Spectral Clustering 推导

Soul Walker

2020 年 7 月 22 日

给定数据矩阵 $\boldsymbol{X} = (\boldsymbol{x}_1; \boldsymbol{x}_2; \cdots; \boldsymbol{x}_n) \in R^{n \times m}$，训练数据集

$$T = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N\} = \{\boldsymbol{x}_i\}_{i=1}^n$$

使用 $G(V, E)$ 来描述图, 对于 $A \subset V$, $|A|$ 为子集 A 中的结点个数, $vol(A) = \sum_{i \in A} d_i$ 即为图中所有点度的和。

0.1 背景知识

解决谱聚类问题我们一般通过图割（无向图）的方法来解决。
首先我们定义相似度矩阵 $\boldsymbol{W} \in R^{n \times n}$，其中若结点 i 与结点 j 相连接则 $w_{ij} = 1$，否则 $w_{ij} = 0$。度矩阵 $\boldsymbol{D} \in R^{n \times n}$，其中 $d_i = \sum_{j=1}^n w_{ij}$。拉普拉斯

矩阵 $\boldsymbol{L}=\boldsymbol{D}-\boldsymbol{W}$。其中拉普拉斯矩阵对于向量 $\boldsymbol{f}$ 有

$$
\begin{aligned}
\boldsymbol{f}^T\boldsymbol{L}\boldsymbol{f} &= \boldsymbol{f}^T\boldsymbol{D}\boldsymbol{f}-\boldsymbol{f}^T\boldsymbol{W}\boldsymbol{f} \\
&= \sum_{i=1}^{n} d_i\boldsymbol{f}_i^2-\sum_{i,j=1}^{n}\boldsymbol{f}_i\boldsymbol{f}_j w_{ij} \\
&= \frac{1}{2}(\sum_{i=1}^{n} d_i\boldsymbol{f}_i^2-2\sum_{i,j=1}^{n} w_{ij}\boldsymbol{f}_i\boldsymbol{f}_j+\sum_{j=1}^{n} d_j\boldsymbol{f}_j^2) \\
&= \frac{1}{2}(\sum_{i=1}^{n}(\sum_{a=1}^{n} w_{ia})\boldsymbol{f}_i^2-2\sum_{i,j=1}^{n} w_{ij}\boldsymbol{f}_i\boldsymbol{f}_j+\sum_{j=1}^{n}(\sum_{b=1}^{n} w_{jb})\boldsymbol{f}_j^2) \\
&= \frac{1}{2}(\sum_{i,a=1}^{n} w_{ia}\boldsymbol{f}_i^2-2\sum_{i,j=1}^{n} w_{ij}\boldsymbol{f}_i\boldsymbol{f}_j+\sum_{j,b=1}^{n} w_{jb}\boldsymbol{f}_j^2) \\
&= \frac{1}{2}(\sum_{i,j=1}^{n} w_{ij}\boldsymbol{f}_i^2-2\sum_{i,j=1}^{n} w_{ij}\boldsymbol{f}_i\boldsymbol{f}_j+\sum_{j,i=1}^{n} w_{ji}\boldsymbol{f}_j^2) \\
&= \frac{1}{2}(\sum_{i,j=1}^{n} w_{ij}\boldsymbol{f}_i^2-2\sum_{i,j=1}^{n} w_{ij}\boldsymbol{f}_i\boldsymbol{f}_j+\sum_{j,i=1}^{n} w_{ij}\boldsymbol{f}_j^2) \\
&= \frac{1}{2}\sum_{i,j=1}^{n} w_{ij}(\boldsymbol{f}_i^2-2\boldsymbol{f}_i\boldsymbol{f}_j+\boldsymbol{f}_j^2) \\
&= \frac{1}{2}\sum_{i,j=1}^{n} w_{ij}(\boldsymbol{f}_i-\boldsymbol{f}_j)^2
\end{aligned}
\tag{0.1}
$$

0.2 切图

对于无向图 G 我们有切图

$$
Cut(A,B)=\sum_{\substack{i\,\in\,A\\ j\,\in\,B}} w_{ij} \tag{0.2}
$$

那么我们需要的是 $Cut(A,B)$ 越小越好，因为 $Cut(A,B)$ 越小代表子图 A 与子图 B 之间的关系越弱，但是这种切图方法由于未考虑类内距离度量而会产生误差较大的情况，所以一般并不采取这种切图方法。

0.3 RatioCut

为了解决上节问题，我们定义 RatioCut 切图

$$RatioCut(A,\bar{A}) = \frac{1}{2}(\frac{Cut(A,\bar{A})}{|A|} + \frac{Cut(\bar{A},A)}{|\bar{A}|}) \tag{0.3}$$

