Spectral Clustering: RBF and Sigma Exploration

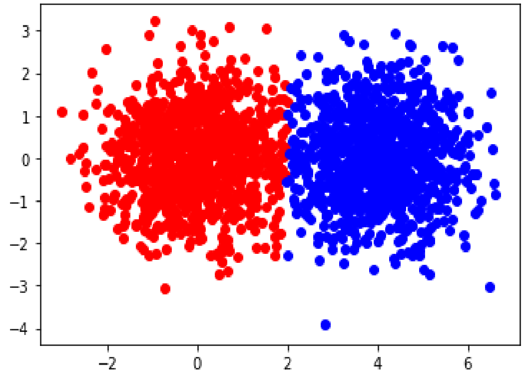
Martin Buck, Tufts University

Outline

- Background and mathematics
  - Why pursue spectral clustering?
    - Comparison with k-means
  - Spectral clustering algorithm
    - Radial Basis Function and sigma parameter
    - Eigenvector embedding and the Fiedler vector
- Effects of sigma parameter on eigenvalues and eigenvectors
  - Connection strength and sigma
    - Eigenvalue and eigenvector distribution vs. sigma
  - Examples with two-spiral data set
- Summary and discussion

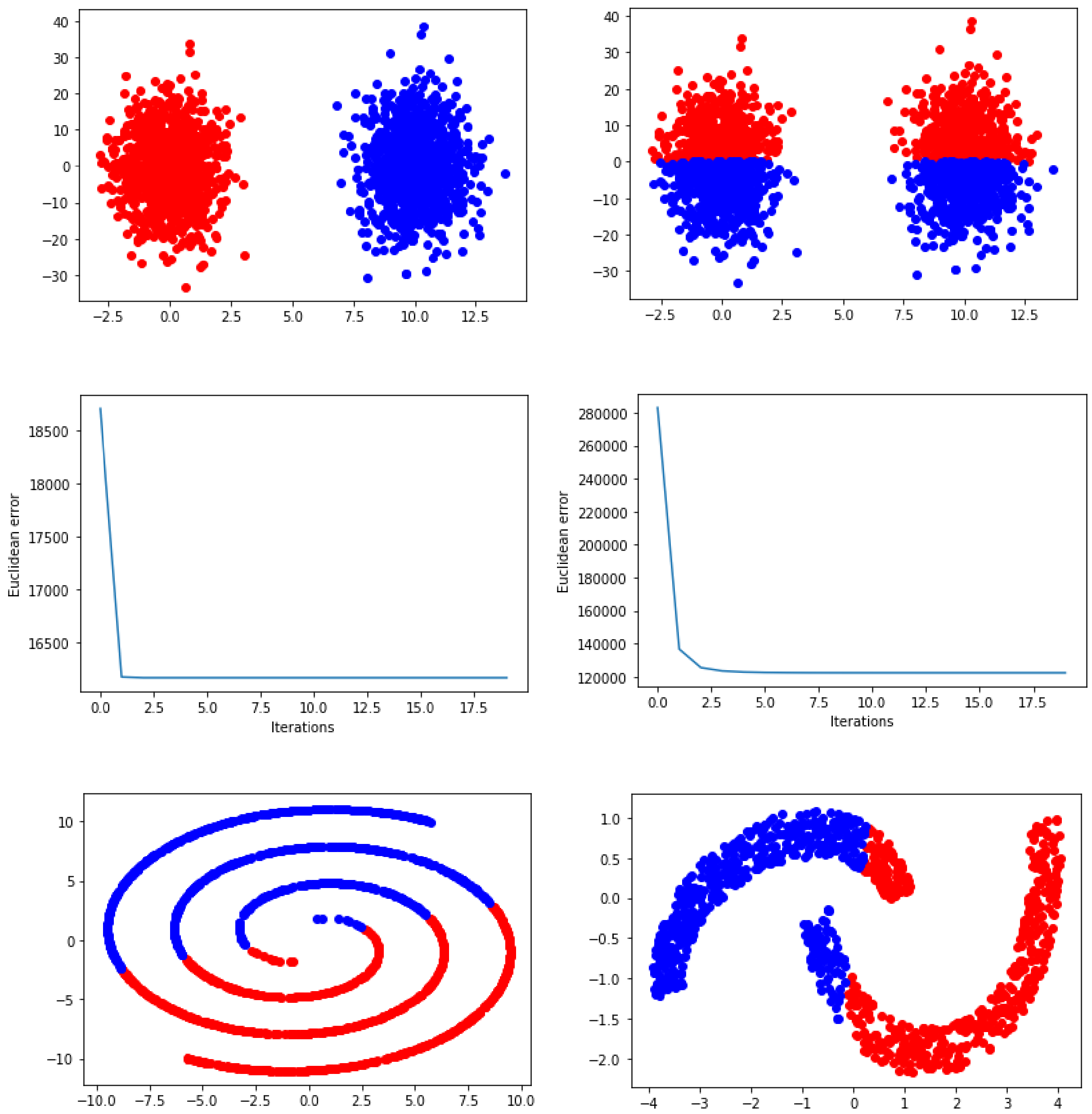
Clustering

- Want to divide a data set into k disjoint groups
  - Given $x_i \in R^d$, assign labels $y_i \in \{1, 2, \ldots, k\}$, where k is given number of clusters
  - Brute force approach is expensive!
- K-means clustering
  - Iterative center-based algorithm that seeks to minimize an objective function
    - Clusters $C_j = \{x_i \mid y_i = j\}$
    - Centers $\mu_c = \frac{1}{n}\sum x_i$
    - $F(C_j) = \sum_{j=1}^{k} \sum_{x_i \in C_j} ||x_i - \mu_j||^2$
  - Algorithm
    - Randomly assign centers μ_j
      - Estimate cluster labels for given centers
        - $y_i = \arg\min ||x_i - \mu_j||^2$
      - Estimate new centers for given cluster labels
        - $\mu_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i$



Problems with K-means

- Non-convex optimization can result in local minimum depending on where centers are initialized
  - Ellipse data has at least two solutions k-means converges to
- If data is "non-radial" then clusters may not correspond to intuition or known labels
  - Two-spirals and moon data have at least one wacky solution



Why Pursue Spectral Clustering?

- Spectral clustering can pre-process the data in a way to handle more complex data shapes
  - Run k-means on the embedded data which hopefully disentangles clusters
  - Local minimum and initial center location still an issue
- The algorithm
  - Create weight matrix W_{ij} and degree matrix $D_{ij} = \sum_{i \neq j}^{|V|} W_{ij}$
  - Subtract the two to get the Laplacian $L = D - W$
  - Find eigenvalues and eigenvectors of L
    - Know a basis of eigenvectors for R^n exist by the spectral theorem
    - L is also positive semi-definite which is important since the smallest eigenvalues play an important role in clustering
  - Choice of eigenvectors "embed" original data in a way that makes k-means converge to a nice solution
    - Multiplicity of the zero eigenvalue gives the number of connected components
    - Fiedler vector is the eigenvector corresponding to the second smallest eigenvalue and is often used to start clustering
      - Fiedler vector is the solution to a relaxed optimization problem that minimizes cuts on a connected graph

