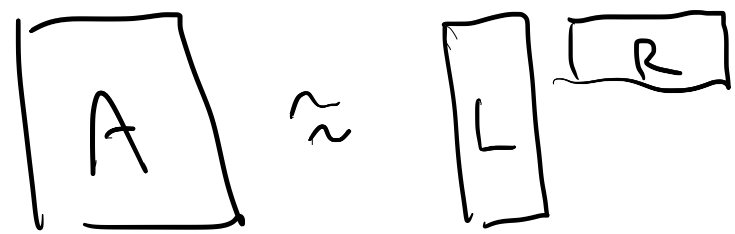
Feb 27, 2020

Lots of dimensionality reduction / latent factors

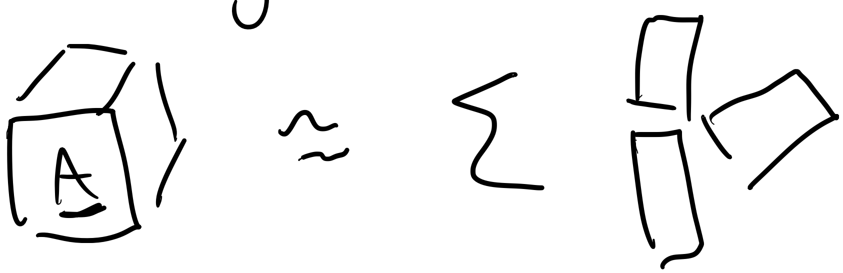
PCA, CUR, NMF, MC

Goal: reconstruct data



Last lecture: generalized to tensors

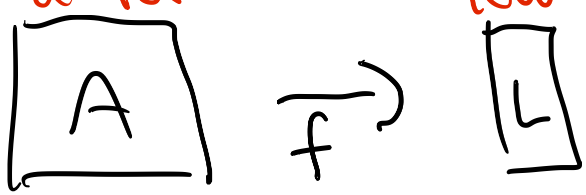
CP:



Today: "nonlinear dimensionality reduction"

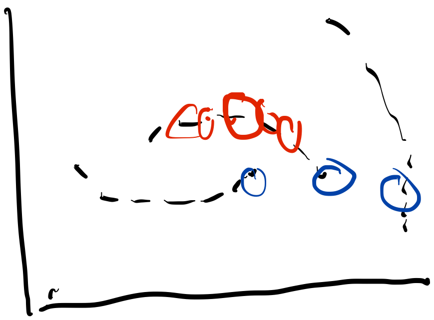
data

low-dim factor



Idea: look locally to find the manifold

Algorithm: Isometric feature mapping (ISOMAP)



(1) form k-nearest neighbors graph

(2) All-pairs shortest paths

(3) normalize + TSVD

(1) Approx who is a neighbor on manifold

(2) Approx geodesic distance: D

(3) $d_{ij}^2 \approx \|x_i - x_j\|_2^2$ (Multidimensional scaling)

x low-dimensional

$$\approx \|x_i\|_2^2 + \|x_j\|_2^2 - 2x_i^T x_j$$

$$D^{(2)} \simeq -2X^TX + z\mathbb{1}^T + \mathbb{1}^T z$$

$$z = \begin{pmatrix} \|x_1\|_2^2 \\ \vdots \\ \|x_n\|_2^2 \end{pmatrix} \qquad H = \left(I - \frac{1}{n}\mathbb{1}\mathbb{1}^T\right)$$

$$H^T D^{(2)} H \simeq -2\underbrace{H^TX^T}_{Y^T}XH + \cancel{H^T(z\mathbb{1}^T)H} + \cancel{H^T(\mathbb{1}z^T)H}$$

$$\mathbb{1}^T H = \mathbb{1}^T - \frac{1}{n}\underbrace{\mathbb{1}^T\mathbb{1}}_{n}\mathbb{1}^T = 0$$

$$\underbrace{-\frac{1}{2}H^TD^{(2)}H}_{M} \simeq Y^TY \quad \Big\} \quad M \text{ SPD}$$

$$\min_B \|M - B\|_F \quad \text{s.t.} \quad B \text{ rank } k$$

$$B = U_k\Sigma_kV_k^T \Rightarrow U_k\Sigma_kU_k^T$$

$$Y = U_k\sqrt{\Sigma_k}$$

Idea: avoid finding "proper" pairwise dists

instead: should be "locally linear" on manifold

reconstruct each point from its neighbors

Locally linear embedding (LLE)

1. Find k nearest neighbors $N(i)$ of point i

2. Local approx:

$$a_i \approx \sum_{j \in N(i)} W_{ij} a_j \qquad a_i \in \mathbb{R}^n$$

$$\text{constraint: } \sum_{j \in N(i)} W_{ij} = 1$$

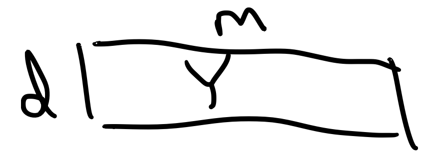
$$\min_W \|A - AW\|_F^2 \quad \text{s.t.} \quad \mathbb{1}^T W = \mathbb{1}$$

