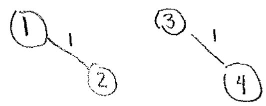
Spectral Clustering

Ex. Two-armed spiral and concentric rings

See Canvas > Files > Numerical Examples > Clustering

Ex: Consider the following simple graph



A = adjacency (affinity) matrix

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

A_{uv} = weight of edge connecting nodes u, v

D = degree matrix

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

$D_{uu} = \sum_v A_{uv}$ = # nodes connected to u

L = unnormalized graph Laplacian $= D - A = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$

Observe:

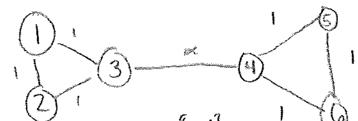
- L is symmetric (diagonalizable)
- Given vector $\vec{x}$, $\vec{x}^T L \vec{x} = \vec{x}^T D \vec{x} - \vec{x}^T L \vec{x}$

$$= \frac{1}{2} \sum_i \sum_j A_{ij} (x_i - x_j)^2 \geq 0 \quad \text{since } A_{ij} \geq 0$$

$\therefore$ L is positive-semidefinite

- 0 is an eigenvalue w/ eigenvector $\mathbb{1}_4$ (also has eigenvector $(1,1,0,0)^T$)
    - 0 is eigenvalue w/ geometric/algebraic multiplicity 2 = # connected components!
- Unchanged if we permute rows/columns (re-label the nodes) but keep the same connections/structure

Ex:



$\alpha \in \{0,1\}$

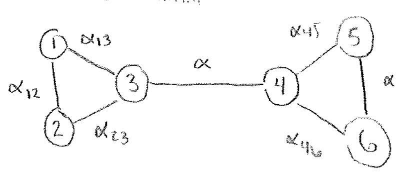
$$L = \begin{bmatrix} 2 & -1 & -1 & & & \\ -1 & 2 & -1 & & 0 & \\ -1 & -1 & 2+\alpha & -\alpha & & \\ & & -\alpha & 2+\alpha & -1 & -1 \\ & 0 & & & 2 & -1 & -1 \\ & & & & -1 & -1 & 2 \end{bmatrix}$$

- L still symmetric, positive-semi-definite
- 0 is an eigenvalue

Eigenvectors	$\alpha =$
$\mathbb{1}_4$	0
$(1,1,1,0,0,0)^T$ $(0,0,0,1,1,1)^T$	1

- If $\alpha = 1$, $\lambda_2 = 0.4384$ is the second smallest eigenvalue
  - λ_2 approximates the sparsest cut of the graph, i.e. the ratio of the fewest # of removed edges (which splits the graph into 2 disconnected pieces) divided by the # of vertices in the smaller of the 2 graphs ($1/3$ in example when $\alpha = 1$)
  - Sparsest cut related to clustering nodes into $K=2$ clusters with minimal removal of edges

Ex. Weighted Graph



$$L = \begin{bmatrix} \alpha_{12}+\alpha_{13} & -\alpha_{12} & -\alpha_{13} & 0 & 0 & 0 \\ & & & 0 & 0 & 0 \\ & & & \alpha & 0 & 0 \\ 0 & 0 & \alpha & & & \\ 0 & 0 & 0 & & & \\ 0 & 0 & 0 & & & \end{bmatrix}$$

- Relationship between λ_2 and α unchanged ($\lambda_2 = 0$ iff $\alpha = 0$)

* Spectrum of L is giving information about # of connected components of graph!

ASIDE: WHY IS IT CALLED THE GRAPH LAPLACIAN?

SUPPOSE $\phi: V \longrightarrow \mathbb{R}^V$ DESCRIBES THE DISTRIBUTION OF HEAT

(V: SET OF NODES; $\mathbb{R}^V$: HEAT IN NODES)

ACROSS A GRAPH $G = (V, \mathcal{E})$; $\phi_i(t)$ IS THE HEAT OF NODE i AT TIME t.

