Session 8: SVD, spectral clustering on graphs

Optimization and Computational Linear Algebra for Data Science

Léo Miolane

Contents

1. Singular Value Decomposition
2. Graphs and Graph Laplacian
3. Spectral clustering

Midterm next week

"Demo" available on Gradescope.

- Thu. Oct. 29, the questions have to be downloaded from Gradescope between 00:01 AM and 9:59 PM. ← Time in NYC
- **Duration:** 1 hour and 40 minutes to work on the problems + 20 minutes to scan and upload your work.
- Upload your work **as a single PDF**.
- In case the upload does not work for you, **email me your work**.

Singular Value Decomposition

Singular Value decomposition

Theorem

Let $A \in \mathbb{R}^{n \times m}$. Then there exists two orthogonal matrices $U \in \mathbb{R}^{n \times n}$ and $V \in \mathbb{R}^{m \times m}$ and a matrix $\Sigma \in \mathbb{R}^{n \times m}$ such that $\Sigma_{1,1} \geq \Sigma_{2,2} \geq \cdots \geq 0$ and $\Sigma_{i,j} = 0$ for $i \neq j$, that verify

$$\boxed{A = U \Sigma V^{\mathsf{T}}.}$$

left singular vectors

singular values

right-singular vectors

$$A = \underbrace{\begin{array}{c|c} & n \\ \hline n & \begin{bmatrix} | & & | \\ u_1 & \cdots & u_n \\ | & & | \end{bmatrix} \end{array}}_{U} \underbrace{\begin{array}{c|c} & m \\ \hline n & \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & (0) & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & 0 & \cdots \; 0 \end{bmatrix} \end{array}}_{\Sigma} \underbrace{\begin{array}{c|c} & m \\ \hline m & \begin{bmatrix} — & v_1^{\mathsf{T}} & — \\ & \vdots & \\ — & v_m^{\mathsf{T}} & — \end{bmatrix} \end{array}}_{V^{\mathsf{T}}}$$

Remarks

$A = U\Sigma V^T$ Sp. Th: $A = P$

- $AA^T = U\Sigma \overbrace{V^T V}^{= Id} \Sigma^T U^T$

$$= U \underbrace{\begin{pmatrix} \sigma_1^2 & & & & \\ & \ddots & & & \\ & & \sigma_r^2 & & \\ & & & 0 & \\ & & & & \ddots \; 0 \end{pmatrix}}_{n} \updownarrow n \; U^T$$

the left singular vectors are eigenvectors of AA^T

- $A^TA = V \underbrace{\begin{pmatrix} \sigma_1^2 & & & & \\ & \ddots & & & \\ & & \sigma_r^2 & & \\ & & & 0 & \\ & & & & \ddots \; 0 \end{pmatrix}}_{m} \updownarrow m \; V^T$

— right —

———

— A^TA

The matrices AA^T and A^TA have the same non-zero eigenvalues.

SVD

$$AA^T : \quad U \begin{pmatrix} \sigma_1^2 & & & & \\ & \ddots & & & \\ & & \sigma_r^2 & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{pmatrix} U^T$$

$\sigma_1^2 \; - - - \; \sigma_r^2 \; 0 \; - - - \; 0$

are eigenval. of AA^T

Spectral theorem:

$$A = P \begin{pmatrix} \lambda_1 & & (0) \\ & \ddots & \\ (0) & & \lambda_n \end{pmatrix} P^T$$

$$AA^T = P \begin{pmatrix} \lambda_1^2 & & \\ & \ddots & \\ & & \lambda_n^2 \end{pmatrix} P^T$$

$\lambda_1^2 \; - - - \; \lambda_n^2$

are eigenval of AA^T

Assume $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|$

$$|\lambda_1| = |\sigma_1| = \sigma_1$$

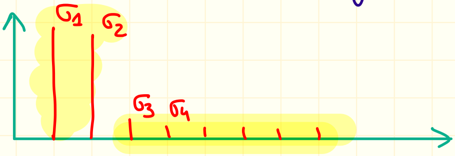
$$\lambda_1^2 = \sigma_1^2$$
$$\vdots$$
$$\lambda_r^2 = \sigma_r^2$$
$$\vdots$$

Low-rank approximation

How can we approximate a matrix A by a matrix of «small» rank ?

