Mathematics of Image and Data Analysis
Math 5467

Spectral Clustering

Instructor: Jeff Calder
Email: jcalder@umn.edu

http://www-users.math.umn.edu/~jwcalder/5467

Announcements

- Projects due Friday

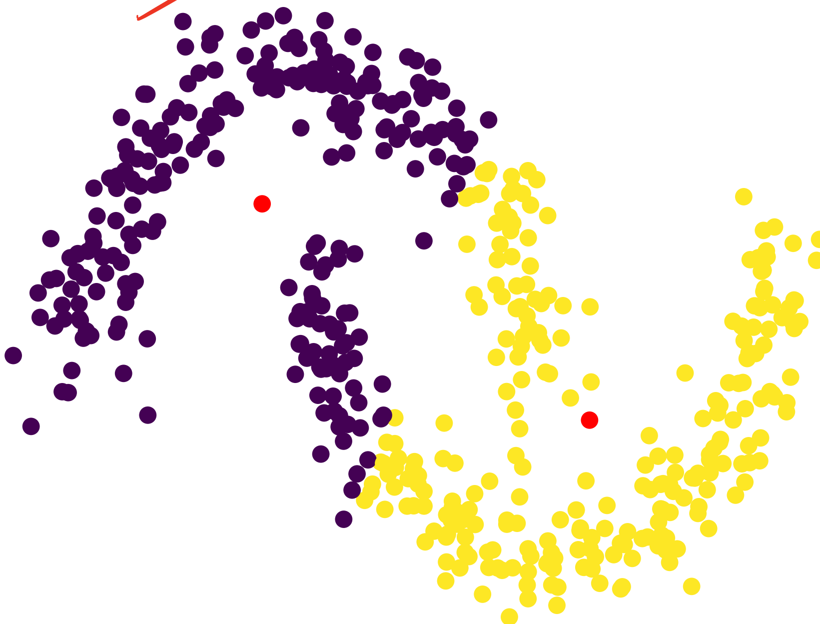
Last time

- k-means clustering

Today

- Spectral Clustering

Two-moons



- Sometimes a single point is not a good representative of a cluster, in Euclidean distance.
- Instead, we can try to cluster points so that nearby points are assigned to the same cluster, without specifying cluster centers.

Weight matrix

Let $x_1, x_2, \ldots, x_m$ be points in $\mathbb{R}^n$. To encode which points are nearby, we construct a weight matrix W, which is an $m \times m$ symmetric matrix where $W(i,j)$ represents the similarity between datapoints x_i and x_j. A common choice for the weight matrix is Gaussian weights

$$W(i,j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), \tag{1}$$

where the σ is a free parameter that controls the scale at which points are connected.

Graph cuts for binary clustering

A graph-cut approach to clustering minimizes the graph cut energy

$$E(z) = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} W(i,j) \underbrace{|z(i) - z(j)|^2}_{0 \text{ or } 1} \tag{2}$$

over label vectors $z \in \{0,1\}^m$.

Notes:

- The value $z(i) \in \{0,1\}$ indicates which cluster x_i belongs to.
- The graph-cut energy is the sum of the edge weights $W(i,j)$ that must be **cut** to separate the dataset into two clusters.

Balanced graph cuts for binary clustering

Minimizing the graph cut energy

$$E(z) = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} W(i,j)|z(i) - z(j)|^2$$

can lead to very unbalanced clusters (e.g., one cluster can have just a single point).

A more useful approach is to minimize a balanced graph cut energy

$$E_{balanced}(z) = \frac{\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} W(i,j)|z(i) - z(j)|^2}{\underbrace{\sum_{i=1}^{n} z(i)}_{\#1}\underbrace{\sum_{j=1}^{n}(1 - z(j))}_{\#0}}. \tag{3}$$

The denominator is the product of the number of points in each cluster, which is maximized when the clusters are balanced.

Balanced graph-cut problems are NP hard.

