Lecture 7: Spectral Clustering; Linear Dimensionality Reduction via Principal Component Analysis

Lester Mackey

April 21, 2014

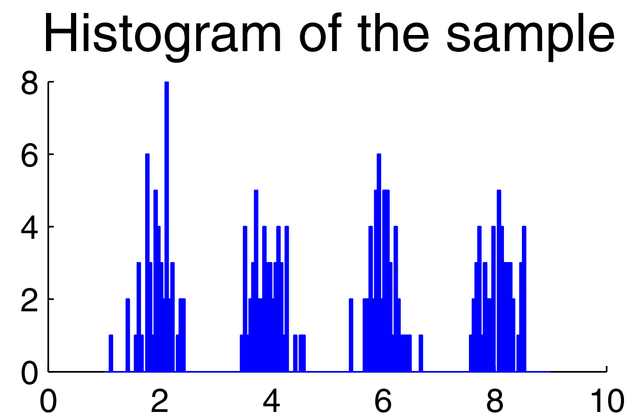
Blackboard discussion

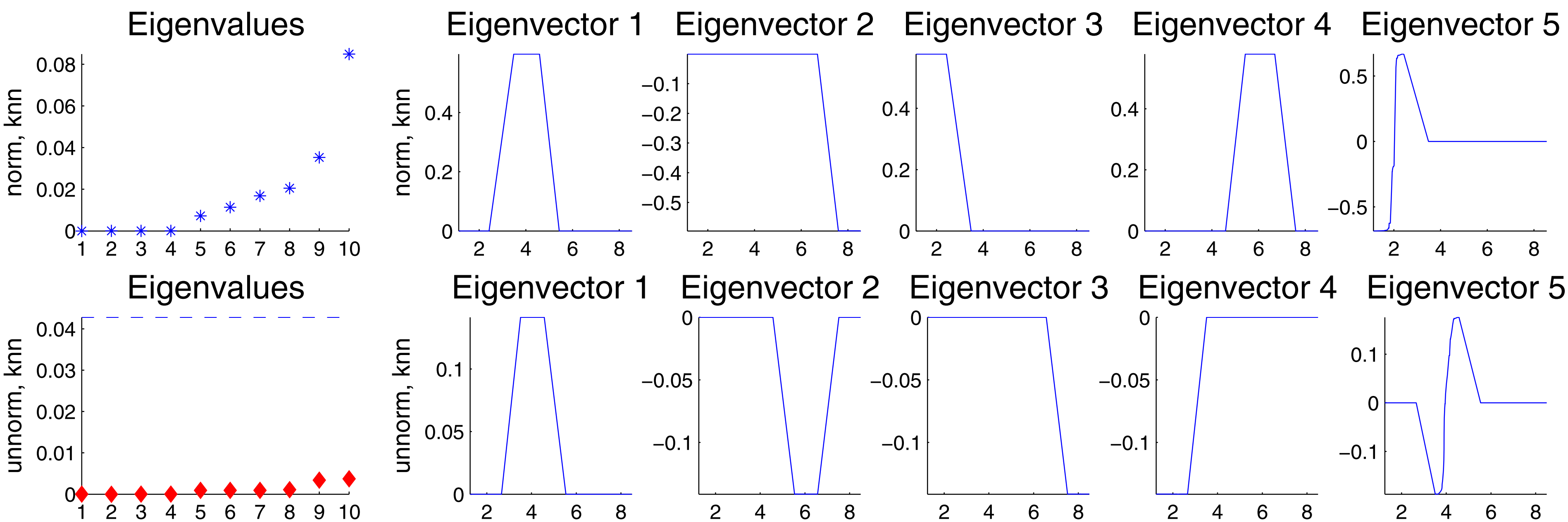
- See lecture notes

Spectral clustering example: GMM

von Luxburg, 2007

- Data generated from a mixture of 4 Gaussians in 1D
- W from 10-nearest neighbors
- Top row: normalized L_{rw}
- Bottom row: unnormalized L_{un}

