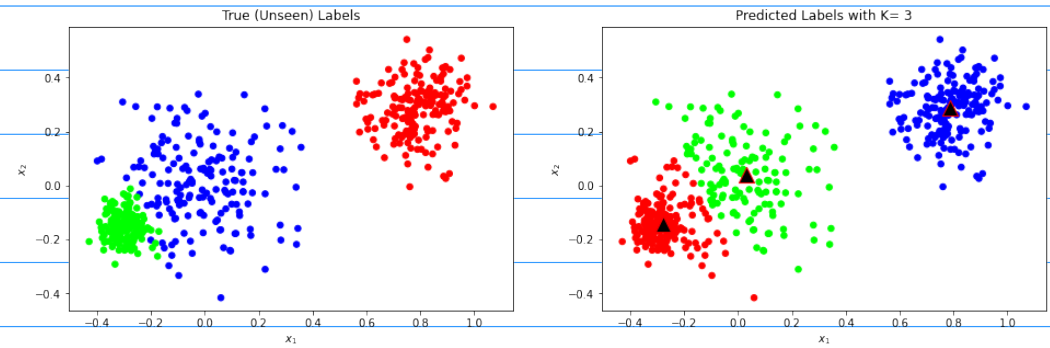
Lecture 19: Introduction to Graph Laplacians

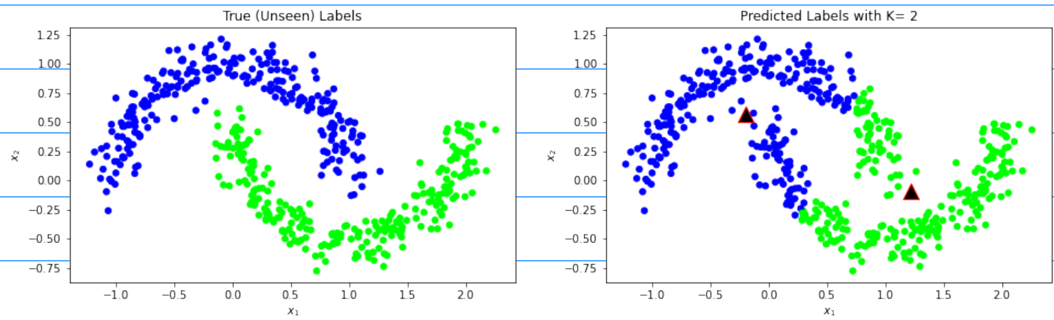
We saw K-means clustering in the last lecture. A simple & effective method for unsupervised learning based on "distances" of points from each other.

But we also saw that K-means has a major drawback!

Good!



Bad!



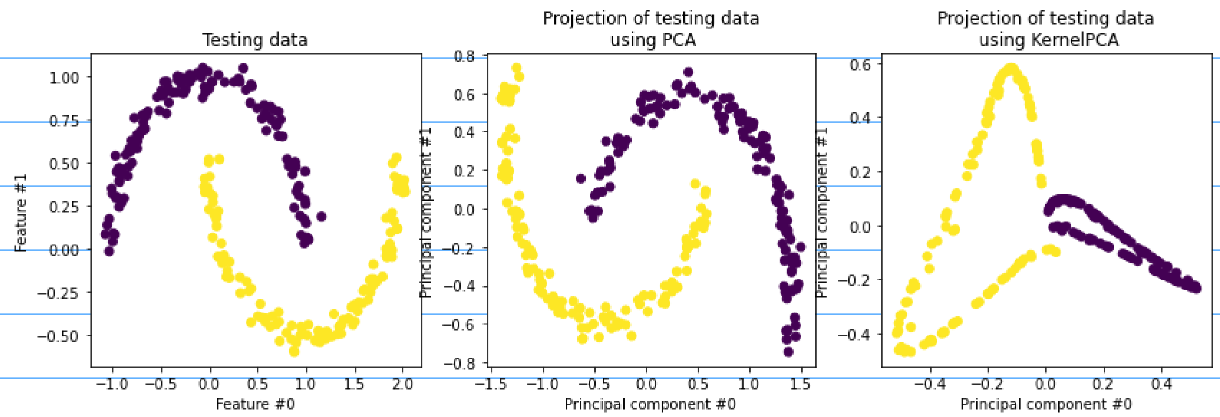
K-means does not see low-dimensional structure in the data!