NEWTON'S LAW OF COOLING

THE HEAT TRANSFERRED BETWEEN NODES i AND j IS PROPORTIONAL (W/ CONSTANT $-A_{ij}$) TO $\phi_i - \phi_j$, THE DIFFERENCE IN HEAT BETWEEN NODES

$$\frac{d\phi_i(t)}{dt} = -\sum_j A_{ij}\left(\phi_i(t) - \phi_j(t)\right)$$

$$= -\phi_i \sum_j A_{ij} - \sum_j A_{ij}\phi_j$$

$$= -\phi_i D_{ii} - \sum_j A_{ij}\phi_j = -\sum_j L_{ij}\phi_j$$

IN MATRIX FORM

$$\frac{d}{dt}\phi = -L\phi$$

THIS IS THE SAME AS THE HEAT EQUATION W/ Δ, THE LAPLACIAN, REPLACED BY $-L$, HENCE "GRAPH LAPLACIAN"

THM: LET $G = (V, \mathcal{E})$ BE A WEIGHTED GRAPH, AND LET $L = D - A$ BE ITS GRAPH LAPLACIAN. THEN

i) THE NUMBER c OF CONNECTED COMPONENTS $K_1 \ldots K_c$, IS EQUAL TO THE DIMENSION OF THE NULLSPACE OF L, WHICH IS EQUAL TO THE GEOMETRIC MULTIPLICITY OF $\lambda = 0$.

ii) THE NULLSPACE HAS A BASIS CONSISTING OF INDICATOR VECTORS OF THE CONNECTED COMPONENTS OF G, THAT IS VECTORS $(v_1 \ldots v_N)^T$ S.T. $v_j = 1$ IFF NODE j IN COMPONENT K_i, 0 OTHERWISE. FOR $i = 1 \ldots, c$

Data Similarity : Graph Laplacians

Given data $\vec{x}_1, \ldots \vec{x}_N$ construct an $N \times N$ affinity matrix, A, so that $A_{ij} = A_{ji}$ is the similarity between $\vec{x}_i$ and $\vec{x}_j$

* Common choice is $A_{ij} = \exp\left(\dfrac{-\|\vec{x}_i - \vec{x}_j\|_2}{2\sigma^2} \right)$

Gaussian radial basis function

> <u>Equivalently</u>
>
> $\exp\left(-\gamma \|x_i - x_j\|^2\right)$
>
> $\exp\left(-\|x_i - x_j\|^2 / c\right)$

Then create graph Laplacian

$$L = D - A$$

, note $A_{ii} = 1$ but cancels in L

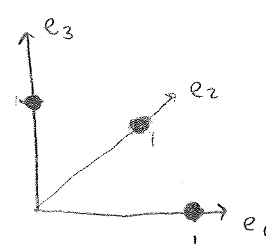
$D_{ii} = \sum_j A_{ij} = 1 + \sum_{j \neq i} A_{ij}$

$L_{ii} = D_{ii} - A_{ii} = \sum_{j \neq i} A_{ij}$

Ideal case:

- $\|x_i - x_j\| = \infty$ when $\vec{x}_i, \vec{x}_j$ in different components
- L has eigenvalue 0 w/ multiplicity K
- $\vec{y}_1, \ldots \vec{y}_k \in \mathbb{R}^N$ linearly independent eigenvectors w/ eigenvalue 0
- Treat $Y = \begin{bmatrix} y_1 & \cdots & y_k \end{bmatrix} \in \mathbb{R}^{N \times k}$ as a data matrix

What does graph of (rows) of Y look like?



All points in cluster i mapped to $\vec{e}_i$! (Ideally?)

Applying Kmeans clustering here works very, very well!!

NOTE: EVEN IN THE IDEAL CASE IT IS UNLIKELY THAT EIGENVALUE SOLVER WILL GIVE

$$Y = \begin{bmatrix} 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & \vdots \\ 0 & 1 & \\ \vdots & \vdots & 0 \\ 0 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \end{bmatrix}$$

HOWEVER, $\vec{y}_1, \ldots, \vec{y}_k$ WILL BE ORTHOGONAL (ORTHONORMAL IF NORMALIZED) BY THE SPECTRAL DECOMPOSITION THM. (LIMITS THE INTERPRETABILITY OF Y)

$\Rightarrow$ CAN CREATE K DISTINCT CLUSTERS ON THE UNIT SPHERE IN $\mathbb{R}^K$ WHICH ARE WELL SEPARATED!

$\Rightarrow$ APPLYING K-MEANS CLUSTERING WILL THEN CAPTURE THE K CLUSTERS IN OUR DATA!

