Feb 28, 2019

So far: lots of dimensionality reduction

main ideas:

- find latent/hidden factors
- reduce noise

PCA, CUR, NMF

linear dimensionality redux

$$A \xrightarrow{P} LR$$

linear transformation

data matrix

low-dim. factors

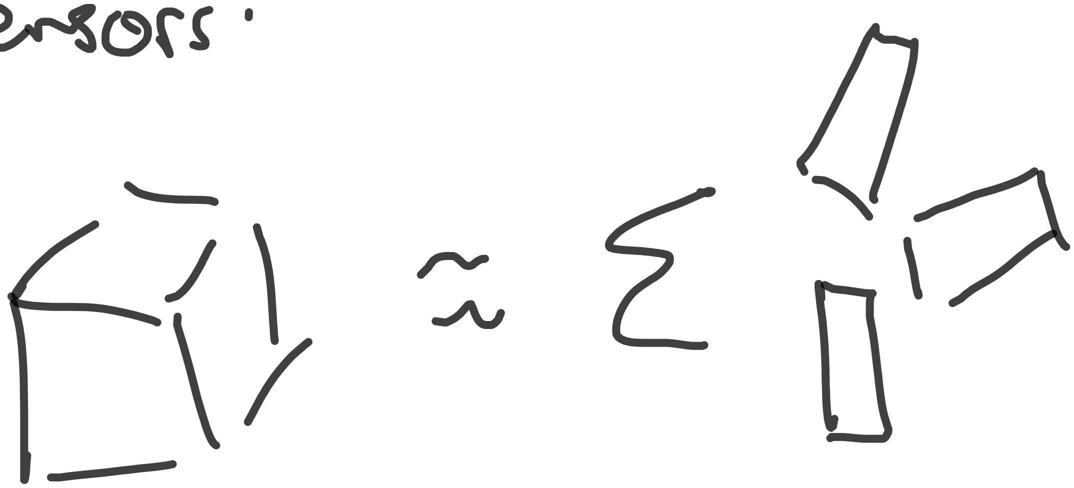
PCA: $A \xrightarrow{V_k} U_k \Sigma_k$

data matrix

PCs

$$A \approx U_k \Sigma_k V_k^T$$

tensors:



CP decomposition

"multilinear dimensionality redux"

Today: "nonlinear" dim redux

$$A \xrightarrow{f} L$$

f: arbitrary function

A: data matrix

L: low-dim representation

- often lose theoretical understanding
- gain increased expressibility

1. ISOMAP
2. Locally Linear Embedding (LLE)
3. SNE / t-SNE

Method: Isometric feature mapping (ISOMAP)

Tenenbaum, de Silva, Langford '00

Idea: look locally

Alg:

① k-nearest neighbor graph with distances

② all pairs shortest paths

→ D_{ij} = shortest path distance

[②.5 take conn. components]

(3) Normalize distances and take top few singular vectors

---

(1) Approximating who is a neighbor on the manifold

(2) Approximating geodesic distance (note: need approx. uniform sampling)

(3) $\tau(D) = -HD(G)^2H/2$

$D(G)$ dist. matrix

$H = (I - \frac{1}{n}\mathbb{1}\mathbb{1}^T)$

$$\min_{Y} \|\tau(D) - \tau(D_Y)\|_F^2$$

$n\ \overset{k}{\boxed{Y}}$ $$(D_Y)_{ij} = \|y_i - y_j\|_2$$

Solution: $\tau(D) = V \Lambda V^T$

$$Y_k = \sqrt{\Lambda_k}\, V_k$$

Method: Locally Linear Embedding (LLE)

(Roweis & Saul '00)

Avoid trying to find proper pairwise distances

Instead, should be "locally linear" on the manifold

Idea:

1. try to "reconstruct" each point from nearest neighbors
2. embed in low-dim. space from the reconstruction map

---

0. Find k-nearest neighbors

$N(i)$ = nearest neighbors of point i

1. $\text{Error}(i) = \|x_i - \sum_{j \in N(i)} W_{ji} x_j\|_2^2$

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

   Opt: $\min_W \sum_i \text{Error}(i)$ s.t. $\circledast$

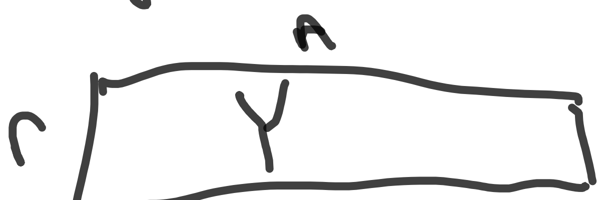
   $= \|XW - X\|_F$ s.t. $\circledast$

   columns of X are the data points

   LLS with simple constraints

   $\Rightarrow$ weights $\hat{W}$

2. Choose some dim. r



$$\min_{Y} \|Y\hat{W} - Y\|_F^2$$

constraints

I $\sum_j \hat{Y}_{ji} = 0$ (0-mean)

