i^n \sum_j^n w_{ij}}_{d_j} f_j^2 - 2 \sum_i^n \sum_j^n w_{ij} f_i f_j \Big) \qquad w_{ij} = w_{ji}$$

$$= \frac{1}{2}\Big(\sum_i^n d_i f_i^2 + \sum_j^n d_j f_j^2 \Big) - \sum_i^n \sum_j^n w_{ij} f_i f_j$$

$$= \sum_{i=1}^n d_i f_i^2 - \sum_i^n \sum_j^n w_{ij} f_i f_j = f^T D f - f^T W f = f^T (D - W) f = f^T L f.$$

Graph Laplacian

Unnormalized graph laplacian matrix:

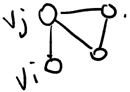
$$L = D - W$$

Properties of L

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

Graph Laplacian

G is undirected:



v1
v2
v3
v4

(vj)

Unnormalized graph laplacian matrix:

$$L = \underline{D} - \underline{W}$$

$w_{ij} = w_{ji}$

Properties of L

- For every $f \in \mathbb{R}^n$, $f^T L f = \frac{1}{2}\sum_{i,j=1}^{n} w_{ij}(f_i - f_j)^2$
- L is symmetric and positive semi-definite

$\forall f \in \mathbb{R}^n$, $\underline{f L f} = \frac{1}{2}\sum_{i,j}^{n} \underbrace{w_{ij}}_{\geq 0} \underbrace{(f_i - f_j)^2}_{\geq 0}$

Graph Laplacian

Unnormalized graph laplacian matrix:

$$L = D - W$$

Properties of L

- For every $f \in \mathbb{R}^n$, $f^T L f = \frac{1}{2} \sum_{i,j=1}^{n} w_{ij}(f_i - f_j)^2$
- L is symmetric and positive semi-definite
- The smallest eigenvalue of L is 0 with eigenvector $\mathbf{1}$

Graph Laplacian

Unnormalized graph laplacian matrix:

$$L = D - W$$

Properties of L

- For every $f \in \mathbb{R}^n$, $f^T L f = \frac{1}{2} \sum_{i,j=1}^{n} w_{ij} (f_i - f_j)^2$
- L is symmetric and positive semi-definite
- The smallest eigenvalue of L is 0 with eigenvector $\mathbf{1}$

Graph Laplacian

Unnormalized graph laplacian matrix:

$$L = D - W$$

Properties of L

- For every $f \in \mathbb{R}^n$, $f^T L f = \frac{1}{2} \sum_{i,j=1}^{n} w_{ij} (f_i - f_j)^2$
- L is symmetric and positive semi-definite
- The smallest eigenvalue of L is 0 with eigenvector $\mathbf{1}$
- L has n real eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$

A Review on Eigenvalue Problem

The Eigenvalue Problem

Nonzero vector $u \in \mathbb{R}^n$ is an **eigenvector** of matrix $A \in \mathbb{R}^{n \times n}$ if

$$Au = \lambda u$$

for some $\lambda \in \mathbb{R}$. We call λ the **eigenvalue** corresponding to u.

- A has at most n distinct eigenvalues

Eigenvalue Decomposition

Let $U = [u_1, \ldots, u_n]$ be the matrix of n linearly independent eigenvectors of A and $\Lambda = diag([\lambda_1, \ldots, \lambda_n])$, then

$$A = U \Lambda U^{-1}$$

$$\begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \left(\begin{bmatrix} u \end{bmatrix}^{-1} \right)$$

- If A is symmetric, A can be decomposed as $A = U \Lambda U^T$ where U is an orthogonal matrix ($U^T U = I$).

Rayleigh-Ritz Theorem

Theorem 1

Given symmetric matrix $A \in \mathbb{R}^{n\times n}$, the solution to the minimization problem is the smallest eigen vector of A

$$\min_{x\in\mathbb{R}^n} \; x^TAx \quad \text{s.t.} \; \|x\|^2 = 1 \tag{1}$$

$\Leftrightarrow$ $\min_{x\in\mathbb{R}^n} x^TAx$ st. $x^Tx - 1 = 0$.

- An equivalent form of (1) is minimizing the **Rayleigh quotient** $\frac{x^TAx}{x^Tx}$

scale x by some c such that $\|x\|^2 = 1$.

$$\min_{x\neq 0\in\mathbb{R}^n} \frac{x^TAx}{x^Tx}$$

$L(x) = x^TAx + \beta(x^Tx - 1)$

$\frac{\partial L}{\partial x} = 2Ax + 2\beta x = 0$.

