Session 8: Linear Algebra for Graphs (& SVD)

Optimization and Computational Linear Algebra for Data Science

Marylou Gabrié (based on material by Léo Miolane)

Contents

- 1. Singular Value Decomposition
- 2. Graphs and Graph Laplacian
 - 2.1 Definitions
 - 2.2 Properties
- 3. Application: Spectral clustering

1. Singular Value Decomposition

PCA

Data matrix $A \in \mathbb{R}^{n \times m}$

• "Covariance matrix" $S = A^{\mathsf{T}}A \in \mathbb{R}^{m \times m}$.

ightharpoonup S is symmetric (positive semi-definite).

Spectral Theorem: there exists an orthonormal basis v_1, \ldots, v_m of \mathbb{R}^m such that the v_i 's are eigenvectors of S associated with the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_m \geq 0$ (positive semi-definite).

Definitions: Singular values/vectors

For $i = 1, \ldots, m$:

lacktriangle We define $\sigma_i=\sqrt{\lambda_i}$, called the $i^{
m th}$ singular value of A .

Let $r = \operatorname{rank}(A) = \operatorname{number}$ of non-zero λ_i 's (exercise!).

For i = 1, ..., r:

- We call $u_i = \frac{1}{\sigma_i} A v_i$ the i^{th} left singular vector of A.
- $u_1, \cdots u_r$ are orhonormal.
- If r < n, we add $u_{r+1}, \cdots u_n$ such that $u_1, \cdots u_n$ is an orthonormal basis of \mathbb{R}^n .

For i = 1, ..., m:

- Observe that we have $Av_i = \sigma_i u_i$ and $A^T u_i = \sigma_i v_i$.
- We call v_i the i^{th} right singular vector of A.

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

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

Remark: While eigendecomposition is for some square matrices, singular value decomposition exists for all rectangular matrices.

Remarks

▶ Right singular vectors v_i 's are eigenvectors of $A^{\top}A \in \mathbb{R}^{m \times m}$, with eigenvalues $\lambda_i = \sigma_i^2$.

Left singular vectors u_i 's are eigenvectors of $AA^{\top} \in \mathbb{R}^{n \times n}$, with eigenvalues $\lambda_i = \sigma_i^2$.

Low-rank approximation

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

Questions?

7/19

Questions?

7/19

2. Graphs and Graph Laplacian

2.1 Definitions: Graphs

Consider a graph G made of n **nodes** with some **edges**:

Definition

The **adjacency matrix** A of G is the $n \times n$ matrix with entries

$$A_{i,j} = \left\{ \begin{array}{l} 1 \text{ if edge between nodes } i \text{ and } j \\ 0 \text{ otherwise} \end{array} \right.$$

Definition

The **degree matrix** $D \in \mathbb{R}^{n \times n}$ of G is the diagonal matrix with

$$D_{i,i} = \#\{\text{neighbors of } i\} = \deg(i)$$

Graph Laplacian

Definition

The Laplacian matrix of ${\cal G}$ is defined as

$$L = D - A.$$

Graph Laplacian

Definition

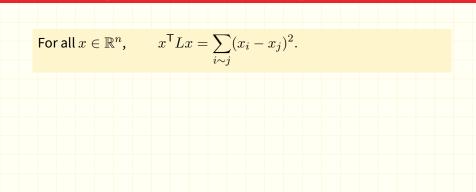
The Laplacian matrix of ${\cal G}$ is defined as

$$L = D - A$$
.

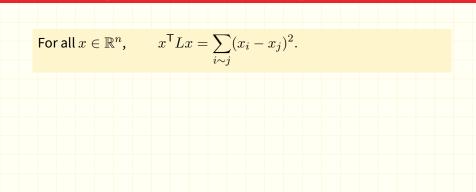
For all
$$x \in \mathbb{R}^n$$
, $x^T L x = \sum_{i \sim j} (x_i - x_j)^2$.

Proof.

2.2 Properties of the Laplacian



2.2 Properties of the Laplacian



Algebraic connectivity

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 $\lambda_2 > 0$.

- 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. $\lambda_2 > 0$.

3. Application: Spectral graph clustering

3.1 Spectral clustering algorithm

Input: Graph Laplacian L, number of clusters k

- 1. Compute the first k orthonormal eigenvectors v_1, \ldots, v_k of the Laplacian matrix L.
- 2. Associate to each node i the vector $x_i = (v_2(i), \dots, v_k(i))$.
- 3. Cluster the points x_1, \ldots, x_n with (for instance) the k-means algorithm.¹
- 4. Deduce a clustering of the nodes of the graph.

14/19

¹ Chap 13 - Elements of Statistical Learning (Hastie, Tibshirani, and Friedman 3. Applica(2001)) 由性使身体身似色的多数由的ford。全体则个hastie/Papers/ESLII.pdf

3.2 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:

$$S = \{i \mid v_2(i) \ge \delta\}$$
 and $S^c = \{i \mid v_2(i) < \delta\},$

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

How does this work?

Let $S \subset \{1, 2, \cdots n\}$.

Definition

The cut of S, denoted $\mathrm{cut}(S)$ is defined as the number of edges between S and S^C .

Ex.

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

Minimal cut problem

Recall
$$x^{\top}Lx = \sum_{i \sim j} (x_i - x_j)^2$$
.

Proposition

For $x \in \{+1, -1\}^n$ representing the subset of nodes S,

$$\operatorname{cut}(S) = \frac{1}{4}x^{\top} L x$$

Minimal cut problem

Recall
$$x^{\top}Lx = \sum_{i \sim j} (x_i - x_j)^2$$
.

Proposition

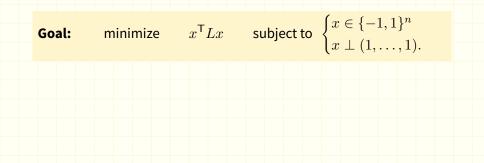
For $x \in \{+1, -1\}^n$ representing the subset of nodes S,

$$\operatorname{cut}(S) = \frac{1}{4}x^{\top}Lx$$

Goal. Find S (or equivalently $x \in \{+1, -1\}^n$) such that

- $\operatorname{cut}(S)$ is small
- $ightharpoonup^{C}$ same number of nodes

« Min-Cut » is NP-Hard

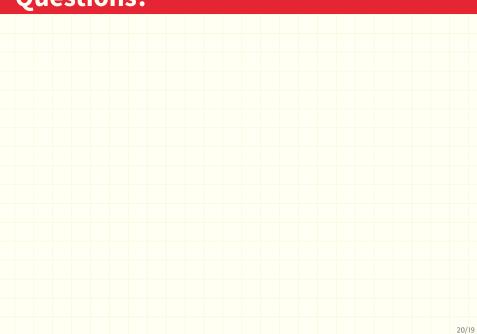


Spectral clustering as a «relaxation»

Idea: We first solve the « relaxed » problem:

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

Questions?



Questions?