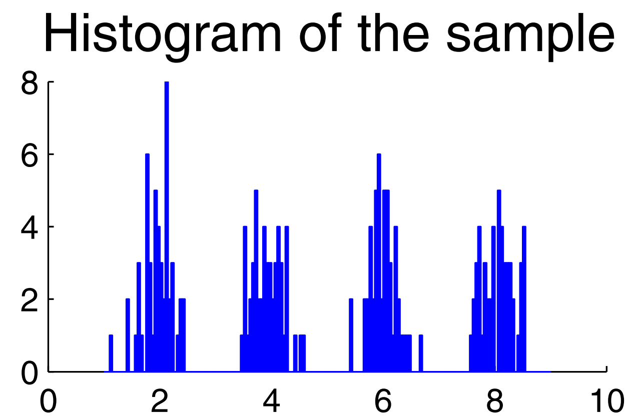


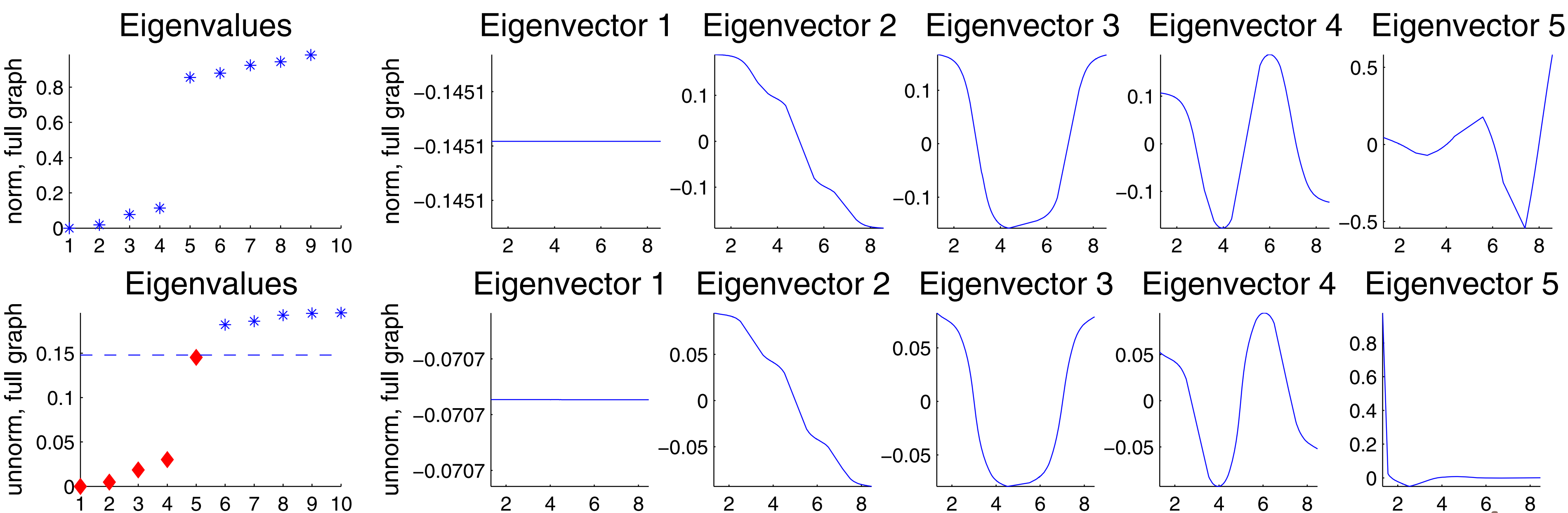


Spectral clustering example: GMM

von Luxburg, 2007

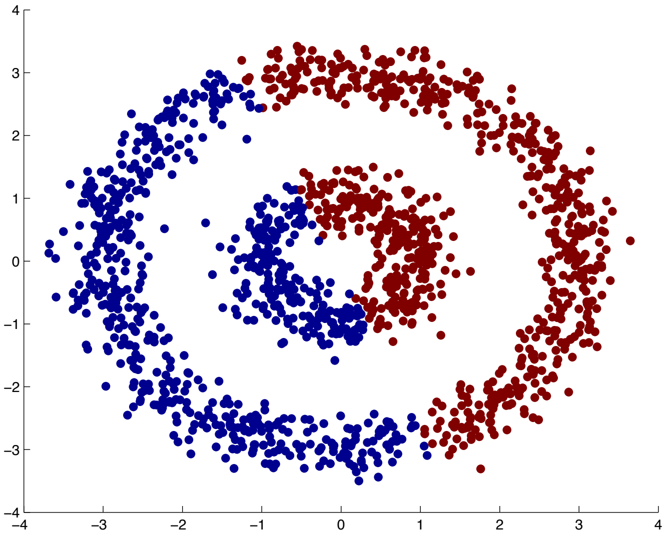
- Data generated from a mixture of 4 Gaussians in 1D
- W = S
- Top row: normalized L_{rw}
- Bottom row: unnormalized L_{un}



