Introduction to Machine Learning

Spectral Clustering

Mingchen Gao

Outline

Contents

1 Spectral Clustering

- An alternate approach to clustering
- Let the data be a set of N points

$$\mathbf{X} = \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$$

- Let $\mathbf{S}$ be a $N \times N$ **similarity matrix**

$$S_{ij} = sim(\mathbf{x}_i, \mathbf{x}_j)$$

- $sim(,)$ is a similarity function
- Construct a weighted undirected graph from $\mathbf{S}$ with adjacency matrix, $\mathbf{W}$

$$W_{ij} = \begin{cases} sim(\mathbf{x}_i, \mathbf{x}_j) & \text{if } \mathbf{x}_i \text{ is nearest neighbor of } \mathbf{x}_j \\ 0 & otherwise \end{cases}$$

- Can use more than 1 nearest neighbors to construct the graph

- Clustering $\mathbf{X}$ into K clusters is equivalent to finding K cuts in the graph $\mathbf{W}$
  - $A_1, A_2, \ldots, A_K$
- Possible objective function

$$cut(A_1, A_2, \ldots, A_K) \triangleq \frac{1}{2}\sum_{k=1}^{K} W(A_k, \bar{A}_k)$$

- where $\bar{A}_k$ denotes the nodes in the graph which are **not in** A_k and

$$W(A, B) \triangleq \sum_{i \in A, j \in B} W_{ij}$$

Normalized Min-cut Problem

$$normcut(A_1, A_2, \ldots, A_K) \triangleq \frac{1}{2}\sum_{k=1}^{K} \frac{W(A_k, \bar{A}_k)}{vol(A_k)}$$

where $vol(A) \triangleq \sum_{i \in A} d_i$, d_i is the weighted degree of the node i

- Equivalent to solving a 0-1 knapsack problem
- Find N binary vectors, $\mathbf{c}_i$ of length K such that $c_{ik} = 1$ only if point i belongs to cluster k
- If we relax constraints to allow c_{ik} to be real-valued, the problem becomes an eigenvector problem
  - Hence the name: **spectral clustering**

1.1 Graph Laplacian

$$\mathbf{L} \triangleq \mathbf{D} - \mathbf{W}$$

- $\mathbf{D}$ is a diagonal matrix with degree of corresponding node as the diagonal value

Properties of Laplacian Matrix

1. Each row sums to 0
2. $\mathbf{1}$ is an eigen vector with eigen value equal to 0
    - This means that $\mathbf{L1} = 0\mathbf{1}$.
3. Symmetric and positive semi-definite
4. Has N non-negative real-valued eigenvalues
5. If the graph ($\mathbf{W}$) has K connected components, then $\mathbf{L}$ has K eigenvectors spanned by $\mathbf{1_{A_1}}, \ldots, \mathbf{1_{A_K}}$ with 0 eigenvalue.

To see why $\mathbf{L}$ is positive semi-definite:

$$\begin{aligned}
\mathbf{xLx}^\top &= \mathbf{xDx}^\top - \mathbf{xWx}^\top \\
&= \sum_i d_i x_i^2 - \sum_i \sum_j x_i x_j w_{ij} \\
&= \frac{1}{2}\left(\sum_i d_i x_i^2 - 2\sum_i \sum_j x_i x_j w_{ij} + \sum_j d_j x_j^2\right) \\
&= \frac{1}{2}\sum_i \sum_j w_{ij}(x_i - x_j)^2
\end{aligned}$$

which is $\geq 0 \forall \mathbf{x}$.

1.2 Spectral Clustering Algorithm

Observation

- In practice, $\mathbf{W}$ might not have K exactly isolated connected components
- By *perturbation theory*, the smallest eigenvectors of $\mathbf{L}$ will be close to the ideal indicator functions

Algorithm

- Compute first (smallest) K eigen vectors of $\mathbf{L}$
- Let $\mathbf{U}$ be the $N \times K$ matrix with eigenvectors as the columns
- Perform kMeans clustering on the rows of $\mathbf{U}$

References

Murphy Book Chapter 21.5

References