Multidimensional scaling
Based in part on slides from textbook, slides of Susan Holmes

©Jonathan Taylor

October 10, 2012
Multidimensional scaling

A visual tool

- Recall the PCA scores were

$$\tilde{X}V = U\Delta$$

where $$\tilde{X} = HXS^{-1/2} = U\Delta V^T$$.

- Above, $$U$$ are eigenvectors of

$$\tilde{X}\tilde{X}^T = HXS^{-1}X^TH$$

- Also, $$XS^{-1}X^T$$ is a measure of similarity between cases.
Multidimensional scaling

Classical multidimensional scaling

- Given a similarity matrix $A$, recall the relationship between a similarity and “distance”

$$D_{ij} = (A_{ii} - 2A_{ij} + A_{jj})^{1/2}$$

- Now, consider the matrix $B_{ij}$ with entries

$$B_{ij} = -\frac{1}{2}D_{ij}^2.$$  

- Finally, consider the matrix

$$C = HBH$$
Multidimensional scaling

### Classical multidimensional scaling

- If \( A = X S^{-1}X^T \) is a similarity matrix for a (scaled) data matrix \( X \).
- Then,
  \[
  B = A - \frac{1}{2}(\nu 1^T + 1\nu^T)
  \]
  \[
  = X S^{-1}X^T - \frac{1}{2}(\nu 1^T + 1\nu^T)
  \]
- Above, \( \nu = \text{diag}(A) \).
- Therefore,
  \[
  C = HAH = \tilde{X}\tilde{X}^T = U\Delta^2U.
  \]
- In short, the eigenvectors of \( C \) are the PCA scores.
Olympic PCA scores
Olympic MDS scores
MDS vs. PCA
MDS vs. PCA
Multidimensional scaling

Classical multidimensional scaling

- We can form $B, C$ for any dissimilarity matrix $D$.
- The matrix $C$ will be symmetric, so it will be diagonalizable as $C = W \Lambda W^T$ with $W_{n \times k}$ and $\text{rank}(C) = k$. In the general case the diagonal entries $\Lambda_{ii}$ are not necessarily non-negative.
- This leads to a Euclidean representation

$$ (W \Lambda^{1/2})_{n \times k} $$

with each row of $W \Lambda^{1/2}$ being a point in Euclidean space.
- Taking the first two columns of $W \Lambda^{1/2}$ gives a two-dimensional representation.
Multidimensional scaling

Classical multidimensional scaling

- The points \((W^{1/2})[1:2]\)

are an optimal Euclidean embedding in the sense that their interpoint distances are chosen to be close to the dissimilarities in \(D\).

- If the dissimilarity is Euclidean and the points lie in a 2-dimensional plane in \(\mathbb{R}^n\), then the interpoint distances will be identical . . .
Distances between US capitals
Distances between US capitals
Map of U.S. capitals
MDS of Iris data, $\ell_2$
Iris PCA scores
MDS of Iris data, $\ell_1$
MDS of Iris data, $\ell_\infty / \sup$ norm
MDS of Iris data, $\ell_{20}$
MDS of Iris data, $\ell_{0.2}$
Other applications

Manifold learning

- Goal of manifold learning is to find “low-dimensional” representation $s$ of $X$ that are not necessarily linear.
- Examples of techniques:
  - ISOMAP
  - Laplacian eigenmaps / diffusion geometry
  - Local linear embedding
Manifold learning
Manifold learning

Graph distance

- In the “bottleneck”, we already have a metric (Euclidean distance), so we can use MDS.
- But we might want to emphasize the fact that the groups are “barely connected” ...
- How can we emphasize this bottleneck in terms of a distance?
- ISOMAP and Diffusion Geometry method does this by creating a graph and creating a new metric based on the graph.
Graph distances

- In our “bottleneck” picture, let’s connect $k$ mutual nearest neighbours to form a graph $G_k$.
- That is, insert an edge $(i, j)$ if either
  1. $i$ is within $j$’s $k$ nearest neighbours;
  2. $j$ is within $i$’s $k$ nearest neighbours.
- Let
  
  $$D_{G_k}(i, j) = \text{length of shortest path between vertices } i \text{ and } j \text{ in } G_k.$$  
- This is a metric on $G_k$. 

Manifold learning

ISOMAP

- Shove $D_{G_k}$ through classical MDS.
- ISOMAP can “flatten” data if there is a low-dimensional Euclidean configuration where the Euclidean distances well approximate the graph distances $D_{G_k}$.
- This means that the data has to be “flat” if ISOMAP is to recover it.
Manifold learning