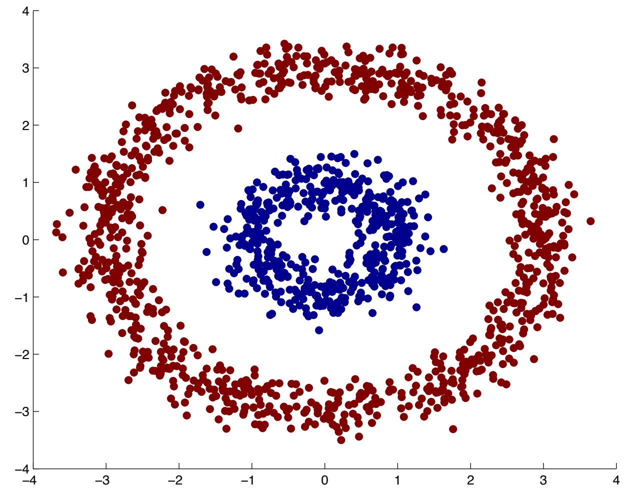


Spectral clustering example: circles

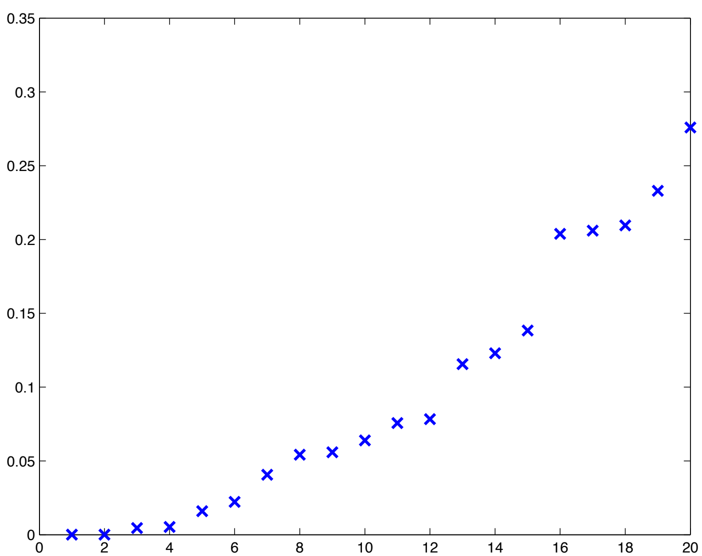
Courtesy: Sriram Sankararaman



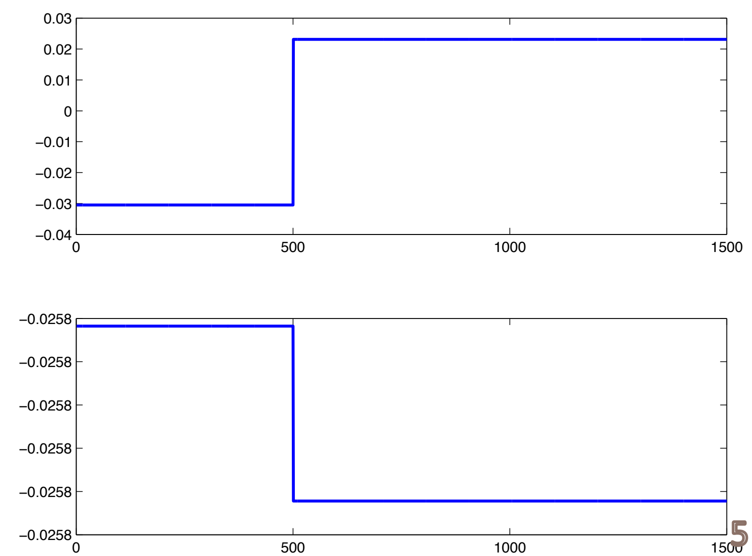
***k*-means, 2 clusters**



Spectral clustering, 2 clusters



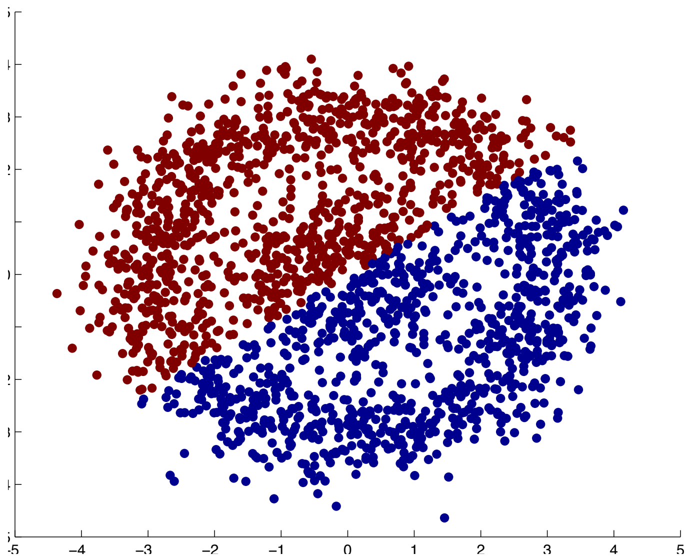
Eigenvalues of L



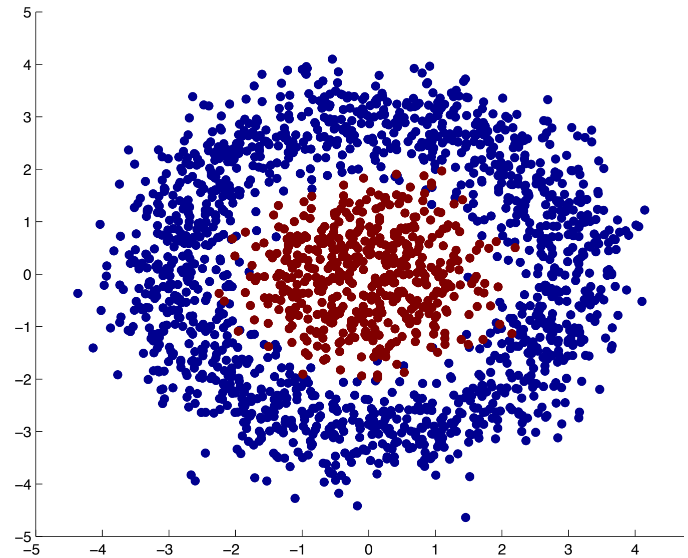
Eigenvectors of L

Spectral clustering example: noisy circles

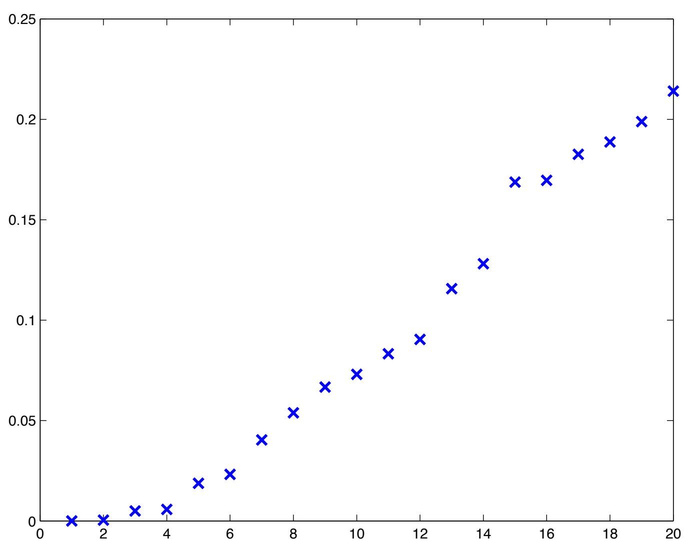
Courtesy: Sriram Sankararaman



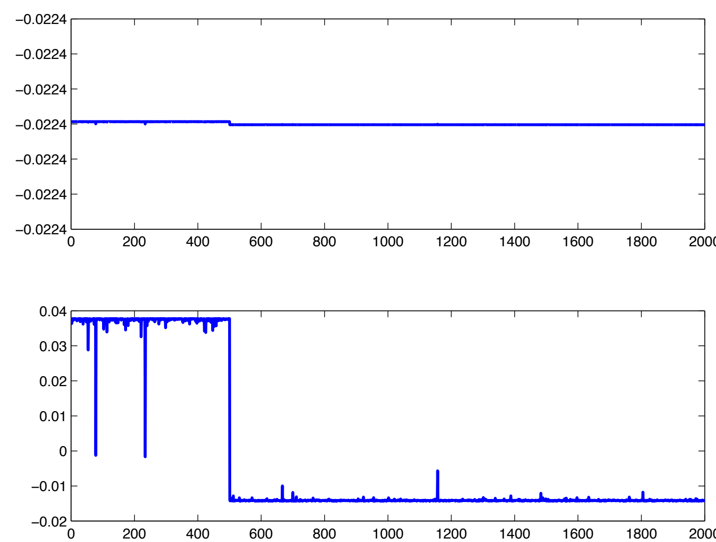
***k*-means, 2 clusters**



Spectral clustering, 2 clusters



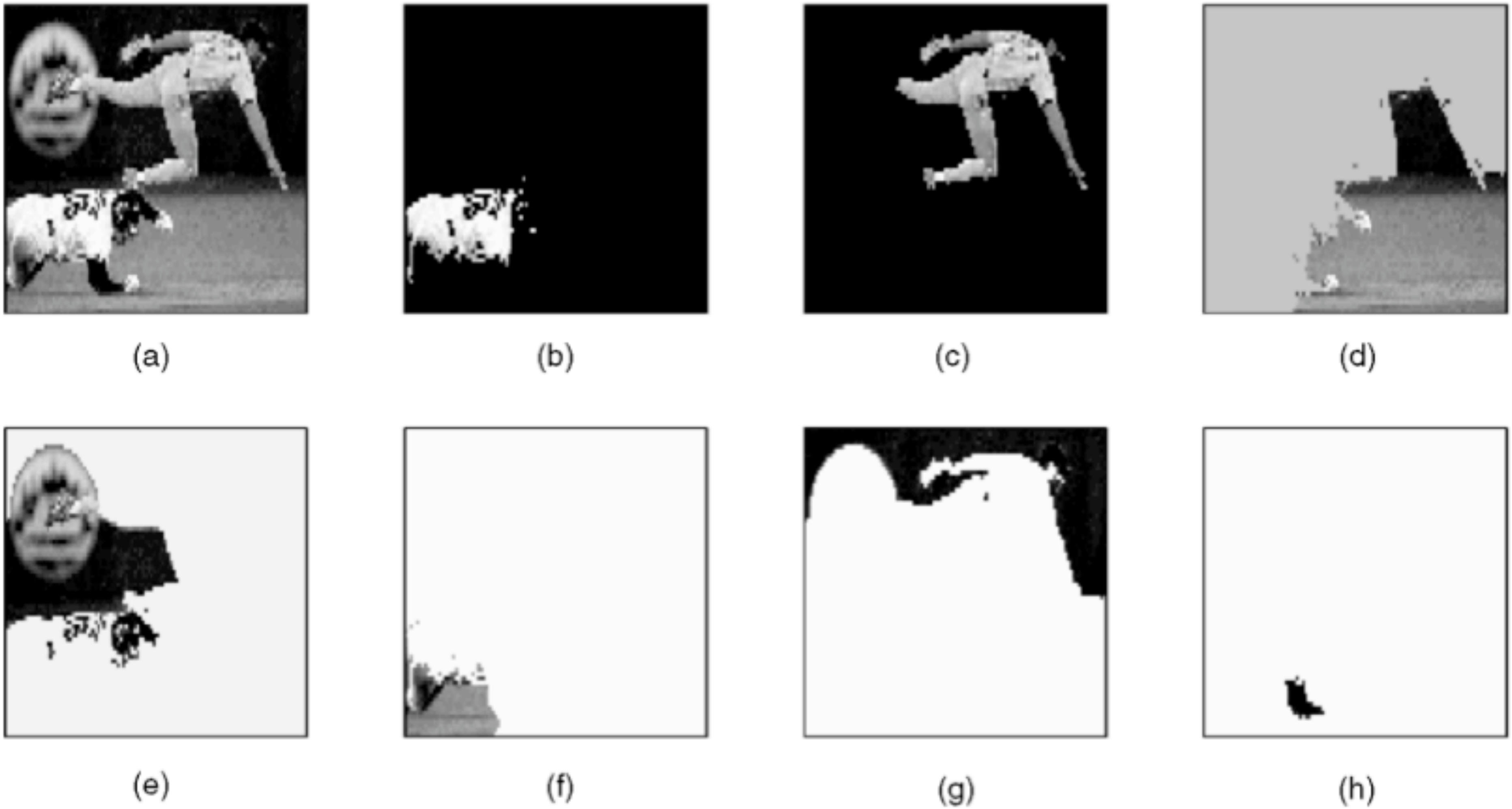
Eigenvalues of L



