Minimum Neighbor Distance Estimators of Intrinsic Dimension

G. Lombardi, A. Rozza, C. Ceruti, E. Casiraghi, and P. Campadelli
{lombardi,rozza,ceruti,casiraghi,campadelli}@dsi.unimi.it

Università degli Studi di Milano

Outline

Motivation

- Many real life signals are high dimensional, but...
- ...the number of their 'useful' degrees of freedom low;
- often the data are assumed drawn from a low-dimensional manifold mapped in a high dimensional space (plus noise): $\mathbf{x} = \psi(\mathbf{z}) + \nu, \qquad \mathbf{x} \in \Re^D, \mathbf{z} \sim \mathcal{M} \equiv \Re^d, \psi : \Re^d \to \Re^D, \nu \sim \mathcal{N}$

Problem definition

- Consider a dataset $\mathbf{X}_N = \{\mathbf{x}_i = \psi(\mathbf{z}_i)\}_{i=1}^N$ sampled from a manifold $\mathcal{M} \equiv \Re^d$ and embedded in $\Re^D$ through a map ψ;
- assume the $\mathbf{z}_i$ sampled from $\mathcal{M}$ by means of a smooth pdf f;
- assume the embedding defined by ψ to be proper;
- our aim is to estimate the intrinsic dimensionality d of $\mathcal{M}$ by means of the samples $\mathbf{X}_N \subset \Re^D$.

Motivation

- Many real life signals are high dimensional, but...
- ...the number of their 'useful' degrees of freedom low;
- often the data are assumed drawn from a low-dimensional manifold mapped in a high dimensional space (plus noise):
  $\mathbf{x} = \psi(\mathbf{z}) + \nu, \qquad \mathbf{x} \in \Re^D, \mathbf{z} \sim \mathcal{M} \equiv \Re^d, \psi : \Re^d \rightarrow \Re^D, \nu \sim \mathcal{N}$

Problem definition

- Consider a dataset $\mathbf{X}_N = \{\mathbf{x}_i = \psi(\mathbf{z}_i)\}_{i=1}^N$ sampled from a manifold $\mathcal{M} \equiv \Re^d$ and embedded in $\Re^D$ through a map ψ;
- assume the $\mathbf{z}_i$ sampled from $\mathcal{M}$ by means of a smooth pdf f;
- assume the embedding defined by ψ to be proper;
- our aim is to estimate the intrinsic dimensionality d of $\mathcal{M}$ by means of the samples $\mathbf{X}_N \subset \Re^D$.

Motivation

- Many real life signals are high dimensional, but...
- ...the number of their 'useful' degrees of freedom low;
- often the data are assumed drawn from a low-dimensional manifold mapped in a high dimensional space (plus noise):
$\mathbf{x} = \psi(\mathbf{z}) + \nu, \qquad \mathbf{x} \in \Re^D, \mathbf{z} \sim \mathcal{M} \equiv \Re^d, \psi : \Re^d \rightarrow \Re^D, \nu \sim \mathcal{N}$

Problem definition

- Consider a dataset $\mathbf{X}_N = \{\mathbf{x}_i = \psi(\mathbf{z}_i)\}_{i=1}^N$ sampled from a manifold $\mathcal{M} \equiv \Re^d$ and embedded in $\Re^D$ through a map ψ;
- assume the $\mathbf{z}_i$ sampled from $\mathcal{M}$ by means of a smooth pdf f;
- assume the embedding defined by ψ to be proper;
- our aim is to estimate the intrinsic dimensionality d of $\mathcal{M}$ by means of the samples $\mathbf{X}_N \subset \Re^D$.

Motivation

- Many real life signals are high dimensional, but...
- ...the number of their 'useful' degrees of freedom low;
- often the data are assumed drawn from a low-dimensional manifold mapped in a high dimensional space (plus noise):
  $\mathbf{x} = \psi(\mathbf{z}) + \nu, \qquad \mathbf{x} \in \Re^D, \mathbf{z} \sim \mathcal{M} \equiv \Re^d, \psi : \Re^d \to \Re^D, \nu \sim \mathcal{N}$

Problem definition

- Consider a dataset $\mathbf{X}_N = \{\mathbf{x}_i = \psi(\mathbf{z}_i)\}_{i=1}^N$ sampled from a manifold $\mathcal{M} \equiv \Re^d$ and embedded in $\Re^D$ through a map ψ;
- assume the $\mathbf{z}_i$ sampled from $\mathcal{M}$ by means of a smooth pdf f;
- assume the embedding defined by ψ to be proper;
- **our aim** is to estimate the intrinsic dimensionality d of $\mathcal{M}$ by means of the samples $\mathbf{X}_N \subset \Re^D$.

