Multidimensional Scaling

Matthew S. Shotwell, Ph.D.

Department of Biostatistics
Vanderbilt University School of Medicine
Nashville, TN, USA

April 15, 2020

Multidimensional scaling (MDS)

- unsupervised learning
- dimension reduction
- map points in high dimension to lower dimension

Multidimensional scaling (MDS)

- $x_1, \ldots, x_N$ in p-dimensions
- let d_{ij} be distance between x_i and x_j
- often $d_{ij} = ||x_i - x_j||$ - Euclidean distance
- only d_{ij} is needed for MDS
- find $z_1, \ldots, z_N \in \mathbb{R}^k$, for $k < p$, to minimize "stress function":

$$S_M(z_1, \ldots, z_N) = \sum_{i \neq i'} (d_{ii'} - ||z_i - z_{i'}||)^2$$

- "Find a lower-dimensional representation $(z_1, \ldots, z_N)$ of the orignal data $(x_1, \ldots, x_N)$ that preserves the pairwise distances among the original data points as well as possible."
- use gradient descent to minimize S_M

Sammon mapping

- "Sammon mapping" is MDS with alternative stress function

$$S_{S_m}(z_1, \ldots, z_N) = \sum_{i \neq i'} \frac{(d_{ii'} - ||z_i - z_{i'}||)^2}{d_{ii'}}$$

- S_{S_m} puts emphasis on preserving smaller pairwise distances

Classical scaling

- Alternative stress function

$$S_C(z_1, \dots, z_N) = \sum_{i \neq i'} (s_{ii'} - (z_i - \bar{z})^T (z_{i'} - \bar{z}))^2$$

  where $s_{ii'}$ is a similarity measure
- called "Classical scaling"
- explicit solution in terms of eigenvectors (ex. 14.11)
- if $s_{ii'} = (x_i - \bar{x})^T (x_{i'} - \bar{x})$ then classical scaling equivalent to PCA

Nonmetric scaling

- Alternative stress function

$$S_{NM}(z_1, \ldots, z_N) = \frac{\sum_{i \neq i'} (\theta(d_{ii'}) - ||z_i - z_{i'}||)^2}{\sum_{i \neq i'} ||z_i - z_{i'}||^2}$$

  where θ is an increasing function
- iterative minimization algorithm
  1. fix θ, minimize over z_i using gradient descent
  2. fix z_i, use monotonic regression to fit $\theta(d_{ii'})$ to $||z_i - z_{i'}||$

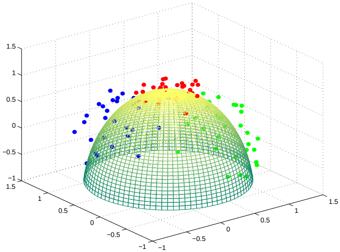


FIGURE 14.15. *Simulated data in three classes, near the surface of a half-sphere.*

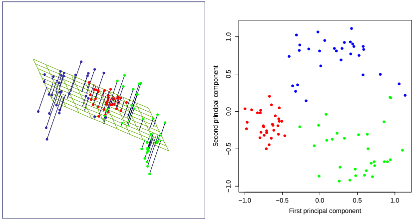


FIGURE 14.21. *The best rank-two linear approximation to the half-sphere data. The right panel shows the projected points with coordinates given by* $\mathbf{U}_2\mathbf{D}_2$*, the first two principal components of the data.*

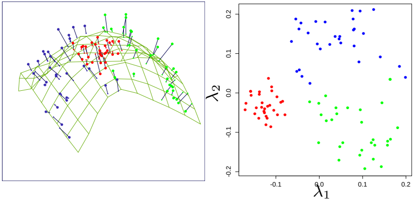


FIGURE 14.28. *Principal surface fit to half-sphere data. (Left panel:) fitted two-dimensional surface. (Right panel:) projections of data points onto the surface, resulting in coordinates* $\hat{\lambda}_1, \hat{\lambda}_2$.