Eigenvectors of L

Spectral clustering image segmentation

- Spectral clustering widely used in image segmentation



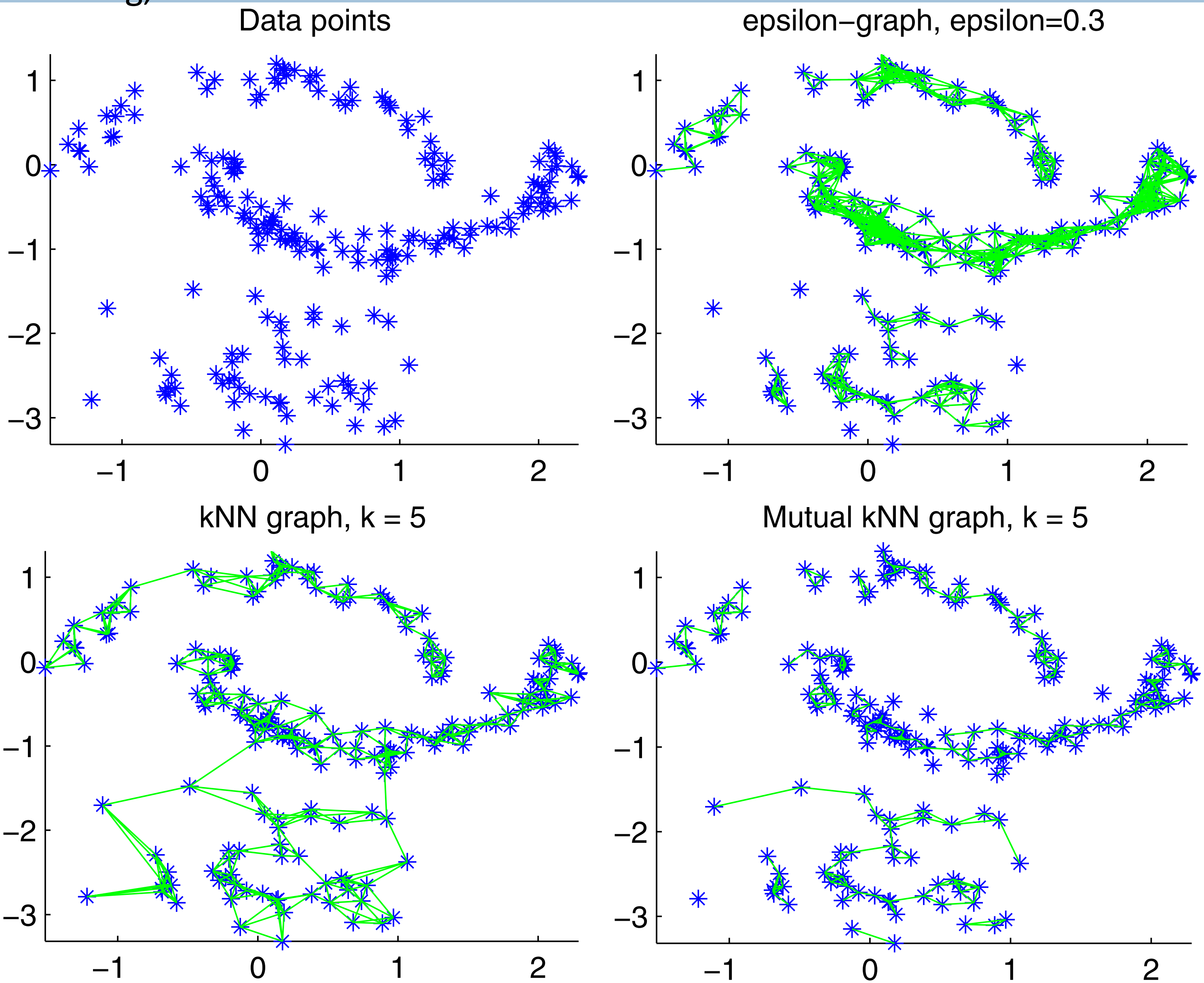
Shi and Malik 2001

How to construct graph weights W?

- **Goal:** capture **local** neighborhood relationships between points / focus on very similar points
- Most common constructions
  - ϵ-neighborhood graph: connect all points with similarity > ϵ
    - Use same weight for all connected points
  - *k*-nearest neighbor graph: connect i and j if j is among the *k*-most similar vertices to i **or** vice-versa
    - Weight retained edges according to similarity