Relaxing the graph cut problem

To relax the graph-cut problem, we consider minimizing the graph cut energy

$$E(z) = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} W(i,j)|z(i) - z(j)|^2$$

over all real-vectors $z \in \mathbb{R}^m$. We still have a balancing issue (here $z = 0$ is a minimizer), so we impose the balancing constraints

$$\mathbf{1}^T z = \sum_{i=1}^{m} z_i = 0 \quad \text{and} \quad \|z\|^2 = \sum_{i=1}^{m} z(i)^2 = 1.$$

Definition 1. The *binary spectral clustering problem* is

$$\text{Minimize } E(z) \text{ over } z \in \mathbb{R}^m, \text{ subject to } \mathbf{1}^T z = 0 \text{ and } \|z\|^2 = 1.$$

The resulting clusters are $C_1 = \{x_i : z(i) > 0\}$ and $C_2 = \{x_i : z(i) \leq 0\}$.

The graph Laplacian and Fiedler vector

Let W be a symmetric $m \times m$ matrix with nonnegative entries.

Definition 2. The *graph Laplacian* matrix L is the $m \times m$ matrix

$$L = D - W \tag{4}$$

where D is the diagonal matrix with diagonal entries

$$D(i,i) = \sum_{j=1}^{m} W(i,j).$$

Lemma 3. *Then the graph cut energy can be expressed as*

$$E(z) = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} W(i,j)|z(i) - z(j)|^2 = z^T L z,$$

where L is the graph Laplacian.

Proof: $W^T = W$, $W(i,j) = W(j,i)$

$$E(z) = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} w(i,j)\left(z(i)^2 - 2z(i)z(j) + z(j)^2\right)$$

$$= \frac{1}{2}\sum_{i=1}^{m}\underbrace{\sum_{j=1}^{m} w(i,j)}_{D(i,i)} z(i)^2 + \frac{1}{2}\sum_{j=1}^{m}\overbrace{\sum_{i=1}^{m} w(i,j)}^{D(j,j)} z(j)^2$$

$$- \sum_{i=1}^{m}\underbrace{\sum_{j=1}^{m} w(i,j) z(j)}_{[Wz]_i} z(i)$$

$$= \frac{1}{2}\underbrace{\sum_{i=1}^{m} D(i,i) z(i)^2}_{z^T D z} + \frac{1}{2}\sum_{j=1}^{m} D(j,j) z(j)^2$$

$$[Dz]_i = D(i,i) z(i)$$

$$- \underbrace{\sum_{i=1}^{m} [Wz]_i z(i)}_{z^T W z}$$

$$= z^T D z - z^T W z$$

$$= z^T (Dz - Wz)$$

$$= z^T (D - W) z = z^T L z. \qquad \blacksquare$$

Properties of the graph Laplacian

Lemma 4. *Let $L = D - W$ be the graph Laplacian corresponding to a symmetric matrix W with nonnegative entries. The following properties hold.*

(i) L is symmetric. $L^T = D^T - W^T = D - W = L$.

(ii) L is positive semi-definite (i.e., $\underline{z^T L z \geq 0}$ for all $z \in \mathbb{R}^m$).

(iii) All eigenvalues of L are nonnegative, and the constant vector $z = \mathbf{1}$ is an eigenvector of L with eigenvalue $\lambda = 0$.

