Spectral Clustering

Bruce Campbell

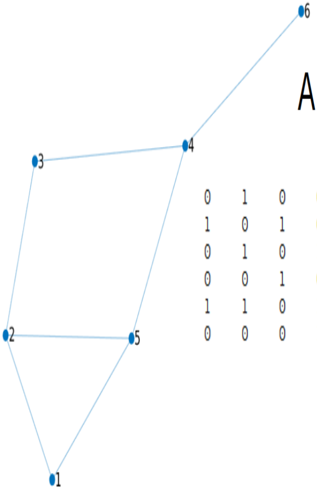
OSU Math 5603 Final Project

2022

Definitions

- X data
- $G = (E, V)$ Graph
- A similarity matrix/ graph adjacency matrix
- D degree matrix
- L graph Laplacian L = D-A
- $\sigma(L)$ the eigenspectrum - here increasing in values

Example Graph



$$A = \begin{array}{cccccc} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{array}$$

$$D = \begin{array}{cccccc} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}$$

$$L = \begin{array}{ccccc} 2 & -1 & 0 & 0 & -1 \\ -1 & 3 & -1 & 0 & -1 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 3 & -1 \\ -1 & -1 & 0 & -1 & 3 \\ 0 & 0 & 0 & -1 & 0 \end{array}$$

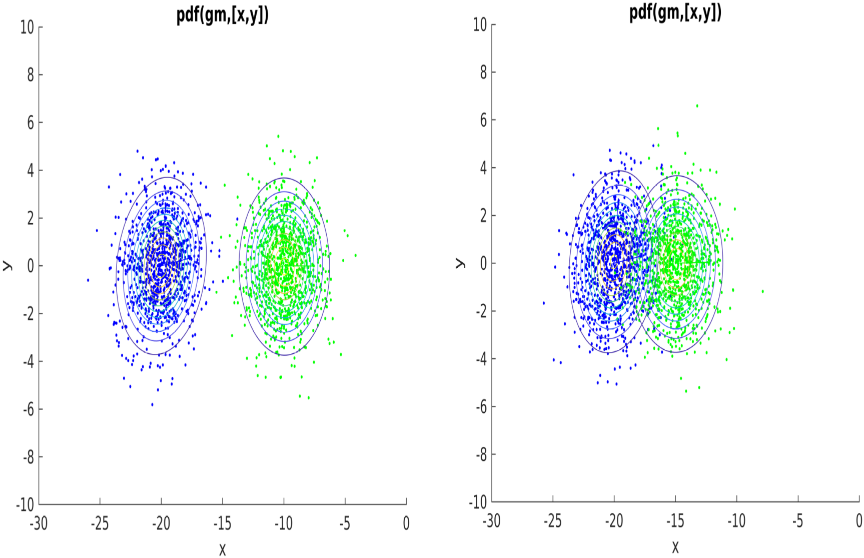
Spectral Clustering Steps

- Calculate similarity matrix A
- Calculate Laplacian L
- Calculate first k eigenvectors U_k
- Using $U_k(i,j)$ as embedded feature values, cluster in $\mathbb{R}^k$

