November 12, 2020 CS49000-VIZ - Fall 2020 Introduction to Data Visualization **Dimensionality Reduction** Lecture 16

High-dimensional data Dimensionality reduction Manifold learning Computational aspects

Why High-Dimensional? •Data samples with many attributes

- Tables with many columns, medical records, ...
- into high-dimensional *feature* vectors

•Amalgamate individual properties

- "Curse of dimensionality"
	- Sampling becomes exponentially costly
	-
	- intractable
- How to visualize?

• Space accumulates in"corners" of hypercube • Data processing becomes extremely expensive /

• Missing intuition

Volume of unit hyper-ball in n-dimensions: $V_n = \frac{n!}{\Gamma(\frac{n}{2} + 1)}$

CS49000-VIZ Intro to Data Visualization / Fall 2020; Lecture 16: Dimensionality Reduction

Curse of Dimensionality

 $\Gamma($

πn/2

n

 $V_n =$

 $2(k!)(4\pi)$ *k* $(2k + 1)!$

$$
V_{2k} = \frac{\pi^k}{k!} \qquad V_{2k+1} = \frac{2}{k}
$$

Ratio to volume of bounding hyperbox

$$
\frac{V_{2k}}{2^k} = \left(\frac{\pi}{4}\right)^k \frac{1}{k!}
$$

Dimension Reduction

- Assume a point cloud $(\mathbf{x}_i)_{i=1,..,n} \in \mathbb{R}^k$
- Interpret these points as observations of a random variable $X \in \mathbb{R}^k$
- \cdot The empirical mean (centroid) is $c=$ • The covariance matrix is given by 1 *n* \sum *i* x*i* 1 $\mathbf{A} = \bar{\mathbf{X}} \bar{\mathbf{X}}^T$
-

 $A_{jl} =$

 $(x_{ij} - c_j)(x_{il} - c_l)$

n

 \sum

i

• The eigenvectors of the covariance matrix form a data-dependent coordinate system

-
- The first *m* eigenvectors (in decreasing order of associated eigenvalues) span the *m* principal dimensions of the point cloud.

Multidimensional Scaling $\forall (i, j) \mid |\mathbf{x}_i - \mathbf{x}_j|| \approx \delta i, j$

•Input: dissimilarity matrix $\Delta_{ij} = \delta_{i,j}$ with $\Delta_{ij} > 0$ and $\Delta_{ii} = 0$ • Goal: find $x_1, x_2, \ldots, x_n \in \mathbb{R}^d$ such that $\Delta \in \mathbb{R}^{n \times n}$ $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \in I\!\!R^d$

HAH where $H = I - \frac{1}{n}$ *n* $\mathbf{11}^T$

$\overline{1}$ 2 $+$

Method:

- Center(*): $B = -\frac{1}{2}H\Delta H$ where $B = -\frac{1}{2}$ 2
- Spectral decomposition: $B = U\Lambda U^T$
- **Clamp(*)**: $(\Lambda_{+})_{ij} = \max(\Lambda_{ij}, 0)$
- Solution: first d columns of $X = U\Lambda$ This is a PCA!

(): if* \triangle *measures Euclidean distances,* B *is positive semidefinite* Δ measures Euclidean distances, B

11

Manifold Learning

- Move away from linear assumption
- Assume curved low-dimensional geometry Δ order topology in order topology in Δ crystallize this notion of dimensionality. *function whose inverse is also a continuous function.* space of R*^D*. Instead, we assume only that the data lies on a *d*-dimensional manifold embedded into R*^D*
- Assume smoothness Definition 2. *A d-dimensional* manifold *M is set that is locally homeomorphic with* R*^d. That is, for* where $\frac{1}{2}$ and the mani-box $\frac{1}{2}$ and the mani-box $\frac{1}{2}$ and the mani-box $\frac{1}{2}$ and the mani-box $\frac{1}{2}$
- ➡ MANIFOLD *^x, ^Nx, and a homeomorphism ^f* : *^N^x* [→] ^R*^d. These* \blacksquare *nieighborhoods are referred* to assume the patches are referred to assume the patches are referred to assume that \blacksquare *and the map is referred to a a* coordinate chart*. The*

• Manifold learning parameter space*.*

JW-UIITICHSIUTIAI AFREE propriate if the data lies on a low-dimensional sub-

CS49000-VIZ Intro to Data Visualization / Fall 2020; Lecture 16: Dimensionality Reduction matics and the above definition to a
Matter and the above defining the above definition of the controller the set of the set of the State definitio We will be interested only in the case where *M* is a Beduction the set of as many R learning, since we are trying to "learn" a manifold

We are given data *^x*1*, x*2*,...,xⁿ* [∈] ^R*^D* and we wish

fold is given by a *single* coordinate chart.¹ We can now describe the problem formally.