Equipped with our intuition from Kernel methods & in particular Kernel PCA we now set out to improve K-means.

We will consider a recent technique called spectral clustering which is based on the following goal:

Let $X = \{\underline{x}_0, \ldots, \underline{x}_{N-1}\} \in \mathbb{R}^d$ be our data set. Can we find a feature map $F: \mathbb{R}^d \longrightarrow \mathbb{R}^m$ so that applying K-means on $F(X)$ rather than X can lead to better clustering?

Recall that we saw something akin to this with kernel PCA:



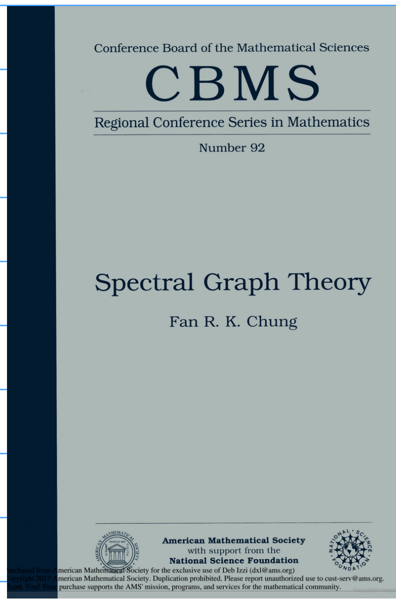
So the problem reduces to finding the nonlinear mapping F. And spectral clustering finds this mapping as the feature map of an implicit kernel that is never computed explicitly!

Note: Since spectral clustering is a relatively new subject most of the literature around it is in the form of papers, tutorials.

Stat Comput (2007) 17: 395–416
DOI 10.1007/s11222-007-9033-z

A tutorial on spectral clustering

Ulrike von Luxburg



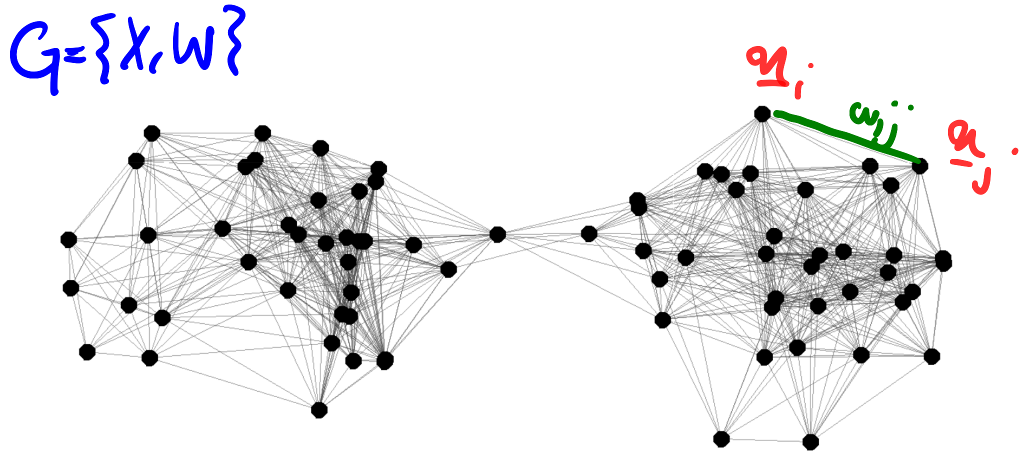
19.1 Similarity Graphs

Consider our data set $X = \{\underline{x}_0, \ldots, \underline{x}_{N-1}\} \in \mathbb{R}^d$ & a symmetric matrix $W \in \mathbb{R}^{N \times N}$ with non-negative entries $w_{ij} \geq 0$.

We then define a (weighted undirected) graph $G = \{X, W\}$ where the $\underline{x}_j \in \mathbb{R}^d$ are the vertices of G & the entries w_{ij} of W denote weights that are associated to edges that connect $\underline{x}_i$ to $\underline{x}_j$.

