Машинное обучение, ФКН ВШЭ
Семинар №17

1 Лапласиан графа

Пусть задан некоторый взвешенный неориентированный граф $G = (V, E)$, где $V = \{v_1, \dots v_n\}$ – множество вершин. Мы также будем обозначать вершину v_i индексом i. Матрицу смежности обозначим $W = (w_{ij})_{i,j=1,\dots,n}$. Степенью вершины v_i назовём

$$d_i = \sum_{j=1}^{n} w_{ij}$$

Также введём матрицу D как диагональную матрицу с элементами $D_{ii} = d_i$.

Ненормированным Лапласианом графа называется матрица

$$L = D - W$$

Лапласиан графа обладает следующими свойствами:

1. Для произвольного вектора $f \in \mathbb{R}^n$

$$f^T L f = \frac{1}{2} \sum_{i,j} w_{ij}(f_i - f_j)^2$$

Докажем это.

$$f^T L f = f^T D f - f^T W f = \sum_{i=1}^{n} d_i f_i^2 - \sum_{i,j} f_i f_j w_{ij} =$$
$$= \frac{1}{2}\left(\sum_{i=1}^{n} d_i f_i^2 - 2\sum_{i,j} f_i f_j w_{ij} + \sum_{j=1}^{n} d_j f_j^2\right) = \frac{1}{2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} f_i^2 - 2\sum_{i,j} f_i f_j w_{ij} + \sum_{j=1}^{n}\sum_{i=1}^{n} w_{ij} f_j^2\right) =$$
$$= \frac{1}{2}\sum_{i,j} w_{ij}(f_i - f_j)^2$$

2. L – симметричная, положительно полуопределённая матрица.

Симметричность следует из того, что граф неориентированный. Положительная полуопределённость следует из того, что все веса неотрицательные $w_{ij} \geqslant 0 \; \forall i, j$, а также из предыдущего пункта:

$$\forall f \in \mathbb{R}^n \; f^T L f \geqslant 0.$$

3. Наименьшее собственное значение равно 0 и соответствует вектору $\mathbf{1} = (1, \dots, 1)^T$.

2 Graph-cut подход

Пусть мы задали на нашем множестве точек взвешенный неориентированный граф, где вес ребра w_{ij} означает степень сходства объектов v_i, v_j (чем больше w_{ij}, тем больше похожи объекты v_i, v_j). Введём обозначения: $W(A, B) = \sum_{i \in A, j \in B} w_{ij}$, $\overline{A}$ – дополнение множества A. Тогда логично решать задачу кластеризации выборки на k непересекающихся кластеров $A_1, \ldots, A_k$, минимизируя функционал

$$\mathrm{cut}(A_1, \ldots, A_k) = \frac{1}{2} \sum_{i=1}^{k} W(A_i, \overline{A_i}). \tag{2.1}$$

На практике минимизация функционала 2.1 зачастую приводит к тому, что в кластера выделяются отдельные объекты, что является нежелательным. Поэтому данный функционал нормируют следующим образом:

$$\mathrm{RatioCut}(A_1, \ldots, A_k) = \frac{1}{2} \sum_{i=1}^{k} \frac{W(A_i, \overline{A_i})}{|A_i|}, \tag{2.2}$$

где $|A|$ – количество вершин во множестве A.

Рассмотрим оптимизационную задачу для случая, когда $k = 2$:

$$\min_{A \subset V} \mathrm{RatioCut}(A, \overline{A}).$$

Переформулируем её через Лапласиан. Для этого введём вектор $f = (f_1, \ldots, f_n)^T$.

$$f_i = \begin{cases} \sqrt{|\overline{A}|/|A|}, v_i \in A \\ -\sqrt{|A|/|\overline{A}|}, v_i \in \overline{A} \end{cases}$$

И запишем

$$\begin{aligned} f^T L f &= \frac{1}{2} \sum_{i,j} w_{ij} (f_i - f_j)^2 \\ &= \frac{1}{2} \sum_{i \in A, j \in \overline{A}} w_{ij} \left(\sqrt{\frac{|\overline{A}|}{|A|}} + \sqrt{\frac{|A|}{|\overline{A}|}} \right)^2 + \frac{1}{2} \sum_{i \in \overline{A}, j \in A} w_{ij} \left(- \sqrt{\frac{|A|}{|\overline{A}|}} - \sqrt{\frac{|\overline{A}|}{|A|}} \right)^2 \\ &= \frac{1}{2} \left(\sqrt{\frac{|\overline{A}|}{|A|}} + \sqrt{\frac{|A|}{|\overline{A}|}} \right)^2 (W(A, \overline{A}) + W(\overline{A}, A)) \\ &= \frac{1}{2} \left(\frac{|\overline{A}|}{|A|} + 2 + \frac{|A|}{|\overline{A}|} \right) (W(A, \overline{A}) + W(\overline{A}, A)) \\ &= \frac{1}{2} \left(\frac{|\overline{A}| + |A|}{|A|} + \frac{|A| + |\overline{A}|}{|\overline{A}|} \right) (W(A, \overline{A}) + W(\overline{A}, A)) \\ &= n \mathrm{RatioCut}(A, \overline{A}) \end{aligned}$$