① Compute the SVD : $A = U \Sigma V^T$

② Look at the singular values.



③ The matrix $\tilde{A} = U \begin{pmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & 0 & & \\ & & & \ddots & \\ & & & & 0 \end{pmatrix} V^T$

is a "good" rank-2 approximation of A.

$$\boxed{A = U \Sigma V^T}$$

10

4

$$\Sigma = \begin{pmatrix} 3 & & & & & \\ & 2 & & & & \\ & & \overset{0}{0{,}1} & & & \\ & & & \overset{0}{0{,}1} & \underbrace{0 \; \cdots \; 0}_{6} \end{pmatrix}$$

$$\mathbb{R}^{10} \longrightarrow \mathbb{R}^{4}$$

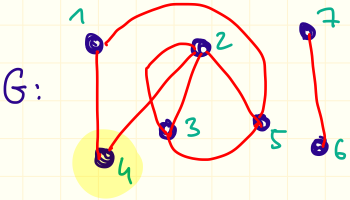
$$A^T A = V \begin{pmatrix} \sigma_1^2 & \\ & \ddots \end{pmatrix} V^T$$

Graphs and Graph Laplacian

Graphs

Graph : "nodes" connected by "edges"

nb. of nodes



We define the adjacency matrix $A \in \mathbb{R}^{n \times n}$ of G.

$$A_{i,j} = \begin{cases} 1 & \text{if } i \sim j \\ 0 & \text{otherwise} \end{cases}$$

"i connected to j"

- The degree $\deg(i)$ of a node i is defined as the number of neighbours of 'i'

$$D = \begin{pmatrix} \deg(1) & & (0) \\ & \ddots & \\ (0) & & \deg(n) \end{pmatrix}$$

"degree matrix"

Graph Laplacian

Definition

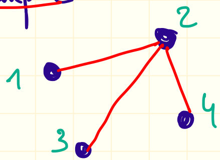
The Laplacian matrix of G is defined as

$$\boxed{L = D - A.}$$

A, D, L are all symmetric

Example

G:



$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \qquad D = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$L = D - A = \begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 3 & -1 & -1 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix}$$

Graph Laplacian

$i \sim j$ = edge between i and j

Definition

The Laplacian matrix of G is defined as

$$L = D - A.$$

sum over all edges of the graph

For all $x \in \mathbb{R}^n$, $\boxed{x^\top L x = \sum_{i \sim j} (x_i - x_j)^2.}$

Proof:

$$L = \sum_{i \sim j} \underbrace{\begin{array}{c|cc|l} & i & j & \\ \hline & 1 & -1 & \leftarrow i \\ & -1 & 1 & \leftarrow j \end{array}}_{B^{(i,j)}}$$

all the other coeffs are 0.

$$x^\top B^{(i,j)} x = x_i^2 + x_j^2 - x_i x_j - x_j x_i = \boxed{(x_i - x_j)^2}$$

Properties of the Laplacian

For all $x \in \mathbb{R}^n$, $\quad x^\top L x = \boxed{\sum_{i\sim j}(x_i - x_j)^2} \geq 0$

- L is positive semi-definite.
- Let $0 \leq \lambda_1 \leq \underline{\lambda_2} \leq \dots \leq \lambda_n$ be its eigenvalues.

$$\lambda_1 = \min_{\|v\|=1} v^\top L v = \min_{x\neq 0} \frac{x^\top L x}{\|x\|^2} \geq 0$$

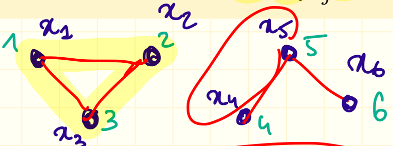
Taking: $\boxed{x = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}}$ we see that $x^\top L x = 0$

$$\longrightarrow \boxed{\lambda_1 = 0}$$

Properties of the Laplacian

For all $x \in \mathbb{R}^n$, $$x^\top L x = \sum_{i\sim j} \boxed{(x_i - x_j)^2}.$$

G:



G has "2 connected components"