Motivation

- Many real life signals are high dimensional, but...
- ...the number of their 'useful' degrees of freedom low;
- often the data are assumed drawn from a low-dimensional manifold mapped in a high dimensional space (plus noise):

$$\mathbf{x} = \psi(\mathbf{z}) + \nu, \qquad \mathbf{x} \in \Re^D, \mathbf{z} \sim \mathcal{M} \equiv \Re^d, \psi : \Re^d \to \Re^D, \nu \sim \mathcal{N}$$

Problem definition

- Consider a dataset $\mathbf{X}_N = \{\mathbf{x}_i = \psi(\mathbf{z}_i)\}_{i=1}^N$ sampled from a manifold $\mathcal{M} \equiv \Re^d$ and embedded in $\Re^D$ through a map ψ;
- assume the $\mathbf{z}_i$ sampled from $\mathcal{M}$ by means of a smooth `pdf` f;
- assume the embedding defined by ψ to be proper;
- **our aim** is to estimate the intrinsic dimensionality d of $\mathcal{M}$ by means of the samples $\mathbf{X}_N \subset \Re^D$.

Motivation

- Many real life signals are high dimensional, but...
- ...the number of their 'useful' degrees of freedom low;
- often the data are assumed drawn from a low-dimensional manifold mapped in a high dimensional space (plus noise):
  $\mathbf{x} = \psi(\mathbf{z}) + \nu, \qquad \mathbf{x} \in \Re^D, \mathbf{z} \sim \mathcal{M} \equiv \Re^d, \psi : \Re^d \to \Re^D, \nu \sim \mathcal{N}$

Problem definition

- Consider a dataset $\mathbf{X}_N = \{\mathbf{x}_i = \psi(\mathbf{z}_i)\}_{i=1}^N$ sampled from a manifold $\mathcal{M} \equiv \Re^d$ and embedded in $\Re^D$ through a map ψ;
- assume the $\mathbf{z}_i$ sampled from $\mathcal{M}$ by means of a smooth pdf f;
- assume the embedding defined by ψ to be proper;
- **our aim** is to estimate the intrinsic dimensionality d of $\mathcal{M}$ by means of the samples $\mathbf{X}_N \subset \Re^D$.

Motivation

- Many real life signals are high dimensional, but...
- ...the number of their 'useful' degrees of freedom low;
- often the data are assumed drawn from a low-dimensional manifold mapped in a high dimensional space (plus noise):
  $\mathbf{x} = \psi(\mathbf{z}) + \nu, \qquad \mathbf{x} \in \Re^D, \mathbf{z} \sim \mathcal{M} \equiv \Re^d, \psi : \Re^d \to \Re^D, \nu \sim \mathcal{N}$

Problem definition

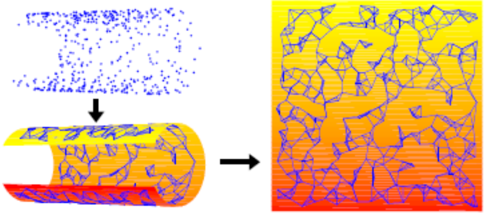
- Consider a dataset $\mathbf{X}_N = \{\mathbf{x}_i = \psi(\mathbf{z}_i)\}_{i=1}^N$ sampled from a manifold $\mathcal{M} \equiv \Re^d$ and embedded in $\Re^D$ through a map ψ;
- assume the $\mathbf{z}_i$ sampled from $\mathcal{M}$ by means of a smooth pdf f;
- assume the embedding defined by ψ to be proper;
- **our aim** is to estimate the intrinsic dimensionality d of $\mathcal{M}$ by means of the samples $\mathbf{X}_N \subset \Re^D$.

Applications

Dimensionality reduction: First step for dimensionality reduction techniques (that generally require d as parameter).

Manifold learning: First step for manifold learning techniques.

Parameter estimation: Estimates the number of eigenvalues to be retained, the number of dimensions for partial whitening algorithms, ...

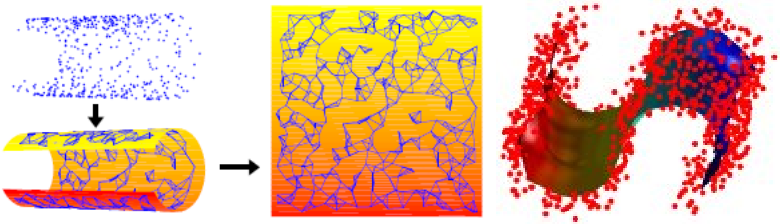


Applications

Dimensionality reduction: First step for dimensionality reduction techniques (that generally require d as parameter).

