6. Ideal Clustering

Bookmark this page

Exercises due Oct 20, 2021 17:29 IST Completed

The goal of clustering that we adopt in this module is that we wish to identify nodes that share edges more among themselves when compared to edges that emanate outside of them.

To this end, let us first consider clustering of nodes into two clusters. Let us split an n-node graph into two subsets each of size n_1 and n_2 such that the goal is to minimize the sum of weights of links between them. Define $s_i, i = 1, \ldots, n$ as

$$s_i = \begin{cases} 1 & \text{if vertex } i \text{ belongs to group 1,} \\ -1 & \text{if vertex } i \text{ belongs to group 2.} \end{cases}$$

Then, let us define the optimization problem to obtain an optimal clustering as

$$C = \min_{\mathbf{s} \in \{-1,1\}^n} \sum_{i,j} A_{i,j} \left(1 - s_i s_j\right), \quad \text{such that } \sum_k s_k = n_1 - n_2.$$

Self exercise: Show that we can equivalently write the above optimization problem as

$$C = \min_{\mathbf{s} \in \{-1,1\}^n} \mathbf{s}^T L \mathbf{s}, \quad \text{such that } \sum_k s_k = n_1 - n_2,$$

where L is the graph Laplacian.

Number of Splits

2/2 points (graded)

In this problem, we explore the complexity of solving the aforementioned clustering optimization problem with the straightforward, naive solution.

1. Say we have 25 nodes in a graph. In how many ways can we split the graph into two distinct subsets with 12 nodes in subset A and 13 in subset B?

5200300 ✔ **Answer:** 5200300

2. How many total ways are there to partition the 25-node graph into two distinct subsets A and B? (Empty sets are valid subsets.)

33554432 ✔ **Answer:** 2^25

Solution:

1. **5200300.** Choosing 12 nodes out of 25 can be done in $\binom{25}{12} = 5200300$ ways.
2. 2^{25}**.** This is a consequence of the binomial theorem.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Complexity of Ideal Clustering

1/1 point (graded)

Pick from the following the theoretical complexity of the naive, straightforward integer programming approach to solving the optimization problem.

- ○ $\mathcal{O}(n)$
- ○ $\mathcal{O}(n^2)$
- ◉ $\mathcal{O}(2^n)$
- ○ $\mathcal{O}(n \log(n))$

✔

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Discussion

Topic: Module 3: Network Analysis:Spectral Clustering / 6. Ideal Clustering

Hide Discussion

Add a Post

All Posts

Self-exercise

discussion posted 2 months ago by anonymous

I was not able to reach it. Did someone do it or perhaps find materials that show how to reach the compact matrix form?

Thanks!

This post is visible to everyone.

Add a Response

1 response

Syed_SB
2 months ago

$s^T Ls = s^T Ds - s^T As$. Now $s^T Ds$ is the sum of diagonal terms which is equal to sum of all entries of the matrix A ($\sum_{ij} A_{ij}$). $s^T As$ is the sum of all entries but each A_{ij} is multiplied by s_i and s_j ($\sum_{ij} A_{ij} s_i s_j$).

Remember $1^T A1$ is equal to sum of all values of the matrix A which is $\sum_{ij} A_{ij}$

posted 2 months ago by **Syed_SB**

Add a comment

Showing all responses

Add a response:

Preview

Submit

‹ Previous Next ›