$x \in \mathrm{Ker}(L) \iff x^\top L x = 0$ because L is PSD

$\iff x_i = x_j$ for all $i \sim j$

$\iff$ $x_i = x_j$ for all i and j in the same "connected component"

The vectors: $x = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$ and $y = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}$ are a basis of $\mathrm{Ker}(L)$

$\dim \mathrm{Ker}(L) = 2$.

if $x \in \ker(L)$ $\quad Lx = 0 \implies x^T L x = 0$

L . PSD

Assume $\boxed{x^T L x = 0}$

then x is a minimizer of $f(v) = \dfrac{v^T L v}{\|v\|^2} \geq 0$

over $\mathbb{R}^n \setminus \{0\}$ because $\begin{cases} f(x) = 0 \\ f(v) \geq 0 \end{cases}$ for all $v \neq 0$

- Minimizers of f are eigenvectors of L associated with the smallest eigenvalue. $\rightarrow$ x is an eigenvector of L associated with $\lambda_1 = 0$

$\rightarrow \boxed{x \in \ker(L)}$

$$\lambda_1 = \min_{\|v\|=1} v^T L v = \min_{u \neq 0} \frac{u^T L u}{\|u\|^2} = \underbrace{\left(\frac{u}{\|u\|}\right)^T}_{} L \underbrace{\left(\frac{u}{\|u\|}\right)}_{}$$

unit vectors

$$v = \frac{u}{\|u\|}$$

$$|\lambda_i| = \sqrt{\sigma_i}$$ Correct

$$-Id = \begin{pmatrix} -1 & & \\ & \ddots & \\ & & -1 \end{pmatrix}$$

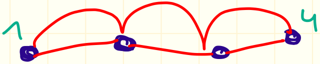
$$-1 + \sqrt{\sigma_i} \geq 0$$

$$\cancel{\lambda_i = \sqrt{\sigma_i}}$$

False

Algebraic connectivity

$\dim \operatorname{Ker}(L)$



Proposition

- The multiplicity of the eigenvalue 0 of L (i.e. the number of i such that $\lambda_i = 0$) is equal to the number of connected components of G.
- In particular, G is connected if and only if $\lambda_2 > 0$.

$$0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$$

- λ_2 is sometimes called the «algebraic connectivity» of G and measures somehow how well G is connected.
- From now, we assume that G is connected, i.e. $\lambda_2 > 0$.

Exercise: show that λ_2 increases when one adds edges to G.

Spectral clustering with the Laplacian

Spectral clustering algorithm

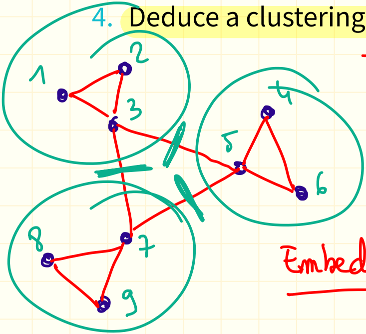
Input: Graph Laplacian L, number of clusters k

ith coord of v_2

1. Compute the first k orthonormal eigenvectors $v_1, \dots, v_k$ of the Laplacian matrix L.

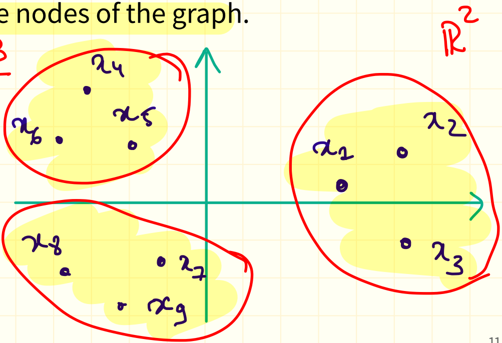
   $\lambda_1 \leq \dots \leq \lambda_k$

2. Associate to each node i the vector $x_i = (v_2(i), \dots, v_k(i))$. $\in \mathbb{R}^{k-1}$
3. Cluster the points $x_1, \dots, x_n$ with (for instance) the k-means algorithm.
4. Deduce a clustering of the nodes of the graph.



