Metric Multidimensional scaling (MDS)

- Problem: from distances $D_{ij}$, how to compute $\bar{x}_i \in \mathbb{R}^d$ such that $\|\bar{x}_i - \bar{x}_j\| \approx D_{ij}$?
- Step #1: derive Gram matrix of inner products
  \[ G_{ij} = \frac{1}{2} \left[ \frac{1}{N} \sum_{k} (D_{ik}^2 + D_{jk}^2) - \frac{1}{N^2} \sum_{k} \sum_{l} D_{kl}^2 - D_{ij}^2 \right] \]
  - Step #2: singular value decomposition (SVD)
  \[ G = V \Sigma V^T \quad \text{with} \quad \Sigma = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_N \end{pmatrix} \text{ and } V^TV = I \]
  \[ \bar{x}_{\alpha} = \sqrt{\sigma_{\alpha}} \cdot V_{\alpha} \text{ for } i=1..N, \alpha=1..d \]

- Estimating the dimensionality $d$
  Singular values $\sigma_{\alpha}$ measure variance in $\alpha$th dimension:
  \[ \text{Var}[x_{\alpha}] = \frac{1}{N} \sum_{i=1}^{N} x_{i\alpha}^2 \]
  b/c $\{\bar{x}_i\}$ are centered
  \[ = \frac{1}{N} \sum_{i=1}^{N} \sigma_{\alpha} \cdot V_{\alpha i}^2 \]
  \[ = \frac{\sigma_{\alpha}}{N} \quad \text{b/c } V^TV = I \]

Rules of thumb:
1) Look for "elbow" in eigenvalue spectrum
2) Choose $d$ to capture $(1-\delta)$ fraction of total variance:
\[ \sum_{\alpha=1}^{d} \sigma_{\alpha} \geq \left( \frac{1-\delta}{\delta} \right) \sum_{\alpha=d+1}^{N} \sigma_{\alpha} \]
Duality of MDS and PCA

- Inputs $\tilde{x}_i \in \mathbb{R}^D$; assume inputs are centered $\bar{\sum x}_i = \vec{0}$

\[
X = \begin{bmatrix}
\tilde{x}_1 & \tilde{x}_2 & \cdots & \tilde{x}_N \\
\end{bmatrix}^{\top} 
\]

- Inner products $(N \times N)$

$G_{ij} = \tilde{x}_i \cdot \tilde{x}_j \iff G = XX^T$

- Covariance matrix $(D \times D)$

$C_{\alpha \beta} = \frac{1}{N} \sum_{i=1}^{N} x_i \tilde{x}_i \iff C = \frac{1}{N} (XX^T)$

- Eigenvalues

$G$ and $C$ have same non-zero eigenvalues up to factor $(\frac{1}{N})$.

# non-zero eigenvalues $\leq \min (D, N)$

- Equivalence

PCA on inputs $\{\tilde{x}_i\}_{i=1}^{N}$ gives same results as MDS on distances $\|\tilde{x}_i - \tilde{x}_j\|$.

Maximum variance subspace = Maximally inner-product preserving subspace
Manifold learning

- Problem: given high dimensional data that lies on (or near) a low dimensional manifold, how to compute a "faithful" embedding?

- Examples of manifolds
  1) Spiral
     extrinsic coordinates: \((x, y)\) in 2D plane
     intrinsic coordinates: arc length along spiral
     no linear projection can "unfold" the spiral
     \((x, y) \rightarrow \Theta\)

  2) Images of car from different viewpoints
     extrinsic coordinates: pixels
     intrinsic coordinates: camera location 3d

  3) Vowel sounds
     extrinsic coordinates: waveform samples
     intrinsic coordinates: articulator positions
     (e.g. tongue, lips, etc.)

- Notation
  inputs \(\tilde{x}_i \in \mathbb{R}^D\) (\(i=1, \ldots, N\))
  outputs \(\tilde{y}_i \in \mathbb{R}^d\) (\(d \ll D\))

- Goals
  - nearby points remain nearby
  - distant points remain distant
  - estimate dimensionality \(d\)
• Non-monotonicity

Rank ordering of Euclidean distances will not be preserved

\[ \|A-C\| > \|A-B\| \]
\[ \|A-C\| < \|A-B\| \]

• Linear vs. nonlinear

What computational price must we pay for nonlinear dimensionality reduction (NLDR)?

• Conventional wisdom (pre-2000)

NLDR requires difficult nonlinear optimization:
- latent variable models (mixture of factor analyzers)
- autoencoder neural networks
- self-organizing maps

Pre-2000: many local minima!

• Since 2000

spectral methods for NLDR:
- based on tractable matrix computations
- not much more complicated than MDS

no local minima
Isomap = "isometric mapping" 
\( \sim \) distance-preserving

**Step #1:** compute neighborhood graph \( \sim O(DN^2) \)
- Nodes represent inputs \( \{x_i\}_{i=1}^N \)
- Edges represent k-nearest neighbor (kNN) relations
- k is ONLY free parameter of algorithm

Ex: spiral graph is discretized skeleton of underlying manifold (k=2)

**Step #2:** compute shortest paths \( \sim O(N^3) \)
- Weight each edge by local distance between kNN
- Compute distance \( D_{ij} \) of shortest path from node \( i \) to node \( j \) (e.g. Dijkstra's algorithm)
- Graph distances approximate "geodesic" distances along manifold
  (Not the same as Euclidean distances, except for kNN)

**Step #3:** Run MDS on graph-based distances
- Derive Gram matrix
- Perform SVD
- Estimate (manifold) dimensionality from eigenvalue spectrum

- Strengths of Isomap
  - polynomial time optimizations
  - no local minima
  - non-iterative (just one pass through data)
  - non-parametric (no explicit assumptions about form of manifold)
  - only heuristic is neighborhood size
Asymptotic convergence

Thm: For data sampled from a manifold that is isometric to a convex subset of Euclidean space, Isomap will recover the subset up to rotation and translation.

Issues
- sensitivity to "short cuts" in graph when data is sparse or noisy
- graph distances $D_{ij}$ may not yield a positive semi-definite Gram matrix
- convexity assumption: manifold has no "holes"

-holes introduce distortions into the embedding
- spurious extra dimensions appear in embedding