II $\frac{1}{n}\sum_i \hat{Y}_i \hat{Y}_i^T = I$

$$Y\underbrace{(I-W)(I-W)}_{M}Y^T$$

$$M = I - W - W^T + WW^T$$

$$1^T M = 1^T - 1^T - w + \cancel{1^T} WW^T$$

1^T

$+ \; w$

$= 0$

II orthogonality constraint

solution: take k smallest eigenvectors of M as the representation

Method: Stochastic Neighborhood Embedding (SNE)

Hinton & Roweis '02

d_{ij} = dist b/w points i and j

x_i and x_j are the vectors

y_i and y_j are representations

$$d_{ij}^2 = \frac{\|x_i - x_j\|_2^2}{2\sigma_i^2}$$

tuning parameter

$$p_{ij} = \frac{\exp(-d_{ij}^2)}{\sum_{k \neq i} \exp(-d_{ik}^2)}$$

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

for each i

$\{p_{ij}\}$, $\{q_{ij}\}$ are prob. dist.

Goal: match those distribs

$$\min_y \sum_i KL(P_i \| Q_i) = f(y)$$

$$= \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

$$\frac{\partial f}{\partial y_i} = 2 \sum_j (y_j - y_i)(p_{ij} - q_{ij} + p_{ji} - q_{ji})$$

minimize over y

$\Rightarrow$ y are SNE representations

Method: t-distributed SNE

(t-SNE)

van der Maaten & Hinton '08

Symmetric SNE

$$KL(P \| Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

$$p_{ij} = p_{ji} \qquad q_{ij} = q_{ji}$$

$$p_{ij} = \frac{\exp\left(-\|x_i - x_j\|^2 / 2\sigma^2\right)}{\sum_{k \neq l} \exp\left(\|x_k - x_l\| / 2\sigma^2\right)}$$

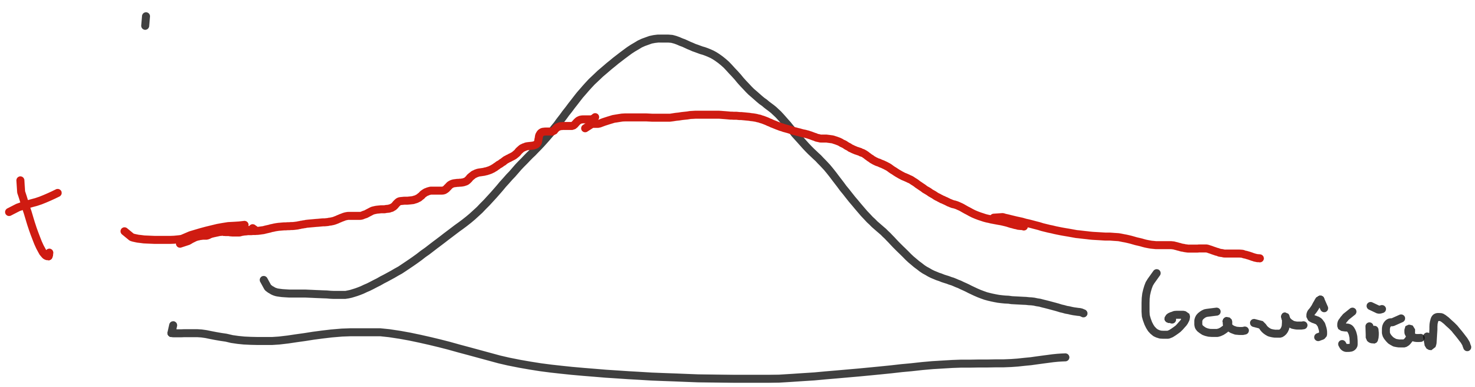
$$q_{ij} = \frac{\exp(-\|y_i - y_j\|^2)}{\sum_{k \neq l} \exp(-\|y_k - y_l\|^2)}$$

$$\frac{\partial f}{\partial y_i} = 4 \sum_j (y_i - y_j)(p_{ij} - q_{ij})$$

"crowding"

manifold in ~10 dimensions and visualize in 2d/3d space, then this is tricky

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



$$\frac{\partial g}{\partial y_i} = 4 \sum_j (y_i - y_j)(p_{ij} - q_{ij})(1 + \|y_i - y_j\|^2)^{-1}$$