Также заметим, что

$$\sum_{i=1}^{n} f_i = |A| \sqrt{\frac{|\overline{A}|}{|A|}} - |\overline{A}| \sqrt{\frac{|A|}{|\overline{A}|}} = 0.$$

И

$$\sum_{i=1}^{n} f_i^2 = |A|\frac{|\overline{A}|}{|A|} + |\overline{A}|\frac{|A|}{|\overline{A}|} = n.$$

Таким образом мы свели задачу минимизации функционала 2.2 к следующей задаче оптимизации:

$$\begin{aligned} & f^T L f \to \min_f \ , \\ & \langle f, \mathbf{1} \rangle = 0 \ , \\ & \|f\| = \sqrt{n} \ . \end{aligned}$$

При этом f обязан принимать дискретные значения, согласно тому, как мы вводили этот вектор. Решать дискретную задачу оптимизации сложно, поэтому будем решать полученную задачу приближённо.

$$\mathcal{L} = f^T L f + \lambda_1 f^T \mathbf{1} + \lambda_2(\|f\| - \sqrt{n})$$

Запишем условие минимума.

$$\nabla_f \mathcal{L} = 2Lf + \lambda_1 \mathbf{1} + \lambda_2 \frac{1}{\|f\|} f = \mathbf{0}$$

Домножим это уравнение слева на $\mathbf{1}^T$, тогда, пользуясь ограничениями получим

$$\begin{gathered} 2\mathbf{1}^T L f + \lambda_1 n = 0 \\ \lambda_1 = \lambda_1^T = -\frac{2}{n} f^T L \mathbf{1} = 0 \end{gathered}$$

В последней строчке мы воспользовались тем, что матрица L симметричная, а также, что $\mathbf{1}$ – собственный вектор матрицы L с собственным значением 0. В таком случае, условие минимума переписывается следующим образом:

$$2Lf + \lambda_2 \frac{1}{\|f\|} f = \mathbf{0}.$$

Учитывая, что $\|f\| = \sqrt{n}$, получаем, что f - собственный вектор матрицы L с собственным значением $æ = (-\lambda_2)/(2\sqrt{n})$. Тогда минимизируемый функционал можно переписать как $f^T L f = æ\|f\|^2 = æn$. Таким образом, минимум достигается на собственном векторе со вторым минимальным собственным значением æ (минимальное собственное значение равно 0 и достигается на собственном векторе $\mathbf{1}$).

Кластеризацию точек можно получить из полученого решения, с помощью решающего правила

$$\begin{cases} v_i \in A, \ \text{если} \ f_i \geqslant 0 \\ v_i \in \overline{A}, \ \text{если} \ f_i < 0 \end{cases}$$

Но, вместо решающего правила, найденные собственные вектора используют как признаки в новом пространстве, в котором можно провести кластеризацию простым метрическим методом, например, k-means. Таким образом, мы получаем следующий алгоритм:

Дано: матрица смежности W, необходимое число кластеров k.

1. Считаем Лапласиан графа $L = D - W$.
2. Находим собственные векторы Лапласиана $u_1, \ldots, u_n$.
3. Выбираем k собственных векторов с наименьшими k собственными значениями. Записываем матрицу $U = (u_1, \ldots, u_k)$, где u_k – k-ый столбец.
4. i-ую строчку матрицы U используем в качестве нового объекта y_i.
5. Кластеризуем новую выборку $\{y_1, \ldots, y_n\}$ с помощью k-means.

Ранее мы рассмотрели оптимизационную задачу для случая, когда $k = 2$. Аналогичным образом можно рассмотреть случай, когда k – произвольное.

Введём матрицу H:

$$H_{ij} = \begin{cases} 1/\sqrt{|A_j|}, \; v_i \in A_j \\ 0, \; \text{иначе} \end{cases}$$

Рассмотрим столбец h_k, он соответствует множеству A_k. В случае, когда множества не пересекаются $H^T H = I$. Запишем

$$\begin{aligned} h_k^T L h_k &= \frac{1}{2} \sum_{i,j} w_{ij} (H_{ik} - H_{jk})^2 \\ &= \frac{1}{2} \sum_{i \in A_k, j \in \overline{A_k}} w_{ij} \frac{1}{|A_k|} + \frac{1}{2} \sum_{i \in \overline{A_k}, j \in A_k} w_{ij} \frac{1}{|A_k|} \\ &= \frac{W(A_k, \overline{A_k})}{|A_k|} \end{aligned}$$

Также мы можем записать $h_k^T L h_k = (H^T L H)_{kk}$. Тогда

$$\mathrm{RatioCut}(A_1, \ldots, A_k) = \mathrm{tr}(H^T L H)$$

Получаем следующую оптимизационную задачу:

$$\begin{aligned} &\mathrm{tr}(H^T L H) \to \min_H \,, \\ &H^T H = I \,. \end{aligned}$$

Можно показать, что в таком случае матрица H должна состоять из первых (по возрастанию собственных значений) собственных векторов матрицы L.