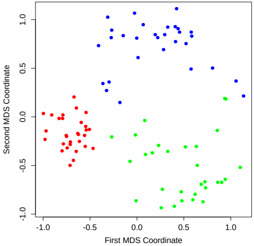


FIGURE 14.43. *First two coordinates for half-sphere data, from classical multi-dimensional scaling.*

Advandage of PCA/principal surfaces vs. MDS

- PCA/principal surfaces: easy to map *new* data into lower dimension
- MDS: no clear mapping of *new* data into lower dimension

ISOMAP and Local MDS

- MDS can behave badly if distance metric not adequate to characterize curvature
- ISOMAP or local MDS can help with this

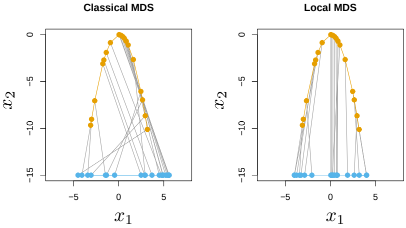
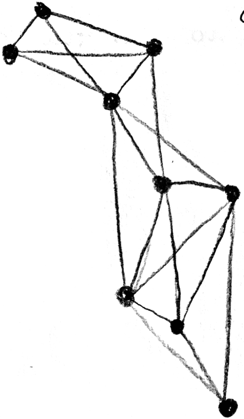


FIGURE 14.44. *The orange points show data lying on a parabola, while the blue points shows multidimensional scaling representations in one dimension. Classical multidimensional scaling (left panel) does not preserve the ordering of the points along the curve, because it judges points on opposite ends of the curve to be close together. In contrast, local multidimensional scaling (right panel) does a good job of preserving the ordering of the points along the curve.*

Isometric feature mapping (ISOMAP)

- create k-NN graph
- calculate geodesic distance between all points
- use classical scaling on geodesic distances

3 - NN Graph



geodesic distance: shortest path between two points

Local MDS

- let $\mathcal{N}$ be pairs of k-NNs
- (i, i') in $\mathcal{N}$ if x_i is a k-NN of $x_{i'}$ or vice versa
- stress function

  $$S_L(z_1, \ldots, z_N) = \sum_{(i,i') \in \mathcal{N}} (d_{ii'} - ||z_i - z_{i'}||)^2 + \sum_{(i,i') \notin \mathcal{N}} w(D - ||z_i - z_{i'}||)^2$$

  where D is large constant and w is a weight
- pairs that are not neighbors considered very far apart
- w small so that non-neighbors don't dominate stress fun

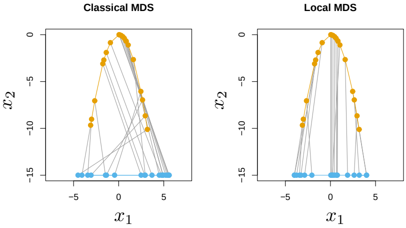


FIGURE 14.44. *The orange points show data lying on a parabola, while the blue points shows multidimensional scaling representations in one dimension. Classical multidimensional scaling (left panel) does not preserve the ordering of the points along the curve, because it judges points on opposite ends of the curve to be close together. In contrast, local multidimensional scaling (right panel) does a good job of preserving the ordering of the points along the curve.*

Local MDS example: faces

- 1965 20×28 grayscale images
- reduce dimension from $20 \times 28 = 560$ to 2 using Local MDS

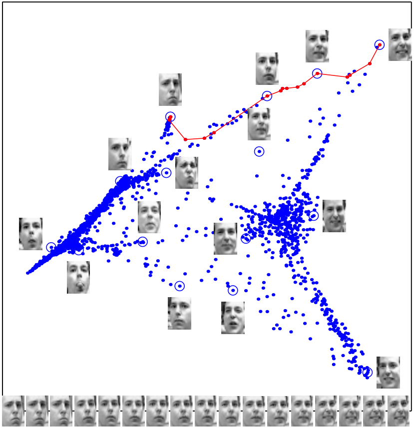


FIGURE 14.45. *Images of faces mapped into the embedding space described by the first two coordinates of LLE. Next to the circled points, representative faces are shown in different parts of the space. The images at the bottom of the plot correspond to points along the top right path (linked by solid line), and illustrate one particular mode of variability in pose and expression.*