* IN CASES OF EXTREME σ OR $\|x_i - x_j\|$, TRADITIONAL EIGENVALUE SOLVERS CAN STRUGGLE BUT L IS POSITIVE SEMIDEFINITE SO THAT USING SVD GIVES MORE RELIABLE RESULTS

THE UNIDEALIZED CASE (AKA REALITY)

- MANY, MANY CHOICES FOR CONSTRUCTING THE AFFINITY MATRIX A
  - $\exp\left(\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$ IS JUST ONE OF MANY DIFFERENT "KERNELS" THAT ONE CAN USE. MORE ON THIS NEXT WEEK.
  - STILL NEED TO TUNE σ^2 IN RADIAL BASIS FUNCTION, $\exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$
  - CAN ALSO CHOOSE TO ONLY USE NEIGHBORHOOD INFORMATION (HASTIE TEXT)

    i.e. SET $A_{ij} = 0$ IF i, j ARE NOT K-NEAREST NEIGHBORS WHICH GIVES A SPARSE AFFINITY MATRIX

  - UNNORMALIZED : $L = D - A$

    SYMMETRIC NORMALIZED : $L^{SYM} = D^{-1/2} L D^{-1/2} = I - D^{-1/2} A D^{-1/2}$

    RANDOM WALK NORMALIZED : $L^{RW} = D^{-1} L = I - D^{-1} A$
- ALSO
  - # OF EIGENVECTORS OF L TO USE (NORMALIZATION)
  - # OF CLUSTERS

CASE STUDY : ON SPECTRAL CLUSTERING (NG, JORDAN, WEISS)

* AUTHORS FOCUS ON $L = D^{-1/2} A D^{-1/2}$ SENDING $\lambda \mapsto 1-\lambda$ FROM OUR CONVENTION
* IMPLEMENTED IN SPECC FUNCTION

ALGORITHM:

GIVEN $\vec{x}_1, \ldots \vec{x}_n \in \mathbb{R}^d$ THAT WE WANT TO CLUSTER INTO k SUBSETS

1) FORM AFFINITY $A \in \mathbb{R}^{n \times n}$

$$A_{ij} = \exp\left(-\|x_i - x_j\|^2 / 2\sigma^2\right) \quad i \neq j \,, \quad A_{ii} = 0$$

2) DEFINE $D \in \mathbb{R}^{n \times n}$ TO BE DIAGONAL MATRIX W/ $D_{ii} = \sum_j A_{ij}$

AND LET $L = D^{-1/2} A D^{-1/2}$

3) FIND THE k-LARGEST EIGENVALUES AND THEIR ASSOCIATED EIGENVECTORS (CHOSEN TO BE ORTHOGONAL IF EIGENVALS ARE REPEATED) $\vec{w}_1, \ldots \vec{w}_k \in \mathbb{R}^n$ AND FORM $W = [\vec{w}_1, \ldots \vec{w}_k]$

4) FORM $Y \in \mathbb{R}^{n \times k}$ FROM W BY

$$Y_{ij} = \frac{W_{ij}}{\left(\sum_j W_{ij}^2\right)^{1/2}} \qquad \text{(PUT ROWS ON UNIT SPHERE)}$$

5) TREAT ROWS OF Y AS POINTS IN $\mathbb{R}^k$, CLUSTER INTO k-CLUSTERS VIA k-MEANS

6) ASSIGN $\vec{x}_i$ TO CLUSTER j IFF ROW i OF Y WAS ASSIGNED TO CLUSTER j.

* ANALYSIS OF NG, JORDAN, AND WEISS GIVES A LESS AD HOC METHOD FOR CHOOSING / TUNING σ!
    * SEARCH OVER A RANGE OF σ^2 (OBTAINED BY LOOKING AT VARIABILITY IN A SUBSET OF DATA) AND CHOOSE VALUE GIVING THE TIGHTEST CLUSTERS! (LOWEST WITHIN CLUSTER SUM OF SQUARES)

* CAN COMPARE PERFORMANCE BY LOOKING AT TOTAL OF WITHINSS AS GLOBAL MEASURE OF ERROR

EX: SPIRAL & CONCENTRIC RINGS

CASSINI (IN-CLASS EXERCISE)