$k=3$

Embeding



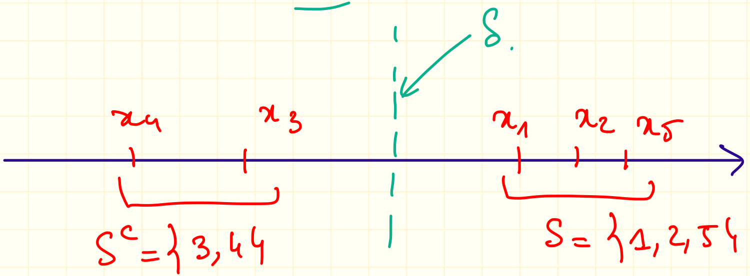
The case of two groups

For $k = 2$ groups:

1. Compute the second eigenvector v_2 of the Laplacian matrix L.
2. Associate to each node i the number $x_i = v_2(i)$.
3. Cluster the nodes in:

$$S = \{i \,|\, v_2(i) \geq \delta\} \qquad \text{and} \qquad S^c = \{i \,|\, v_2(i) < \delta\},$$

for some $\delta \in \mathbb{R}$. $\delta = 0$

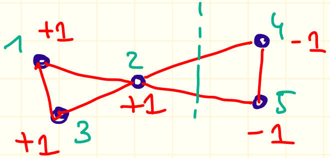


Cut of a partition

Let $S \subset \{1, \dots, n\}$ we define the "cut" of S by:

$\text{cut}(S)$ = " number of edges between S and S^c "

Ex



$S = \{1, 2, 3\}$

$\text{cut}(S) = 2.$

We encode S by a vector $x \in \{+1, -1\}^n$

defined by:

$$\boxed{x_i = \begin{cases} 1 & \text{if } i \in S \\ -1 & \text{otherwise} \end{cases}}$$

in the example

$$x = \begin{pmatrix} 1 \\ 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}$$

Minimal cut problem

Recall $\quad x^T L x = \sum_{i \sim j} (x_i - x_j)^2$

Hence : $\quad \boxed{\mathrm{cut}(S) = \frac{1}{4} \, x^T L x .}$

Goal : find S such that $\mathrm{cut}(S)$ is small

or equivalently find $x \in \{-1, 1\}^n$ such that $x^T L x$ is small

We ask moreover that $\# S = \# S^c$

i.e. $\quad \boxed{x \perp \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}}$ $\quad$ "two clusters of equal sizes"

« Min-Cut » is NP-Hard

Goal: minimize $x^{\mathsf{T}} L x$ subject to $\begin{cases} x \in \{-1, 1\}^n \\ x \perp (1, \ldots, 1). \end{cases}$

Issue: this problem is NP − Hard.

we basically have to compute $\underline{\underline{x^T L x}}$

for each $x \in \{-1, 1\}^n$) 2^n elements

Spectral clustering as a «relaxation»

Idea: We first solve the « relaxed » problem:

minimize $\underbrace{v^\top L v}$ subject to $\begin{cases} \|v\| = \sqrt{n} \\ v \perp (1,\dots,1). \end{cases}$

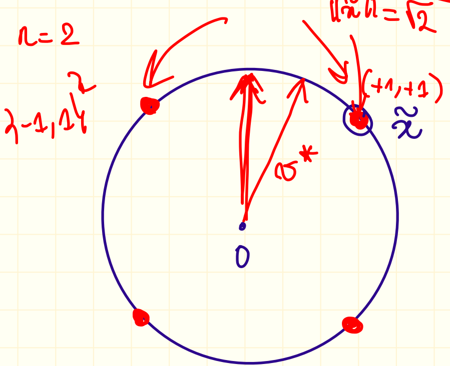
$/\!/ \ v_1$

eig as with λ_2.

The vector $v^* = \sqrt{n}\, v_2$ is a solution

We then define $\tilde{x} \in \{-1,1\}^n$ by

$$\boxed{\tilde{x}_i = \begin{cases} +1 & \text{if } v_2(i) \geq 0 \\ -1 & \text{otherwise} \end{cases}}$$



We "hope" that $\tilde{x}$ leads to a small cut.

Questions?

v_2 is solution to

minimize $v^T L v$ subj. to $\begin{cases} \|v\| = 1 \\ v \perp v_1 \end{cases}$

$v_1 = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$

Questions?