If $Lz = \lambda z$, $z^T L z = z^T \lambda z = \lambda \|z\|^2 \geq 0$

(iii) $z = 1 = (1, 1, \ldots, 1)$, Claim $Lz = 0 = 0 \cdot z$

$$[Lz]_i = [Dz]_i - [Wz]_i$$

$$= D(i,i)\, z(i) - \sum_{j=1}^{m} W(i,j)\, z(j)$$

$$= \sum_{j=1}^{m} w(i,j)$$

$$= \sum_{j=1}^{m} w(i,j)\, z(i) - \sum_{j=1}^{m} w(i,j)\, z(j)$$

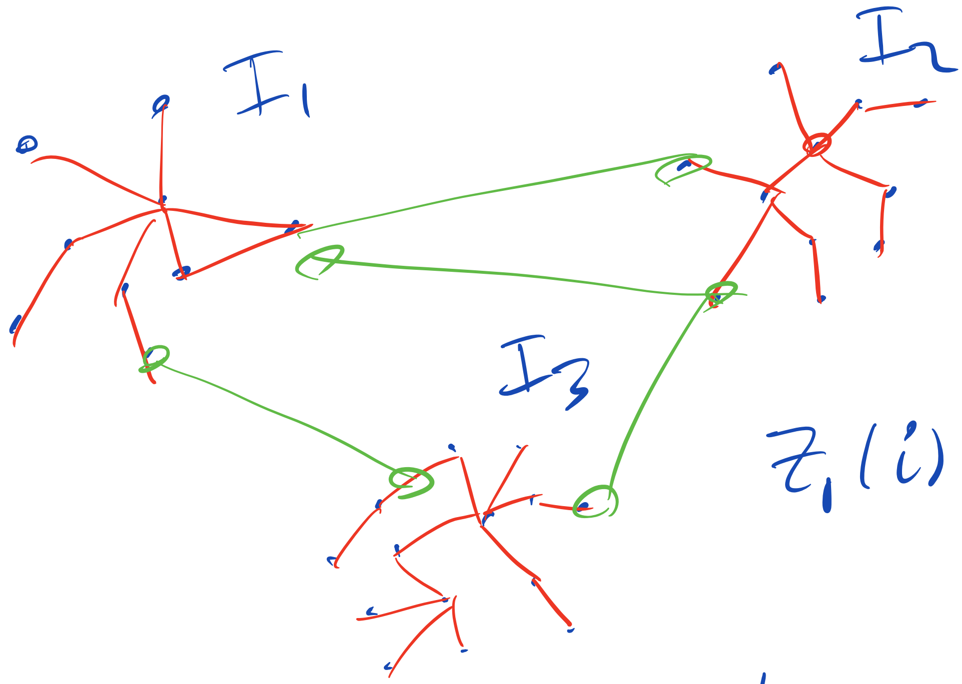
$$[Lz]_i = \sum_{j=1}^{m} w(i,j)\big(z(i) - z(j)\big)$$

Since $z(i) = z(j) = 1 \quad \forall i,j$

$$Lz = 0.$$

$\blacksquare$

Q: Is $z = 1$ the only vector in $\mathrm{Ker}(L)$?



$w(i,j) > 0$
edges

$$z_1(i) = \begin{cases} 1, & i \in I_1 \\ 0, & i \notin I_1 \end{cases}$$

$$L z_1 = 0$$

$$[Lz]_i = \sum_{j=1}^{m} w(i,j)\big(z(i) - z(j)\big)$$

$$z_j(i) = \begin{cases} 1, & i \in I_j \\ 0, & i \notin I_j \end{cases}$$

Then $\quad Lz_j = 0 \quad$ for $\quad j = 1, 2, 3.$

Fiedler vector

Let $v_1, v_2, \ldots, v_m$ be the eigenvectors of the graph Laplacian, with corresponding eigenvalues

$$0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_m.$$

Definition 5. The second eigenvector v_2 of the graph Laplacian L is called the *Fiedler vector.*

Theorem 6. *The Fiedler vector $z = v_2$ solves the binary spectral clustering problem*

$$\textit{Minimize } E(z) \textit{ over } z \in \mathbb{R}^m, \textit{ subject to } \mathbf{1}^T z = 0 \textit{ and } \|z\|^2 = 1.$$

Proof: Let z solve the binary spectral clustering problem. Write

$$z = \sum_{i=1}^{m} a_i v_i$$

① $1^T z = 0$, $\quad 0 = 1^T \sum_{i=1}^{m} a_i v_i$

$$= \sum_{i=1}^{m} a_i 1^T v_i$$

$$= a_1 1^T v_1$$

$$= a_1 \cdot \frac{1}{\sqrt{m}}$$

$$\longrightarrow a_1 = 0.$$

$v_1 = \frac{1}{\sqrt{m}}$

$\|v_1\| = 0$

$v_1^T v_j = 0$

for $j = 2, \cdots, m$.