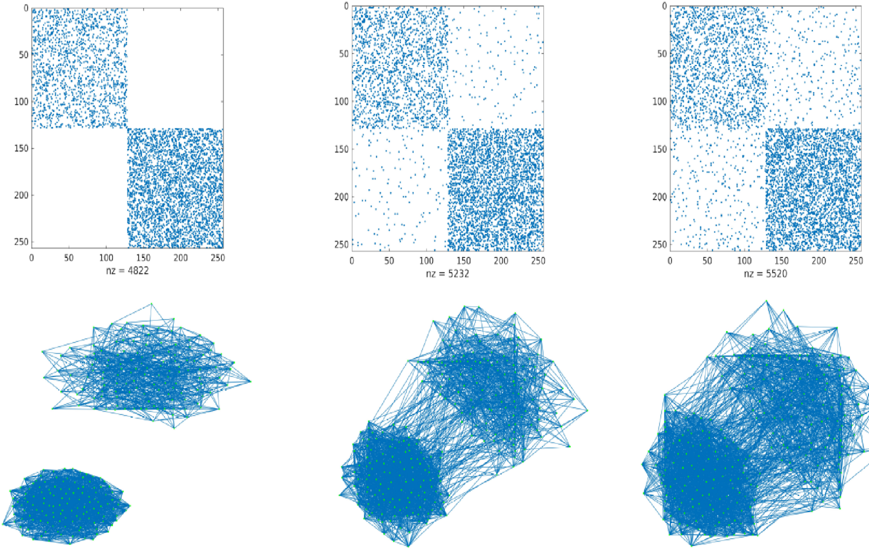
Variations

- A can be weighted
- Neighbors can be all to nearest k
- L can be normalized $\mathcal{L} = D^{\frac{1}{2}} L D^{\frac{-1}{2}}$

X with increasing $p(Class2\ Class1)$



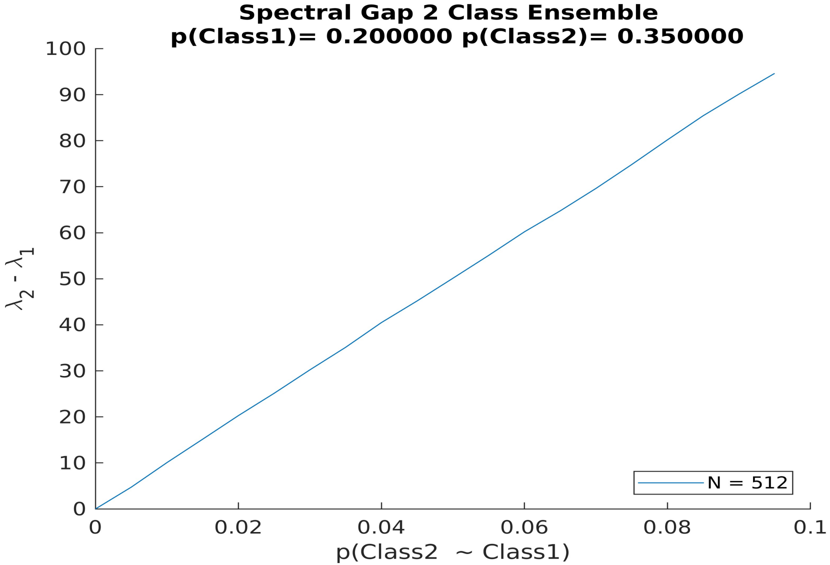
A, G with increasing $p(Class2\ Class1)$



The Spectral Gap

- Cheeger's Inequality - relationship between second eigenvalue and the size of the smallest cut of the graph that separates the classes
- We normalize Laplacian because of the close connection to geometry and stochastic processes.
- If we use $\mathcal{L}$ then $D^{-1}A$ is the transition matrix of a random walk on the graph
- Close relationship to heat flow problem on a manifold

