CSE 575: Statistical Machine Learning (Spring 2021)
Instructor: Nupur Thakur

Spectral Clustering

Arizona State University

Table of contents

Other types of cuts

- To overcome the drawback of MinCut, one obvious solution is to specify that the partitions are reasonably large.

$$\mathrm{RatioCut}(A_1, \ldots, A_k) := \frac{1}{2}\sum_{i=1}^{k} \frac{W(A_i, \overline{A}_i)}{|A_i|} = \sum_{i=1}^{k} \frac{\mathrm{cut}(A_i, \overline{A}_i)}{|A_i|}$$

No. of vertices in a cluster A_i

$$\mathrm{Ncut}(A_1, \ldots, A_k) := \frac{1}{2}\sum_{i=1}^{k} \frac{W(A_i, \overline{A}_i)}{\mathrm{vol}(A_i)} = \sum_{i=1}^{k} \frac{\mathrm{cut}(A_i, \overline{A}_i)}{\mathrm{vol}(A_i)}.$$

$$\mathrm{MinMaxCut}(A_1, \ldots, A_k) := \sum_{i=1}^{k} \frac{\mathrm{cut}(A_i, \overline{A}_i)}{W(A_i, \overline{A}_i)}.$$

Page No. 25 of the doc in Recommended Resources

Properties of Graph Laplacian

- **L = D - W** → Adjacency Matrix
  - D → Degree matrix

$$D = \underbrace{\begin{bmatrix} d_1 & \cdots & & 0 \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{bmatrix}}_{n \times n}$$

- Properties of **L** -
  - *For every vector $f \in \mathbb{R}^n$ we have*

$$\boxed{f'Lf \geq 0}$$

$$f'Lf = \frac{1}{2}\sum_{i,j=1}^{n} \underbrace{w_{ij}}_{+ve}\underbrace{(f_i - f_j)^2}_{+ve}$$

$f' \quad f$

Proof - By the definition of d_i,

$$f'Lf = f'Df - f'Wf = \sum_{i=1}^{n} \underbrace{d_i}_{\text{Element of } D} f_i^2 - \sum_{i,j=1}^{n} f_i f_j \underbrace{w_{ij}}_{\text{Element in } W}$$

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

$(a-b)^2 = a^2 - 2ab + b^2$

Reference: https://arxiv.org/pdf/0711.0189.pdf and https://sites.cs.ucsb.edu/~veronika/SpectralClustering.pdf

Properties of Graph Laplacian

- Properties of **L** -
  - **L** is symmetric and positive semi-definite

D - symmetric
W - Symmetric

From the first property, $f' L f \geq 0$

Positive semi-definite matrix.

Reference: https://arxiv.org/pdf/0711.0189.pdf and https://sites.cs.ucsb.edu/~veronika/SpectralClustering.pdf

Properties of Graph Laplacian

$$d(0)=2 \qquad \begin{array}{c|cccc} & 0 & 1 & 2 & 3 \\ \hline 0 & 0 & 1 & 1 & 0 \\ 1 & & & & \\ 2 & & & & \\ 3 & & & & \end{array}$$

- Properties of **L** -
  - The smallest eigenvalue is 0, the corresponding eigenvector is the 1-vector (all elements being 1)

Proof –

$$L = D - W = \begin{pmatrix} d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_n \end{pmatrix} - \begin{pmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & w_{nn} \end{pmatrix} = \begin{pmatrix} d_1 - w_{11} & -w_{12} & \cdots & -w_{1n} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{j=1}^{n} w_{1j} - w_{11} & \cdots & -w_{1n} \\ \vdots & \ddots & \vdots \\ -w_{n1} & \cdots & \sum_{j=1}^{n} w_{nj} - w_{nn} \end{pmatrix}.$$

$$\Rightarrow \underbrace{\begin{pmatrix} \sum_{j=1}^{n} w_{1j} - w_{11} & \cdots & -w_{1n} \\ \vdots & \ddots & \vdots \\ -w_{n1} & \cdots & \sum_{j=1}^{n} w_{nj} - w_{nn} \end{pmatrix}}_{n \times n} \underbrace{\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}}_{n \times 1}$$

$$n \times 1 \quad \begin{pmatrix} \sum_{j=1}^{n} w_{1j} - w_{11} - \sum_{j=2}^{n} w_{1j} \\ \sum_{j=1}^{n-1}(-w_{nj}) - \sum_{j=1}^{n} w_{nj} - w_{nn} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\boxed{Lx = \lambda x}, \quad \lambda = 0$$

Reference: https://arxiv.org/pdf/0711.0189.pdf and https://sites.cs.ucsb.edu/~veronika/SpectralClustering.pdf

Properties of Graph Laplacian

- Properties of **L** -
    - L has n non-negative, real-valued eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$

Property 1 & 2

Property 3

Refer for all more details on the properties: https://arxiv.org/pdf/0711.0189.pdf and https://sites.cs.ucsb.edu/~veronika/SpectralClustering.pdf

MinCut Problem

- Given a similarity graph with adjacency matrix W, the simplest and most direct way to construct a partition of the graph is to solve the mincut problem.

Minimize $$\underline{\underline{\operatorname{cut}(A_1, \ldots, A_k) := \frac{1}{2} \sum_{i=1}^{k} W(A_i, \overline{A}_i)}}$$

MinCut for k=2

- Given **W** and a cluster membership vector **q**,

$$q_i = \begin{cases} 1 & i \in Cluster\ A \\ -1 & i \in Cluster\ B \end{cases}$$

$$\boxed{\mathbf{q} = \underset{\mathbf{q}\in[-1,\ 1]^n}{\operatorname{argmin}} \frac{1}{4} \sum_{i,j} (q_i - q_j)^2 w_{i,j}}$$

2^n → n → no. of samples

⇒ exponential complexity.

Cut size

Relaxation Approach

$$\begin{cases} \mathbf{q} = \underset{\mathbf{q}}{\operatorname{argmin}}\ J = \underset{\mathbf{q}}{\operatorname{argmin}}\ \boxed{\mathbf{q}^T(\mathbf{D}-\mathbf{W})\mathbf{q}} \\ \text{subject to } \boxed{\sum_{i=1}^{n} q_i^2 = n} \end{cases}$$

$n \to$ no. of samples

$q^T L q$

eigenvector corresponding to second smallest eigenvalue of L is the soln

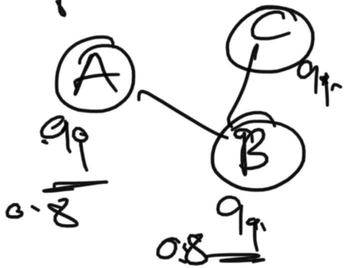
$q_i \Rightarrow$ any real value betn $-1, 1$

$q = [0.8 \quad -0.2 \quad 0.1]$

$\downarrow$

$k = 2 \quad \{ q > 0 \Rightarrow 1 \quad q < 0 \Rightarrow -1 \}$

$q_i = 0.8$



Questions?