② $\|z\|^2 = 1$, $\quad \|z\|^2 = \sum_{i=2}^{m} a_i^2 = 1$

③ $$E(z) = z^T L z = z^T L \sum_{i=2}^{m} a_i v_i$$

$$= z^T \sum_{i=2}^{m} a_i L v_i$$

$$= z^T \sum_{i=2}^{m} a_i \lambda_i v_i$$

$$z = \sum_{i=2}^{m} a_i v_i$$

$$= \sum_{i=2}^{m} a_i \lambda_i z^T v_i$$

$$z^T v_j = a_j \underbrace{v_j^T v_j}_{=1}$$

$$= \sum_{i=2}^{m} a_i^2 \lambda_i$$

So $$E(z) = \sum_{i=2}^{m} a_i^2 \lambda_i$$

We want to minimize this subject to $\sum_{i=2}^{m} a_i^2 = 1$

$[\lambda_1 \le \lambda_2 \le \lambda_3 \le \cdots \le \lambda_m]$

($\star$) <u>Claim:</u> $a_2 = 1,\ a_3 = a_4 = \cdots = a_m = 0$

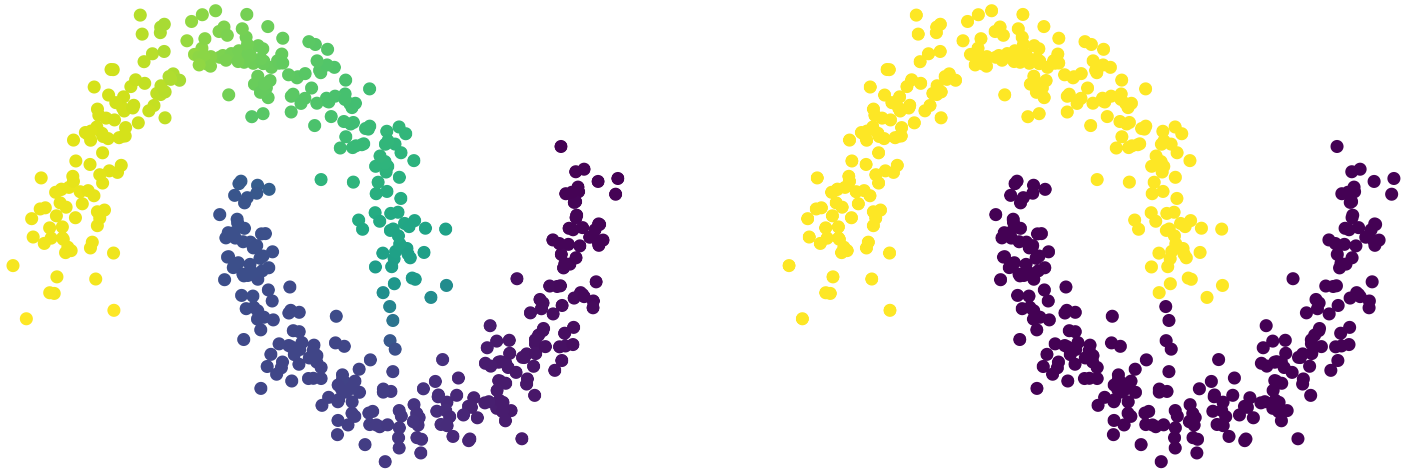
<u>To see this:</u>

$$E(z) = \sum_{i=2}^{m} a_i^2 \lambda_i \geq \sum_{i=2}^{m} a_i^2 \lambda_2$$

$$= \lambda_2 \sum_{i=2}^{m} a_i^2 = \lambda_2$$

$(*) \Longrightarrow \quad E(z) = \lambda_2 \qquad \blacksquare$

Example



(a) Fiedler vector (b) Spectral Clustering

Figure 1: (a) The Fiedler vector and (b) spectral clustering on the 2-moons dataset.

k-nearest neighbor graph

The Gaussian weights

$$W(i,j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right),$$

are not always useful in practice, since the matrix W is dense (all entries are non-zero), and the connectivity length σ is the same across the whole graph.

It is more common to use a k-nearest neighbor graph. Let $d_{k,i}$ denote the Euclidean distance between x_i and its k^{th} nearest Euclidean neighboring point from $x_1, \ldots, x_m$. A k-nearest neighbor graph uses the weights

$$W(i,j) = \begin{cases} 1, & \text{if } \|x_i - x_j\| \leq \max\{d_{k,i}, d_{k,j}\} \\ 0, & \text{otherwise.} \end{cases}$$

The weights need not be binary, and can depend on $\|x_i - x_j\|$, similar to the Gaussian weights. The k-nearest neighbor graph weight matrix W is very sparse (most entries are zero), so it can be stored and computed with efficiently.

Spectral clustering in Python (.ipynb)