- If $w_{ij} = 0$ for some i, j then $\underline{x}_i$ & $\underline{x}_j$ are not connected by an edge.
- Since W is symmetric, then $w_{ij} = w_{ji}$, ie, the edges have no direction.

- We generally think of w_{ij} as the strength of the connection between $\underline{x}_i$ & $\underline{x}_j$. ie, larger weights mean stronger connection.



We will consider a particular family of weighted graphs called proximity graphs

Let $\eta : [0, +\infty) \to [0, +\infty)$ be a non-negative, non-increasing & continuous function at zero (we call it the weight function)

we then take $w_{ij} = \eta(\|\underline{x}_i - \underline{x}_j\|_p)$
for $1 \leq p \leq +\infty$.

A typical example is simply to choose

$$\eta = \exp\left(-\frac{t^2}{2\sigma^2}\right) \quad \& \quad p=2$$

which leads to

$$w_{ij} = \exp\left(-\frac{\|\underline{x}_i - \underline{x}_j\|_2^2}{2\sigma^2}\right)$$

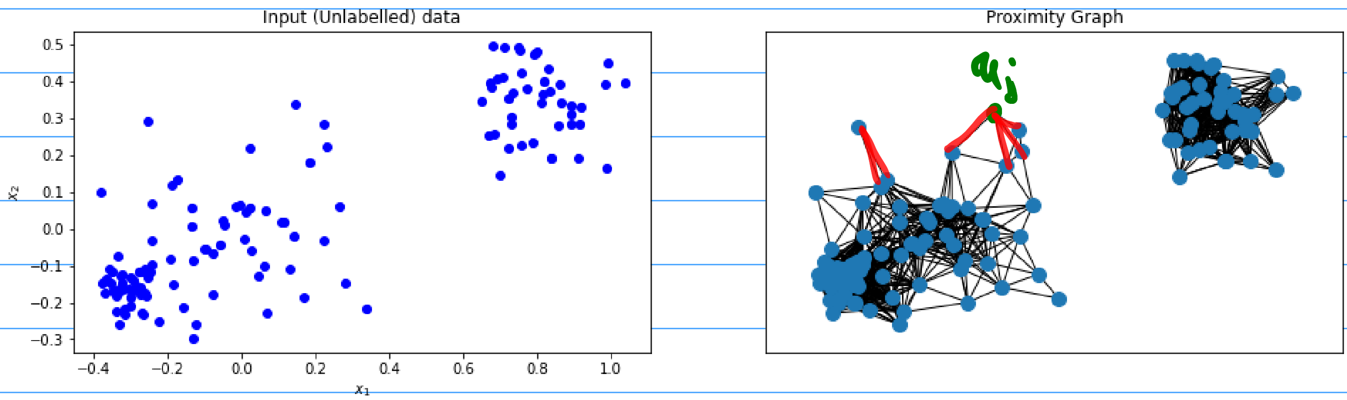
another popular choice for η is,

$$\eta(t) := \begin{cases} 0 & \text{if } t \geq r \\ 1 & \text{if } t < r, \end{cases}$$

which leads to,

$$w_{ij} = \begin{cases} 0 & \text{if } \|\underline{x}_i - \underline{x}_j\|_2 \geq r \\ 1 & \text{if } \|\underline{x}_i - \underline{x}_j\|_2 < r, \end{cases}$$

we call such graphs "proximity" or "similarity" graphs since the weights w_{ij} encode the similarity/closeness of the vertices.



With the matrix W at hand we now define the graph Laplacian matrix of G:

- Define the degree vector
  $\underline{d} \in \mathbb{R}^N$, $\quad d_j = \sum_{i=0}^{N-1} W_{ji}$ ← sum of rows of W
- Define the diagonal degree matrix
  $$D = \mathrm{diag}(\underline{d}) = \begin{bmatrix} d_0 & & 0 \\ & d_1 & \\ 0 & & \ddots & d_{N-1} \end{bmatrix}$$
- Define the (Unnormalized) graph Laplacian
  $$\tilde{L} = D - W$$
- As well as the normalized graph Laplacian
  $$L = D^{-1/2}(D-W)D^{-1/2} = I - D^{-1/2} W D^{-1/2}$$

19.2 Properties of Graph Laplacians

The graph Laplacian matrices $L, \tilde{L}$ introduced above have a lot of useful properties & have applications in many areas (see refs).

