Mathematics of Image and Data Analysis
Math 5467

Lecture 7: Spectral Clustering

Instructor: Jeff Calder
Email: jcalder@umn.edu

http://www-users.math.umn.edu/~jwcalder/5467S21

Announcements

- Projects due Friday
- HW 1 Graded, in canvas soon.

Last time

(*) Out of 12 points.

Grade-point scale

$12 = A$

$11 = A-$

$10 = B+$

$\vdots$

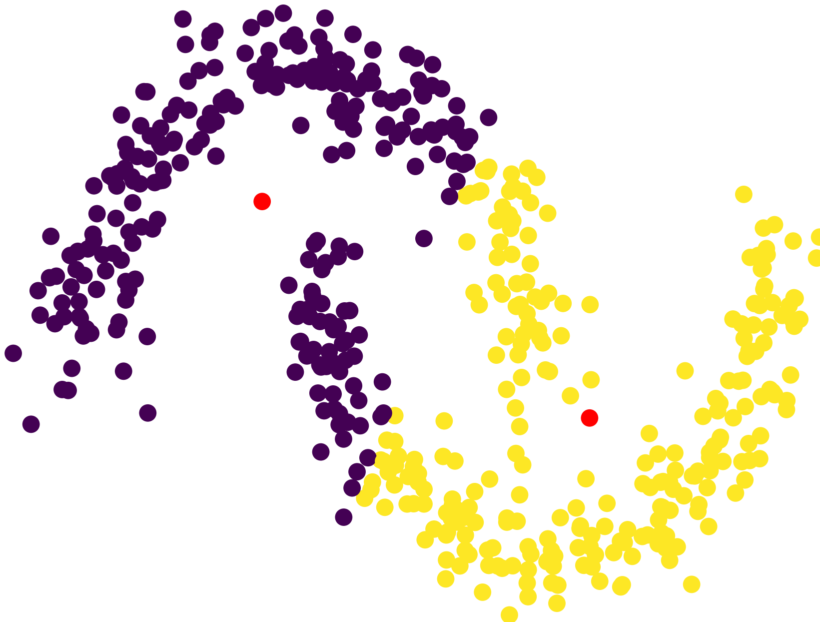
- k-means clustering

Today

- Wrap up k-means
- Spectral Clustering

(*) HW 1 Solutions online

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

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

$\exp(x) = e^x$

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

Connects x_i, x_j if $\|x_i - x_j\| \approx \sigma$

Graph cuts for binary clustering

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

(2)
$$E(z) = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} W(i,j) \underbrace{|z(i) - z(j)|^2}_{= \begin{cases} 1, & \text{if } z(i) \neq z(j) \\ 0, & \text{if } z(i) = z(j) \end{cases}}$$

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

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

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

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

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

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

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

$D(j,j)$

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

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

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

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

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

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

Since $L = D - W$. $\blacksquare$

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.

(ii) L is positive semi-definite (i.e., $z^T Lz \geq 0$ 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 $\lambda = 0$.

(i) follows from symmetry of W.

(ii) Follows from Lemma

$$z^T L z = E(z) \geq 0$$

(iii) L is symmetric, $\exists$ eigenvectors

$v_1, v_2, \dots, v_m,$ and eigenvalues

$$\lambda_1 \le \lambda_2 \le \dots \le \lambda_m$$

$$0 \le v_i^T L v_i = v_i^T \lambda_i v_i = \lambda_i \underbrace{\|v_i\|^2}_{=1}$$

$\Rightarrow \quad \lambda_i \ge 0$ for all i.

Check that $E(\mathbb{1}) = 0$

So $\quad \mathbb{1}^T L \mathbb{1} = 0.$

Check $L\mathbb{1} = D\mathbb{1} - W\mathbb{1}$

$$(L\mathbb{1})(i) = D(i,i) - \underbrace{\sum_{j=1}^{n} W(i,j)\mathbb{1}}_{D(i,i)}$$

$$= 0.$$

$$\Rightarrow \quad v_1 = \frac{1}{\sqrt{m}}, \quad \lambda_1 = 0.$$

$\blacksquare$

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: Write a minimizer z as

$$z = \sum_{i=1}^{m} a_i v_i \quad \text{for some } a_i$$

(1) $1 = \|z\|^2 = \sum_{i=1}^{m} a_i^2$ since v_i are orthonormal.

(2) Since $v_1 = \frac{1}{\sqrt{m}}$ we have

$$0 = 1^T z = \sqrt{m}\, v_1^T \sum_{i=1}^{m} a_i v_i$$

$$= \sqrt{m} \sum_{i=1}^{m} a_i v_1^T v_i$$

$$= \sqrt{m}\, a_1 \|v_1\|^2 = \sqrt{m}\, a_1$$

$$\Rightarrow \boxed{a_1 = 0.}$$

Now write

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

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

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

$$= \sum_{i=1}^{m} a_i \lambda_i \underbrace{z^T v_i}_{= a_i}$$

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

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

is optimal.

<u>To see this:</u> $E(z) \geq \lambda_2 \sum_{i=2}^{m} a_i^2 = \lambda_2$

Since $\sum_{i=2}^{m} a_i^2 = 1$

and if we set $a_2 = 1$, $a_3 = a_4 = \cdots = a_m = 0$

$$E(z) = a_2^2 \lambda_2 = \lambda_2 .$$

$\square$

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)