$Ax = -\beta x$.

x is an eigenvector of A.

- Rayleigh quotient $\frac{x^TAx}{x^Tx}$ is scale invariant.

Let $x' = cx$ $(c \in \mathbb{R})$

$\frac{x'^TAx'}{x'^Tx'} = \frac{(cx)^TA(cx)}{(cx)^T(cx)} = \frac{x^TAx}{x^Tx}$

$x^TAx = (A^Tx)^Tx = (-\beta x)^Tx = -\beta = \lambda$

Then $\lambda = \lambda_{smallest}$.

Rayleigh-Ritz Theorem

$x_i \perp x_j$

$$\begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_k \\ | & | & & | \end{bmatrix}$$

Generalization to multiple vectors:

Theorem 2

$(n \times k)$ $\quad \|x_i\|^2 = 1. \quad k \le n$

Given symmetric matrix $A \in \mathbb{R}^{n\times n}$, $\vec{X} = [x_1, \ldots, x_k]$, $x_i \in \mathbb{R}^n$, *the solution to the minimization problem are k smallest eigen vector of A:*

$$\min_{X\in\mathbb{R}^{n\times k}} \quad tr(X^T A X) \tag{2}$$

$$s.t. \quad \underbrace{X^T X}_{(k\times n)(n\times k)} = \underbrace{I}_{(k\times k)}$$

$$\min_{x_1,\ldots,x_k} \sum_{i=1}^{k} x_i^T A x_i$$

$$\text{s.t.} \quad x_i^T x_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \Rightarrow x_i \perp x_j$$

Graph Laplacian

Unnormalized graph laplacian matrix:

$$L = D - W$$

Properties of L

- The smallest eigenvalue of L is 0 with eigenvector $\mathbf{1}$

$$L\mathbb{1} = (D - W)\mathbb{1} = D\mathbb{1} - W\mathbb{1} = D - D = 0 \cdot \mathbb{1}.$$

Since L is positive semi-definite. $\Rightarrow$ $\lambda_i \geq 0$. for all eigenvalues λ_i $i = 1, \dots, n$.

Graph Laplacian

Unnormalized graph laplacian matrix:

$$L = D - W$$

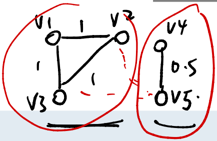
Properties of L

- The smallest eigenvalue of L is 0 with eigenvector $\mathbf{1}$
- L has n real eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$

- symmetric. $\Rightarrow$ n real eigenvalues

- semi positive definite $\Rightarrow$. $\lambda_i \geq 0$.

Graph Laplacian



$$W=\begin{array}{c} \begin{array}{ccccc} v_1 & v_2 & v_3 & v_4 & v_5 \end{array} \\ \left[\begin{array}{ccc|cc} 0 & 1 & 1 & & \\ 1 & 0 & 1 & \multicolumn{2}{c}{O} \\ 1 & 1 & 0 & & \\ \hline & O & & 0 & 0.5 \\ & & & 0.5 & 0 \end{array}\right] \end{array} \quad D=\begin{bmatrix} D_1 & O \\ O & D_2 \end{bmatrix} \quad \mathbf{1}_{A_1}=\begin{bmatrix}1\\1\\1\\0\\0\end{bmatrix}$$

$$L = D - W = \begin{bmatrix} D_1-W_1 & 0 \\ O & D_2-W_2 \end{bmatrix} \quad \mathbf{1}_{A_2}=\begin{bmatrix}0\\0\\0\\1\\1\end{bmatrix}$$

Proposition 1

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

1). When $k=1$. G has one connected component.

$$\mathbf{1}_{A_j}\{i\} = \begin{cases} 1 & v_i \in A_j \\ 0 & v_i \notin A_j \end{cases}$$

Suppose f is the eigenvector of L with eigenvalue 0.

$$f^T L f = 0 \leftarrow \lambda$$

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

for all $(v_i, v_j) \in E$, $w_{ij} > 0$. $(f_i - f_j)^2 = 0$.