我们希望的是 Cut 越小（$|A|$ 越大），那么我们对上式做如下变换

$$\begin{aligned} RatioCut(A,\bar{A}) &= \frac{1}{2}(\frac{Cut(A,\bar{A})}{|A|} + \frac{Cut(\bar{A},A)}{|\bar{A}|}) \\ &= \frac{1}{2}(\sum_{\substack{i\in A \\ j\in B}} w_{ij}(\frac{1}{\sqrt{|A|}} - 0)^2 + \sum_{\substack{i\in A \\ j\in B}} w_{ij}(0 - \frac{1}{|A|})^2) \end{aligned} \tag{0.4}$$

在此处我们使用拉普拉斯矩阵的性质（式 (1)），定义 $\boldsymbol{f} \in R^n$，称 $\boldsymbol{f}$ 为指示向量，定义

$$\boldsymbol{f}_i = \begin{cases} 0, i \in \bar{A} \\ \dfrac{1}{\sqrt{|A|}}, i \in A \end{cases}$$

那么

$$式\ (4) = \frac{1}{2}\sum_{i,j=1}^{n} w_{ij}(\boldsymbol{f}_i - \boldsymbol{f}_j)^2 \tag{0.5}$$

因此，我们可以得出完整函数表达式

$$\begin{aligned} RatioCut(A_1,A_2,\cdots,A_k) &= RatioCut(A_1,\bar{A}_1) + \cdots + RatioCut(A_k,\bar{A}_k) \\ &= \boldsymbol{f}^{1T}\boldsymbol{L}\boldsymbol{f}^1 + \boldsymbol{f}^{2T}\boldsymbol{L}\boldsymbol{f}^2 + \cdots + \boldsymbol{f}^{kT}\boldsymbol{L}\boldsymbol{f}^k \\ &= \sum_{i=1}^{k}\boldsymbol{f}^{iT}\boldsymbol{L}\boldsymbol{f}^i \\ &= tr(\boldsymbol{H}^T\boldsymbol{L}\boldsymbol{H}) \end{aligned} \tag{0.6}$$

其中 $\boldsymbol{H} = (\boldsymbol{f}^1, \boldsymbol{f}^2, \cdots, \boldsymbol{f}^k) \in R^{n\times k}$

那么我们现在的目标就是

$$\begin{cases} \min\limits_{\boldsymbol{H}} tr(\boldsymbol{H}^T\boldsymbol{L}\boldsymbol{H}) \\ s.t. \quad \boldsymbol{H}^T\boldsymbol{H} = \boldsymbol{I} \end{cases} \tag{0.7}$$

那最大值就是 $\boldsymbol{H}^T\boldsymbol{L}\boldsymbol{H}$ 的最大特征值。但是这种表示不准确所以一般得到 $n\times k$ 维的 $\boldsymbol{H}$ 矩阵之后需要对每一行进行 K-Means 算法。

0.4 NormalizedCut

我们定义 $NormalizedCut$ 切图

$$NormalizedCut(A,\bar{A})=\frac{Cut(A,\bar{A})}{vol(A)} \tag{0.8}$$

同理我们使用拉普拉斯矩阵的性质（式 (1)），定义 $\boldsymbol{f}\in R^n$，定义

$$\boldsymbol{f}_i=\begin{cases}0, i\in\bar{A}\\ \frac{1}{\sqrt{vol(A)}}, i\in A\end{cases}$$

我们可以得到优化问题

$$\begin{cases}\min\limits_{\boldsymbol{H}} tr(\boldsymbol{H}^T\boldsymbol{L}\boldsymbol{H})\\ s.t.\quad \boldsymbol{H}^T\boldsymbol{D}\boldsymbol{H}=\boldsymbol{I}\end{cases} \tag{0.9}$$

这里对 $\boldsymbol{H}^T\boldsymbol{D}\boldsymbol{H}=\boldsymbol{I}$ 的推导与式 (7) 中约束条件推导类似如下

$$\begin{aligned}\boldsymbol{H}^T\boldsymbol{D}\boldsymbol{H}&=\sum_{i=1}^{k}((\frac{1}{\sqrt{vol(A_i)}})^2\sum_{j\in A_i}d_j)\\&=\sum_{i=1}^{k}(\frac{1}{vol(A_i)}\sum_{j\in A_i}d_j)\\&=\sum_{i=1}^{k}\frac{1}{vol(A_i)}vol(A_i)\\&=\boldsymbol{I}\end{aligned} \tag{0.10}$$

我们定义 $\boldsymbol{F}=\boldsymbol{D}^{\frac{1}{2}}\boldsymbol{H}$，那么 $\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{F}=\boldsymbol{H}$，那么最终的优化问题为

$$\begin{cases}\min\limits_{F} tr(\boldsymbol{F}^T\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{L}\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{F})\\ s.t.\boldsymbol{F}^T\boldsymbol{F}=\boldsymbol{I}\end{cases} \tag{0.11}$$

那么 $\boldsymbol{F}$ 矩阵的求解由 $\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{L}\boldsymbol{D}^{-\frac{1}{2}}$ 的最小的 k 个特征向量组成。同样在进行 K-Means 算法。