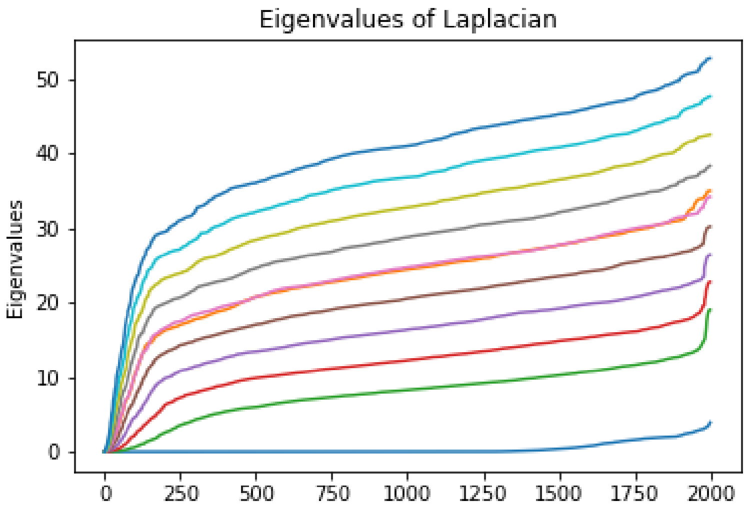
RBF and the Sigma Parameter

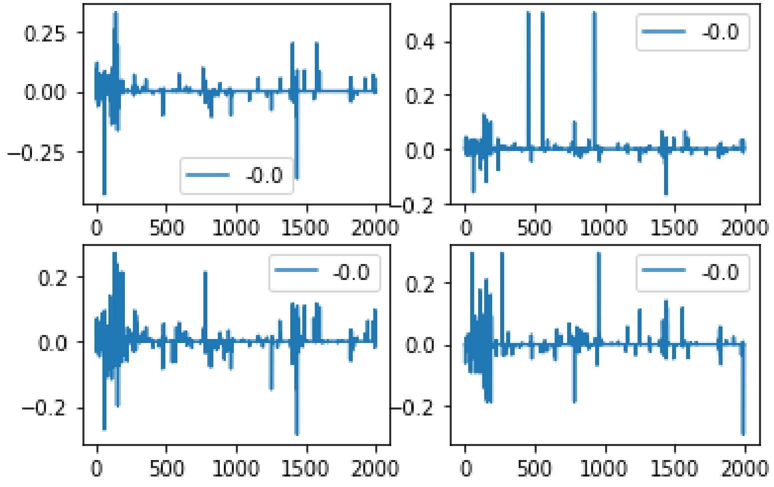
- Common weight matrix to use calculates a distance or similarity measure between each pair of points
  - $W_{ij} = e^{\frac{-||x_i - x_j||^2}{2\sigma^2}}$
  - Called the Radial Basis Function
    - Symmetric gaussian shape around a point with tuning parameter sigma
    - Bounded above by one and below by zero
- What affect does the sigma parameter have on the eigenvalues and eigenvectors of the Laplacian matrix and clustering performance?
  - As sigma increases, the distance between two points matters little. Each entry in the weight matrix approaches one and the connections on our graph are all strong and roughly equal
    - Expect clustering to perform poorly here since each point is indistinguishable
  - As sigma decreases, the distance between two points matters a lot. Each entry in the weight matrix approaches zero and the connections on our graph are all weak
    - Again, expect clustering to perform poorly here since each point is its own cluster

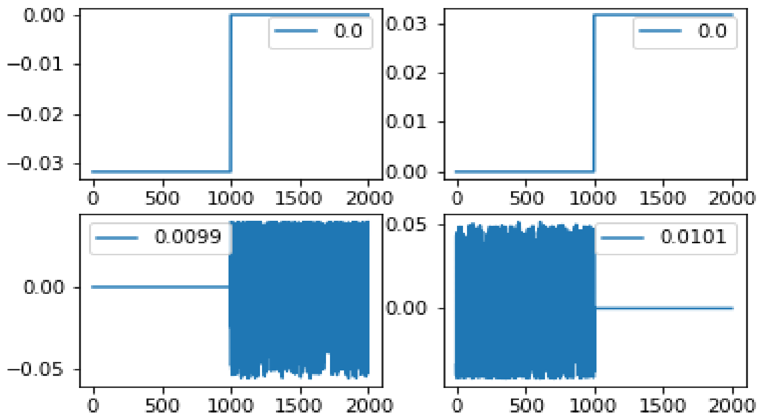
Spectral Clustering on Two-Spirals Data

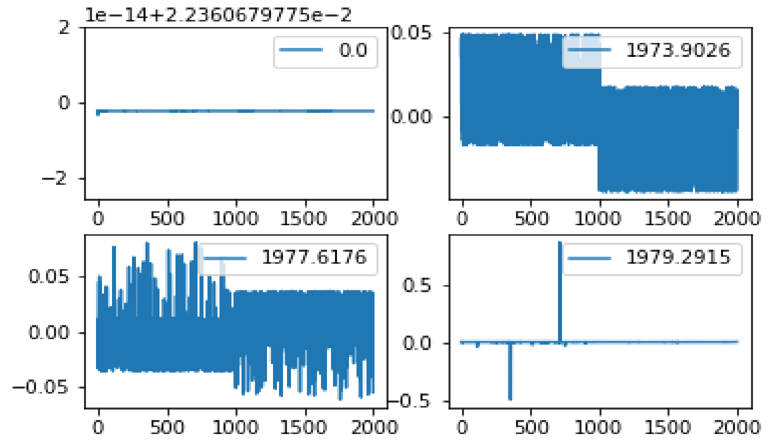
- Demonstrate that this method works where k-means alone fails but performance does depend on the sigma parameter
- Cases for varying sigma
  - For very small sigma ($\sigma = .001$), the multiplicity of the zero eigenvalue is large and we have lots of eigenvectors in its eigenspace
    - Found many "roughly" connected components since small sigma creates weak connections between all data points
  - For very large sigma ($\sigma = 100$), the multiplicity of the zero eigenvalue is one, and we have the only the constant eigenvector in its eigenspace
    - Found only one connected component since large sigma creates strong connections between all data points
  - For intermediate values of sigma ($\sigma \sim .1$), the multiplicity of the zero eigenvalue is two, and we have two eigenvectors in its eigenspace
    - Found two "roughly" connected components. Just right!