$f_i = f_j$ for $i,j \in G$ since G is connected.

f is a constant vector. $f = c \cdot \mathbf{1} \rightarrow c\mathbf{1}_A \leftarrow \underbrace{\begin{bmatrix}1\\1\\1\\1\end{bmatrix}}_{|V|}$

2) When $k > 1$, G has k connected components $A_1, \dots A_k$.

$$\underline{L} = \begin{bmatrix} \boxed{L_1} & & \\ & \boxed{L_2} & \\ & & \ddots \ \boxed{L_k} \end{bmatrix}$$ where each L_i is the laplacian of $\underline{A_i}$

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

let λ_{ij}, f_{ij} be jth eigenvalue, eigenvector. of $\underline{L_i}$

$$L \begin{bmatrix} & & \\ & \boxed{L_i} & \\ & & \end{bmatrix} \begin{bmatrix} 0 \\ f_{ij} \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ L_i v_{ij} \\ 0 \end{bmatrix} = \underline{\lambda_{ij}} \begin{bmatrix} 0 \\ v_{ij} \\ 0 \end{bmatrix},$$

$\begin{bmatrix} 0 \\ v_{ij} \\ 0 \end{bmatrix}$ is an eigenvector of L.

(k=1) Each L_i has eigenvalue 0 with <u>multiplicity</u> 1,

Therefore, L has total multiplicity k, its eigenvectors are spanned by $\mathbb{1}_{A_i}$.

$\square$

---

unnormalized graph laplacian

$$L = D - W.$$

(Normalized) Graph Laplacian

Normalized graph laplacian (Chung 1997) [1]:

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

(D-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_{rw} 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

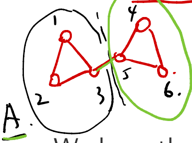
Let G be an undirected graph with non-negative weights W, the multiplicity k of eigenvalue 0 of L_{rw} is the number of connected components $A_1, \ldots, A_k$ in G.
The eigenspace of eigenvalue 0 is spanned by vectors $\mathbf{1}_{A_1}, \ldots, \mathbf{1}_{A_k}$

[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

$A_1, \dots, A_k.$

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



$$\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

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

We have that $\|f\|^2 = |A|$.

$$\|f\|^2 = \sum_{\substack{i=1 \\ i\in A}}^{n} f_i^2 = \sum_{\substack{i=1 \\ i\in A}}^{n} 1 = |A|.$$

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

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

$(v_i \in A, v_j \in \bar{A})$ or $(v_i \in \bar{A}, v_j \in A)$

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

Let $f_{(1)}, \dots, f_{(k)}$ be k indicator functions $\mathbf{1}_{A_i}, \dots, \mathbf{1}_{A_k}$. 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:

RatioCut$(A_1, \dots, A_k)$.

$$\min_{A_1,\dots,A_k} \sum_i^k \frac{cut(A_i, \bar{A}_i)}{|A_i|} = \|f\|^2 \tag{3}$$

$$\Longrightarrow \min_{A_1,\dots,A_k} \sum_i^k \frac{f_{(i)}^T L f_{(i)}}{f_{(i)}^T f_{(i)}} \tag{4}$$

$f_i = 1_{A_i}$

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

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

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

$\min_{f_1 \cdots f_k} \sum_i^k f_i^T L f_i$

s.t. $\|f_i\|^2 = 1$. for all i

Solving graph cut

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

$$\min_{f_{(1)},\dots,f_{(k)}\in\mathbb{R}^n} \sum_{i}^{k} f_{(i)}^T L f_{(i)} \tag{6}$$
$$s.t.\ f_{(i)}^T f_{(j)} = 0,\ \text{for all } i \neq j$$
$$f_{(i)}^T f_{(i)} = 1,\ \text{for all} i = 1,\dots,k$$

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

$k=3$

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

$$\begin{array}{cc} v1 & \\ v2 & \\ v3 & \\ v4 & \\ v5 & \end{array}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

$(v_1, v_3, v_5) \Rightarrow C_1$

$\rightarrow (v_2) \Rightarrow C_2$

$(v_4) \Rightarrow C_3.$

- By Theorem $\boxed{2}$, 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 $x^{(1)}, \ldots, x^{(n)}$ 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 L
- 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 $A_i = \{j | y_j = C_i\}$

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

Spectral Clustering Algorithm

Normalized spectral clustering (Ng, Shi and Malik 2000)

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

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

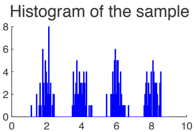
Output: $A_1, \dots, A_k$ where $A_i = \{j | y_j = C_i\}$

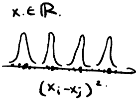
- Normalized spectral clustering (L_{rw}) is a relaxed solution to the NCut problem.