Spectral Gap with increasing $p(Class2\ Class1)$

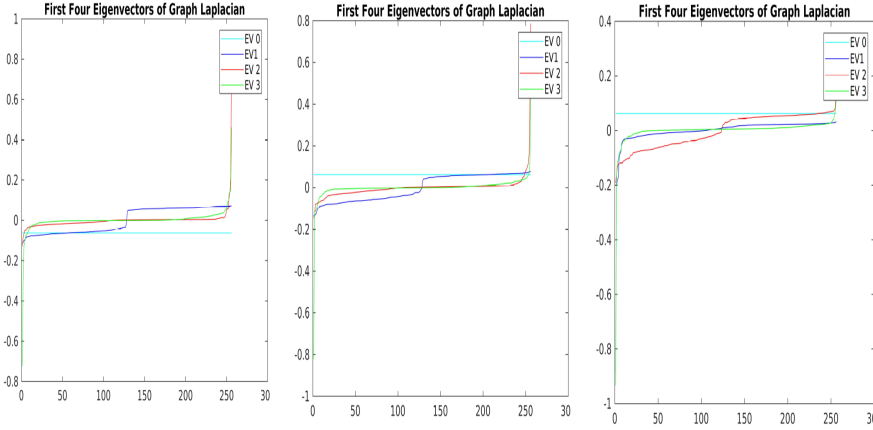


We see Cheeger's inequality $\lambda_2 \leq \phi(G)$ in action. As crosstalk increases, it becomes harder to partition the data.

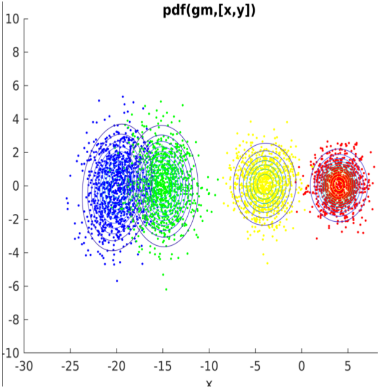
So How Do We Actually Separate The Classes?

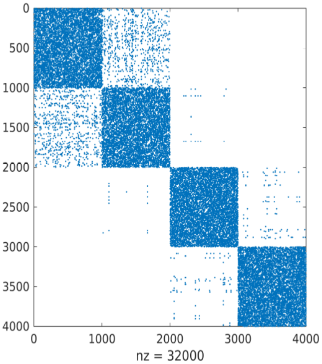
- The second eigenvector - the Fiedler Vector - of the graph Laplacian can be considered the first 'fundamental mode' of the graph.
- The spectrum is ordered from lowest to highest values
- The eigenvectors are ordered from global to local features.
- We can use the sign of the Fiedler vector to partition

Eigenvectors with increasing $p(Class2\ Class1)$

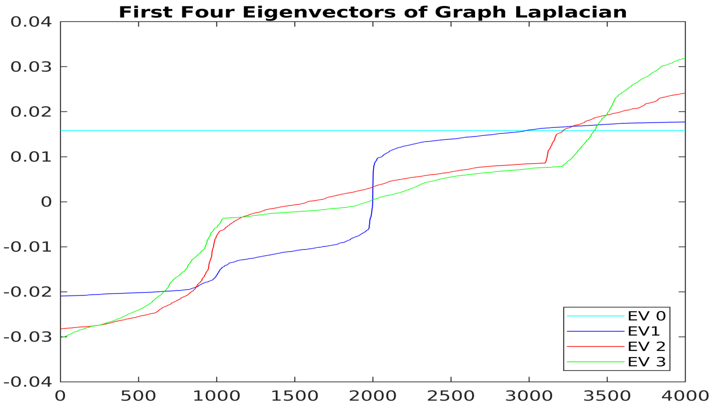


More Classes





Eigenvectors : 4 Classes



We notice the eigenvectors associated with non zero eigenvalues all have structure

Blessings and Curses of High Dimension

Blessings

- We may have redundant features - serial correlation in signals etc.
- Nice theorems about retaining structure after redacting data (Johnson Lindenstrauss)

Curses

- High dimensional data is sparse
- The structure we're interested in may be hidden
- Inference can be difficult due to correlated features and noise dimensions
- Counter-intuitive geometry of high dimensional data - most points are orthogonal etc.

Bibliography

- Spectral Graph Theory, Fan Chung, American Mathematical Soc., 1997
- On Spectral Clustering: Analysis and an algorithm, Ng, Andrew and Jordan, Michael and Weiss, Yair, Advances in Neural Information Processing Systems,2001
- Git Repository, https://github.com/campbell-2589/OSU_MATH_5603/tree/main/Project/Report