We will account a small number of these properties as they are very useful in the context of clustering.

- Both L & $\tilde{L}$ are non-negative definite & symm. (NDS). Thus, they have real eigenvalues & eigenvectors $\{\lambda_j, \underline{u}_j\}_{j=0}^{N-1}$
  $0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{N-1}$
- Both L & $\tilde{L}$ have at least one zero eigenvalue,

$$\tilde{L}\mathbb{1} = (D - W)\mathbb{1} = \underline{d} - \underline{d} = 0$$

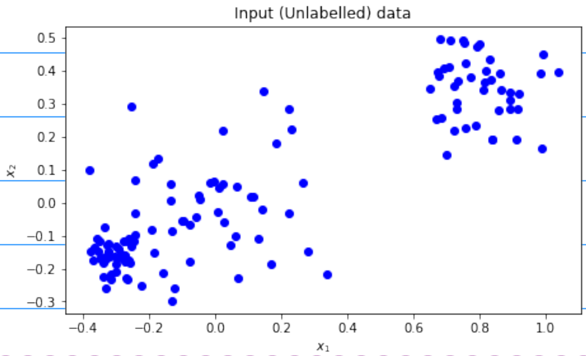
$$L(D^{1/2}\mathbb{1}) = D^{-1/2}\tilde{L}D^{-1/2}D^{1/2}\mathbb{1} = D^{-1/2}\tilde{L}\mathbb{1} = 0$$

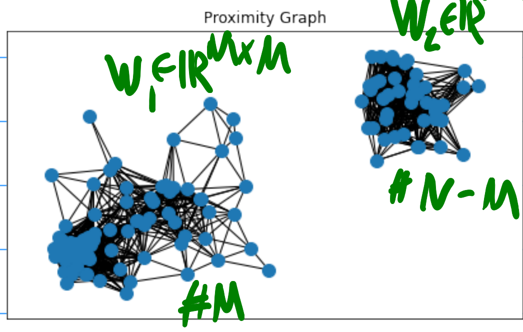
- If $\underline{u}$ is an eigenvector of $\tilde{L}$ then $D^{1/2}\underline{u}$ solve the generalized eigenvalue problem $L(D^{1/2}\underline{u}) = \lambda D^{-1}(D^{1/2}\underline{u})$

(Most important for us)

- If the graph G has K-disconnected components (ie, subgraphs that are not connected by any edges) then,

$$0=\lambda_0=\lambda_1=\dots=\lambda_{K-1}<\lambda_K\leq\lambda_{K+1}\leq\dots$$





$W_2\in\mathbb{R}^{(N-M)\times(N-M)}$

$W_1\in\mathbb{R}^{M\times M}$

$\# N-M$

$\# M$

In other words, # of zero eigenvalues of $L, \tilde{L}$ correspond to "clusters" in the graph!

$$\tilde{L}_1 = D_1 - W_1, \quad \tilde{L}_1 \mathbb{1} = \underline{d}_1 - \underline{d}_1 = 0$$

$$\tilde{L}_2 = D_2 - W_2, \quad \tilde{L}_2 \mathbb{1} = 0$$

$$W = \left[\begin{array}{c|c} W_1 & \emptyset \\ \hline \emptyset & W_2 \end{array}\right], \quad \tilde{L} = \left[\begin{array}{c|c} \tilde{L}_1 & \emptyset \\ \hline \emptyset & \tilde{L}_2 \end{array}\right]$$

$$\overbrace{\left[\begin{array}{c|c} \tilde{C}_1 & 0 \\ \hline 0 & \tilde{C}_2 \end{array}\right]}^{\tilde{C}} \left[\begin{array}{c} 1 \\ \hline 0 \\ 0 \\ \vdots \end{array}\right] = 0$$

$$\overbrace{\left[\begin{array}{c|c} \tilde{C}_1 & 0 \\ \hline 0 & \tilde{C}_2 \end{array}\right]}^{\tilde{C}} \left[\begin{array}{c} 0 \\ \hline 1 \end{array}\right] = 0$$