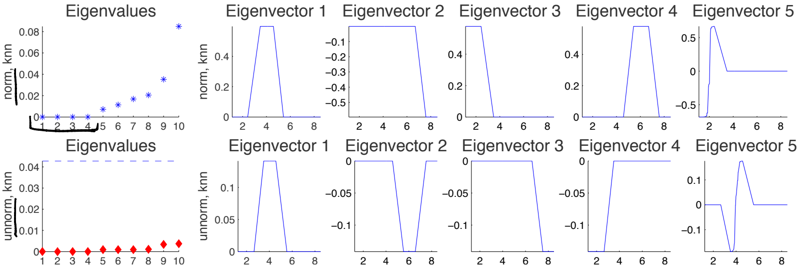
$$\sum_{i=1}^{k} \frac{cut(A_i, \bar{A_i})}{vol(A_i)}.$$

Toy Example

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







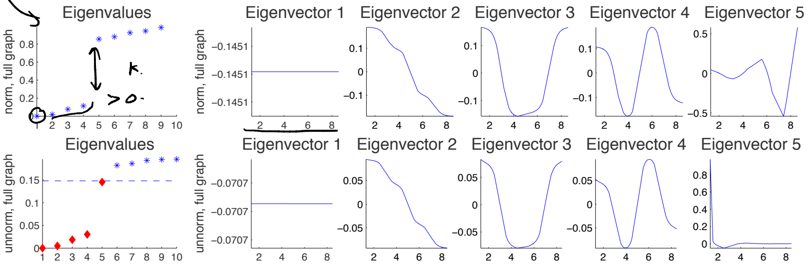
First 4 eigenvalues are 0 with eigenvectors 1_{A_i} , $i = 1, \ldots, 4$

Toy Example

choose k based on eigenvalue gap

$$d(x^{(i)}, x^{(j)}) = e^{-\frac{(x^{(i)} - x^{(j)})^2}{\sigma}}$$

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



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

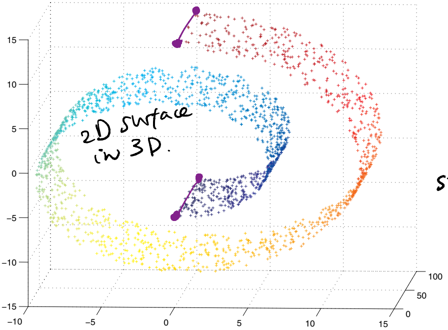
Spectral Embedding

Also known as Laplacian Eigenmaps [Belkin et. al., 2003]:

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

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

normalized graph laplacian.



3D, X

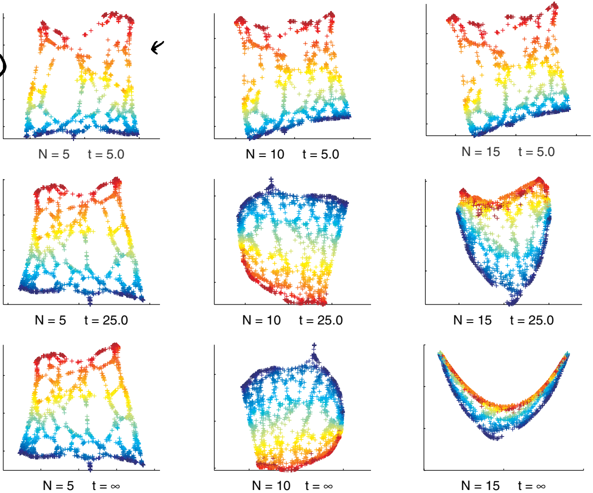
swiss-roll

Spectral Embedding

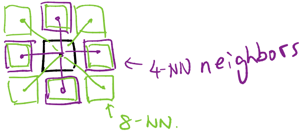
Example: 2D embedding results:

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

unrolling

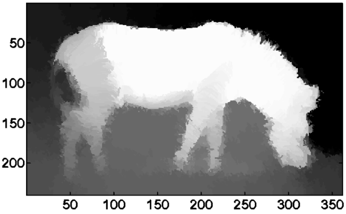


Additional topics of graph Laplacian methods



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



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