  - mutual *k*-nearest neighbor graph: connect i and j if j is among the *k*-most similar vertices to I **and** vice-versa
    - Weight retained edges according to similarity
  - fully connected graph: connect all nodes
    - Only useful when "local" similarity measure used like $s_{ij} = \exp(-||x_i - x_j||^2 / (2\sigma^2))$, which decays rapidly

Spectral clustering graph examples

von Luxburg, 2007



Spectral clustering and optimality

- Is spectral clustering optimal in any sense? If so, for what objective?
  - One variant minimizes a relaxation of the normalized cut graph partitioning criterion (Shi and Malik, 2000)
    - Same variant, based on L_{rw}, approximately minimizes probability that a random walk on the weighted graph transitions from one cluster to another
  - Consistency studied under certain statistical models (e.g., Rohe/Chatterjee/Yu, 2010 - Spectral clustering and the high-dimensional stochastic blockmodel)

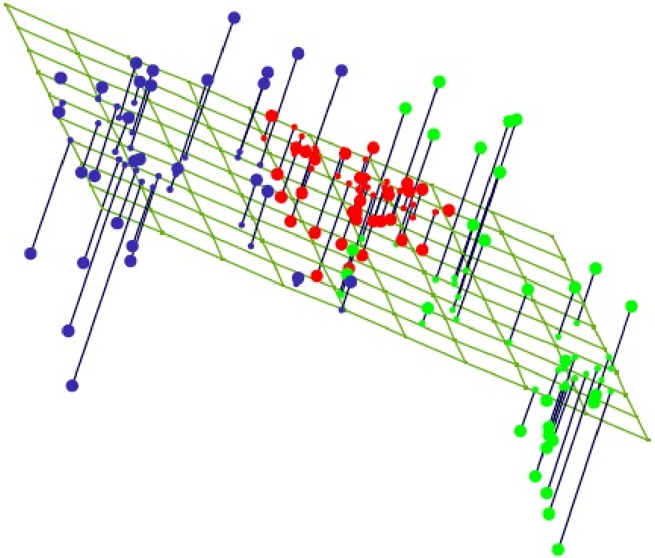
Dimensionality reduction

- **Goal:** Find a low-dimensional representation that captures the "essence" of higher-dimensional data points
  - Also known as **latent feature modeling**
- **Motivation**
  - Compression for improved storage and computational complexity
  - Visualization for improved human understanding of data
    - Difficult to plot / interpret data in more than 3 dimensions
  - Noise reduction
    - Ameliorates noisy and infrequent measurements, missingness
  - Preprocessing for supervised learning task
    - Reduced / denoised representations may lead to better performance or act as regularization for reduced overfitting
  - Anomaly detection
    - Characterize normal data and distinguish from outliers

Linear dimensionality reduction

- **Given:** High-dimensional datapoints $x_i \in R^p$
  - e.g., images of faces in R^{361}

- **Goal:** Assign useful representations $z_i = U^T x_i \in R^k$, where a $U^T \in R^{k \times p}$ is a **linear mapping** into a low-dimensional space



- How to choose a useful U?

Blackboard discussion

- See lecture notes