Problem: Given points $x_1, \ldots, x_n \in \mathbb{R}^D$ that lie on a *d*-dimensional manifold *M* that can be described by a single coordinate chart $f: M \to \mathbb{R}^d$, find $y_1, \ldots, y_n \in$ \mathbb{R}^d , where $y_i \stackrel{\text{def}}{=}$ $\stackrel{\text{def}}{=}$ $f(x_i)$.

, 100

- •Form k-NN graph of input points
- Form dissimilarity matrix Δ as distance between points •Compute MDS!

square of approximated geodesic

ion from source to v al path from source o source are unoptimized - thus are in Q

en removed from O. in 1st step – dist[u] is infinity yet

•Computing geodesic distance • standard graph problem in CS

Dijkstra Algorithm $O(|E| + |V| \log |V|)$

X. Two simple methods are to connect each R EPORTS

• Successful in computer vise B \blacksquare ing neighboring data points. p^2 The complete isometric feature mapping, In its monthly or Islam Isomap, algorithm has the Isomap Series steps in geodesic distances *dM* (*i*,*j*) between all pairs are detail of points on the manifold *M* by computing their shortest path distances *dG*(*i*,*j*) in the problems A **Fig. 1.** (**A**) A canonical dimensionality reduction problem from visual perception. The input consists $\sigma = 4096-4096$

Varuina pose

16. This procedure, known as Floyd's algorithm, requires

 $g_{\rm eff}$ will high probability $h_{\rm eff}$ and $h_{\rm eff}$ and $h_{\rm eff}$ are a path notation of $h_{\rm eff}$

much longer than the true geodesic, but small

enough to prevent edges that \mathbf{s}

geometry of the manifold. More precisely, given ar-

bitrarily small values of '1, '2, and (, we can guar-

will hold uniformly over all pairs of data points *i*, *j*. For

 $|z|$

 \mathcal{Q}

 $\overline{2}$

CS49000-VIZ Intro to Data Visualization / Fall 2020; Lecture 16: Dimenzionally Reduction

ⁱ be the *i*-th

|고,

 \sim

Intuition

• smooth manifold is locally close to linear

Method

- such that $||\mathbf{x}_i \sum W_{ij} \mathbf{x}_j||^2$ is minimized
- Find d-dimensional y's that minimize • Solution is obtained as first d eigenvectors of $j \in \mathcal{N}(i)$ y $\sum ||\mathbf{y}_i - \sum$ *i j* W_{ij} y $_j$ ||² $(I - W)^T (I - W)$
-

• Characterize local linearity through weight matrix *W*

CS49000-VIZ Intro to Data

Intuition

• smooth manifold is locally close to linear

Method

- such that $||\mathbf{x}_i \sum W_{ij} \mathbf{x}_j||^2$ is minimized
- Find d-dimensional y's that minimize • Solution is obtained as first d eigenvectors of $j \in \mathcal{N}(i)$ y $\sum ||\mathbf{y}_i - \sum$ *i j* W_{ij} y $_j$ ||² $(I - W)^T (I - W)$
-

• Characterize local linearity through weight matrix *W*

LLE involve local minima. By exploiting the local symmetries of linear reconstructions, LLE is able to learn the global structure of nonlinear manifolds, such as those generated by images of faces or documents of text. coordinates as observed modes of variability. maps the high-dimensional coordinates of the high-dimensional coordinates of the higheach neighborhood to global internal coordinates and coordinates and coordinates and coordinates and coordinates and nates on the manifold. By design, the reconstruction weights *W*ij reflect intrinsic geometric properties of the data that are invariant to the data that are invariant to the data that are in exactly such transformations. We therefore the transformations. We therefore the transformations. We therefore

How do we judge similarity? Our mental

- Graph Laplacian of W is with D diagonal and $D_{ii} = \sum W_{ij}$
- Define W as either binary indicator of connectivity or through heat kernel *W*
- Solve for $\sqrt{ }$ *i,j* $W_{ij} ||y_i - y_j||^2 = tr(Y^T L Y)$
- zero eigenvalue

j

• Compute eigenvectors y of Laplacian associated with non-

Figure 2: Two-dimensional representations of the "swiss roll" data, for different values of the number of nearest neighbors N and the heat kernel parameter t. $t = \infty$ corresponds to the discrete weights.

Laplacian Eigenmaps

•Spectral decomposition intractable

for large problems

- •Approximate method can be used •Nyström method + column
- sampling