Spectral Clustering on Two-Spirals Data



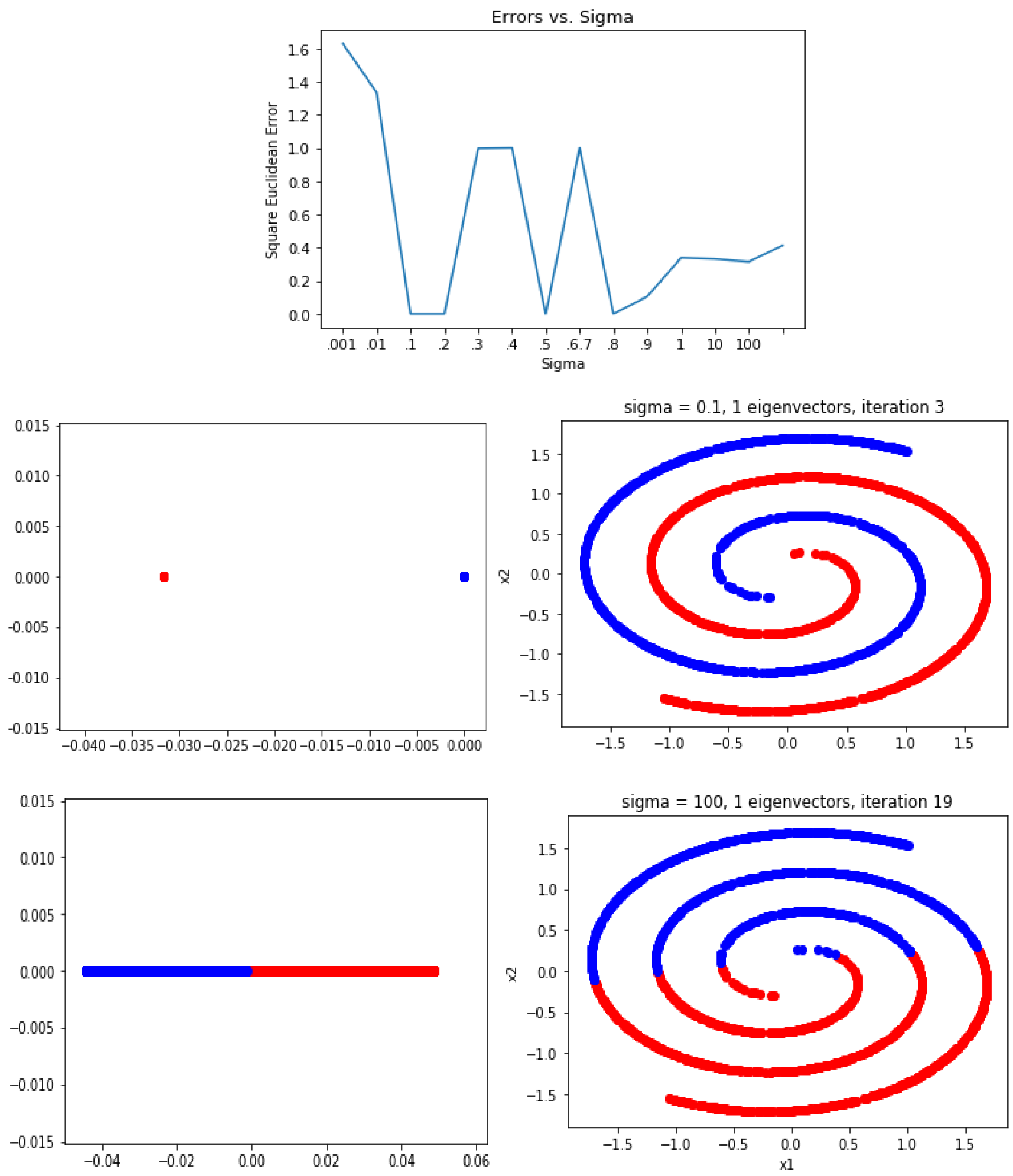






Spectral Clustering on Two-Spirals Data

- How does the previous slide translate into clustering performance?
  - For intermediate values of sigma, we have great clustering results using either eigenvector for $\lambda = 0$
    - No need to use Fiedler vector since we have two “roughly” connected components and two eigenvectors
  - For small values of sigma, we have poor clustering results
    - Difficult to even identify the multiplicity of $\lambda = 0$ or find the Fiedler vector
  - For large values of sigma, also have poor clustering results
    - Can see on previous slide there is a signal in the Fiedler vector that differentiates the two clusters but post-processing needed



Summary, Further Directions, Drawbacks

- We can use the eigenvectors of the Laplacian matrix to embed or transform data in a way that makes life easier for k-means
- For certain values of sigma the graph "disconnects" in a way that highlights the clusters
  - This is reflected in the multiplicity of the zero eigenvalue
- The value of sigma can seriously change the eigenvalue distribution and the values of the corresponding eigenvectors
  - Picking which eigenvectors to use is challenging for non-optimal sigma
    - Is there a choice of eigenvectors that will still separate the two spirals?
- How does this generalize to other data sets?
  - More complex data shapes and clusters
    - Other weight matrices may be more appropriate

References

- Friedman, J., Hastie, T. and Tibshirani, R., 2001. The elements of statistical learning (Vol. 1, No. 10). New York: Springer series in statistics.
- James, G., Witten, D., Hastie, T. and Tibshirani, R., 2013. An introduction to statistical learning (Vol. 112, p. 18). New York: springer.
- Russell, S. and Norvig, P., 2002. Artificial intelligence: a modern approach.