③ Reconstruction

$a_i \in \mathbb{R}^n$

$y_i \in \mathbb{R}^d \quad d << n$

$$\min_Y \|Y - YW\|_F^2$$



$$= \min_Y \|Y(I-W)\|_F^2$$

$\sum y_i = 0$ unit-covariance: $\frac{1}{d} Y Y^T = I$

$$\min_Y \operatorname{tr}\left(Y (I-W)(I-W)^T Y^T\right) \qquad \operatorname{tr}(X^T A X)$$

$$\text{s.t.} \quad \frac{1}{N} Y Y^T = I \qquad\qquad \text{s.t.} \quad X^T X = I$$

$Y =$ last eigenvectors of $(I-W)(I-W)^T$

t-SNE (2008)

Based on stochastic neighborhood embedding (SNE) 2003

"distance" proportional to probabilities

$$d_{ij}^2 = \|x_i - x_j\|_2^2 / 2\sigma_i^2 \leftarrow \text{tuning param}$$

$$P_{ij} = \exp(-d_{ij}^2) \Big/ \sum_{k \neq i} \exp(-d_{ik}^2)$$

$$d_{ij}^2 = \|x_i - x_j\|_2^2 = \|x_i\|_2^2 + \|x_j\|_2^2 - 2x_i^T x_j$$

$$\sigma_i = 1, \quad \|x_k\|_2^2 = 1 \ \forall k$$

$$-d_{ij}^2 = \tfrac{1}{2}(-2 + 2x_i^T x_j) = -1 + x_i^T x_j$$

$$P_{ij} = \exp(\cancel{-1+}\, x_i^T x_j) \Big/ \sum_{k \neq i} \exp(\cancel{-1+}\, x_i^T x_k)$$

$$Q_{ij} = \frac{\exp(-\|y_i - y_j\|_2^2)}{\sum_{k \neq i} \exp(-\|y_i - y_k\|_2^2)}$$

learn $y_i \in \mathbb{R}^k$

Goal: match $P(i,:)$ and $Q(i,:)$

$$\min_{Y} \sum_i KL(P(i,:) \,\|\, Q(i,:))$$

$$= \sum_i \sum_j P_{ij} \log\left(\frac{P_{ij}}{Q_{ij}}\right) = f(Y)$$

$$\frac{\partial f}{\partial y_i} = 2 \sum_j \underbrace{(y_j - y_i)}_{\text{closeness of points}} \overbrace{(p_{ij} - q_{ij} + p_{ji} - q_{ji})}^{\text{closeness in prob.}}$$

t-SNE t-distributed SNE

① SNE difficult to optimize

② "crowding problem"

$\Rightarrow$ really cool figures

Symmetric SNE distribution over (i,j)

$$\tilde{q}_{ij} = \exp(-\|y_i - y_j\|_2^2) / \sum_{k \neq l} \exp(-\|y_k - y_l\|_2^2)$$

$$= \tilde{q}_{ji}$$

$$\tilde{p}_{ij} = \frac{\exp(-\|x_i - x_j\|_2^2)}{\sum_{k \neq l} \exp(-\|x_k - x_l\|_2^2)}$$

outlier x_i

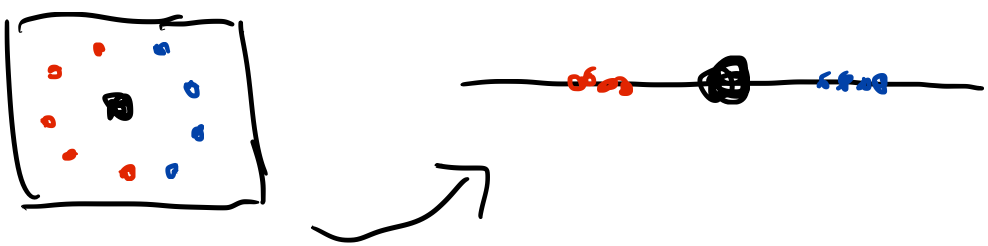
$\tilde{p}_{ij}$ small for all j

$$\tilde{p}_{ij} = \frac{p_{ij} + p_{ji}}{2n}$$

$$\sum_j \tilde{p}_{ij} = \frac{1}{2n}\Big(\underbrace{\sum_j p_{ij}}_{1} + \underbrace{\sum_j p_{ji}}_{>0}\Big)$$

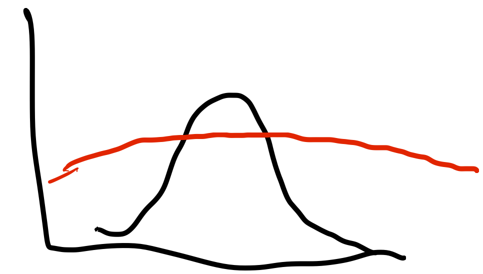
$$> \frac{1}{2n}$$

"crowding problem"



$2D \to 1D$

$$\tilde{q}_{ij} = \frac{\exp(-\|y_i - y_j\|_2^2)}{\sum_{k \neq l} \exp(-\|y_k - y_l\|_2^2)}$$



$$p(x) \propto \exp(-x^2)$$

$$p(x) \propto \frac{1}{1+x^2}$$

$$\tilde{q}^{t}_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l} (1 + \|y_k - y_l\|^2)^{-1}}$$

$$\min_{Y} KL(\tilde{p} \,\|\, \tilde{q}^t)$$

$$\frac{\partial g}{\partial y_i} = 4 \sum_j (y_i - y_j)(\tilde{p}_{ij} - \tilde{q}_{ij})(1 + \|y_i - y_j\|_2^2)^{-1}$$