GI07/COMPM012: Mathematical Programming and Research Methods (Part 2)

3. Elements of graph theory and applications to data analysis

Massimiliano Pontil

Outline

- Graph Laplacian

- Graph embedding

- Spectral clustering

Graph

An undirected graph is an ordered pair $G := (V, E)$, where $V = \{1, \ldots, n\}$ is a set of vertices (nodes) and $E \subseteq V \times V$ is a set of edges

Adjacency matrix $A \in \mathbb{R}^{n \times n}$: $A_{ij} = 1$ if vertices i and j are connected and $A_{ij} = 0$ otherwise

Degree of vertex i: $d_i = \sum_{j=1}^{n} A_{ij}$

Degree matrix: $D = \text{diag}(d_1, \ldots, d_n)$

Graph Laplacian

The matrix A induces a natural quadratic form on G:

$$\frac{1}{2}\sum_{i,j=1}^{n} A_{ij}(x_i - x_j)^2$$

We have the important identity:

$$\begin{aligned}
\frac{1}{2}\sum_{i,j=1}^{n} A_{ij}(x_i - x_j)^2 &= \frac{1}{2}\sum_{i,j=1}^{n} A_{ij}(x_i^2 + x_j^2 - 2x_i x_j) \\
&= \sum_{i=1}^{n} x_i^2 d_i - \sum_{ij=1}^{n} A_{ij} x_i x_j \\
&= \sum_{i,j=1}^{n} x_i (D_{ij} - A_{ij}) x_j = x^\top L x
\end{aligned}$$

The matrix $L = D - A$ is called the **graph Laplacian**

Graph Laplacian (cont.)

L is positive semidefinite because $x^\top Lx = \sum_{i,j=1}^{n} A_{ij}(x_i - x_j)^2 \geq 0$

Let λ_i be the eigenvalues of L and let u_i be the corresponding eigenvectors:

$$Lu_i = \lambda_i u_i$$

Here we use the convention that $0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$

Note that if x is a constant vector (ie. all components are equal) then $Lx = 0$

Thus $\lambda_1 = 0$ and $u_1 = \frac{1}{\sqrt{n}}$

Interpretation of eigenvectors

$$Lx = \lambda x \qquad \Longleftrightarrow \qquad x_i = \frac{1}{d_i - \lambda} \sum_{j=1}^{n} A_{ij} x_j$$

If λ is small then each component of the corresponding eigenvector is close to the average of the "neighbor components"

The smaller λ the smoother the corresponding eigenvector (ie. higher eigenvectors tend to be more irregular)

Do eigenvalues and eigenvectors of L help us understanding the structure of the graph?

Connected graphs

A graph G is called *connected* if there is a path between any two vertices

The number of connected components of G is the smallest partition of G in connected subgraphs

Theorem: G in connected if and only if $\lambda_2 > 0$. Moreover, the number of non-zero eigenvalues of G equal the number of connected components of G

Building the graph from data

Given data points $t_1, \ldots, t_n \in \mathbb{R}^d$ we may build a graph as follows:

Let $N_k(t_i)$ be the set of k nearest neighbors of point t_i

Then set $A_{ij} = 1$ if either $j \in N_k(t_i)$ or $i \in N_k(t_j)$ (that is, at least one the two points is one of the k nearest neighbors of the other)

Graph construction may be extended to give weighted graphs: A_{ij} is a non-negative number (weight) indicating the similarity between “objects” i and j

Graph embedding

Linear embedding: map V into the line:

Consider the problem

$$\min\left\{\sum_{i,j=1}^{n} A_{ij}(x_i - x_j)^2 : x \in \mathbb{R}^n, \sum_{i=1}^{n} x_i = 0.\ \sum_{i=1}^{n} x_i^2 = 1\right\}$$

The constraints say that the embedded points have zero mean and unit variance

The solution is $x = u_2$

The line embedding maps vertex the i-th vertex to the point u_{2i} on the line

Planar embedding

Map V into the plane:

$$\min\left\{\sum_{i,j=1}^{n} A_{ij}\|x_i - x_j\|^2 : x_1, \ldots, x_n \in \mathbb{R}^2\right\}$$

We require that $x^1 = (x_{11}, \ldots, x_{n1})$ and $x^2 = (x_{12}, \ldots, x_{n2}) \perp \mathbf{1}$ (i.e. sum to zero) and that $(x^1)^\top x^2 = 0$. We also require that $\|x^1\| = \|x^2\| = 1$

We get the solution $x^1 = u_2$, $x^2 = u_3$

The planar embedding maps the i-th vertex to the point (u_{2i}, u_{3i}) in the plane

Data

The above embeddings give a representation of the data which reflects the similarity between the underlying datapoint/vertices

The embedding is *non-linear*

Very different from PCA!

Spectral clustering

Goal: to find a partition of the vertices of a graph into different groups, such that there are as few edges as possible between nodes in different groups and as many nodes as possible within each group

We discuss the case of two groups, but these ideas can be extended to more groups with no much further difficulty

Ratio Cut

$$\text{cut}(J) = \sum_{i \in J} \sum_{j \in J_c} A_{ij}$$

An optimal partition of the graph into two groups should minimize the quantity

$$\rho(J) = \text{cut}(J) \left(\frac{1}{|J|} + \frac{1}{|J_c|} \right)$$

over all subsets $J \subseteq \{1, \ldots, n\}$

The term $1/|J| + 1/|J_c|$ encourages balanced cuts (eg. it avoids that we simply disconnect only one vertex of small degree)

The above problem is known to be *NP-hard*

Optimization formulation

Let $\gamma = \sqrt{\frac{|J_c|}{|J|}}$ and define the vector $f(J)$ as $f(J)_i = \begin{cases} \gamma & \text{if } i \in J \\ -1/\gamma & \text{if } i \in J_c \end{cases}$

We see that $\sum_{i=1}^n f_i = 0$ and $\|f\|^2 = \sum_{i=1}^n f_i^2 = n$. Moreover:

$$\begin{aligned} f^\top L f &= \frac{1}{2}\sum_{i,j=1}^n A_{ij}(f_i - f_j)^2 \\ &= \frac{1}{2}\sum_{i\in J}\sum_{j\in J_c} A_{ij}(\gamma + 1/\gamma)^2 + \frac{1}{2}\sum_{j\in J}\sum_{i\in J_c} A_{ij}(-1/\gamma - \gamma)^2 \\ &= \text{cut}(J)\left(\gamma^2 + 1/\gamma^2 + 2\right) \\ &= \text{cut}(J)\left(\frac{|J_c|}{|J|} + \frac{|J|}{|J_c|} + 2\right) = n\rho(J) \end{aligned}$$

Optimization formulation

The above observations imply that our problem can be formulated as the problem of minimizing the function

$$f^\top L f, \quad f \in \mathbb{R}^n$$

subject to the constraints

$$1^\top f = 0$$

$$\|f\|^2 = n$$

$$f = f(J),\ J \subseteq \{1, \dots, n\}$$

The problem is difficult due to the last (combinatorial) constraint

Convex relaxation

A simple relaxation to the above problem is obtained by dropping the combinatorial constraint:

$$\min\{f^\top L f : f \in \mathbb{R}^n, 1^\top f = 0,\ \|f\|^2 = n\}$$

The solution is second eigenvector of the Laplacian

There is no guarantee on the quality of the solution relative to the optimal (combinatorial) solution