Manifold learning: First step for manifold learning techniques.

Parameter estimation: Estimates the number of eigenvalues to be retained, the number of dimensions for partial whitening algorithms, ...

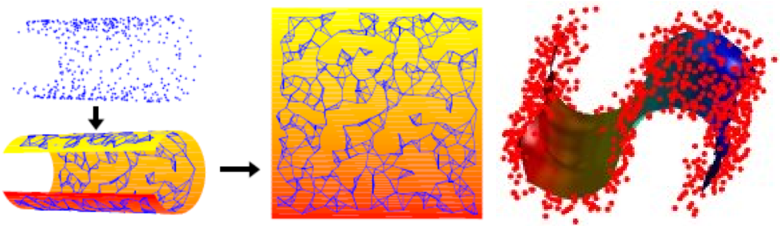


Applications

Dimensionality reduction: First step for dimensionality reduction techniques (that generally require d as parameter).

Manifold learning: First step for manifold learning techniques.

Parameter estimation: Estimates the number of eigenvalues to be retained, the number of dimensions for partial whitening algorithms, . . .



Problems arising with dimensionality

Curse of dimensionality: The number of samples N required for manifold learning grows exponentially with d;

Empty space: If D is high enough, splitting the space with a regular grid leaves most of the 'boxes' empty;

Lack of geometry: If D increases, geometry "disappears" and statistical properties arise; e.g. compression of norms.

Problems arising with dimensionality

Curse of dimensionality: The number of samples N required for manifold learning grows exponentially with d;

Empty space: If D is high enough, splitting the space with a regular grid leaves most of the 'boxes' empty;

Lack of geometry: If D increases, geometry "disappears" and statistical properties arise; e.g. compression of norms.

Problems arising with dimensionality

Curse of dimensionality: The number of samples N required for manifold learning grows exponentially with d;

Empty space: If D is high enough, splitting the space with a regular grid leaves most of the 'boxes' empty;

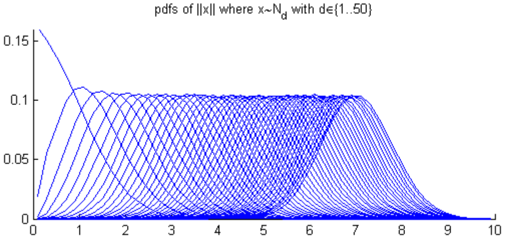
Lack of geometry: If D increases, geometry "disappears" and statistical properties arise; e.g. compression of norms.

Problems arising with dimensionality

Curse of dimensionality: The number of samples N required for manifold learning grows exponentially with d;

Empty space: If D is high enough, splitting the space with a regular grid leaves most of the 'boxes' empty;

Lack of geometry: If D increases, geometry "disappears" and statistical properties arise; e.g. compression of norms.



Dimensionality estimation algorithms

Global/local

Global: The i.d. is estimated for the whole dataset.

Local: The i.d. is estimated for each point.

Linear/nonlinear

Linear: Assumes $\mathcal{M}$ linearly embedded in $\Re^D$.

Nonlinear: Assumes the embedding proper (may be non-linear).

Geometrical/statistical

Geometrical: Uses geometric informations such as tangent space estimation (e.g. Tensor Voting Framework).

Statistical: Uses statistics on measures (e.g. Maximum Likelihood Estimation based on distances).

Dimensionality estimation algorithms

Global/local

Global: The i.d. is estimated for the whole dataset.

Local: The i.d. is estimated for each point.

Linear/nonlinear

Linear: Assumes $\mathcal{M}$ linearly embedded in $\Re^D$.

Nonlinear: Assumes the embedding proper (may be non-linear).

Geometrical/statistical

Geometrical: Uses geometric informations such as tangent space estimation (e.g. Tensor Voting Framework).

Statistical: Uses statistics on measures (e.g. Maximum Likelihood Estimation based on distances).

Dimensionality estimation algorithms

Global/local

Global: The i.d. is estimated for the whole dataset.

Local: The i.d. is estimated for each point.

Linear/nonlinear

Linear: Assumes $\mathcal{M}$ linearly embedded in $\Re^D$.

Nonlinear: Assumes the embedding proper (may be non-linear).

Geometrical/statistical

Geometrical: Uses geometric informations such as tangent space estimation (e.g. Tensor Voting Framework).

Statistical: Uses statistics on measures (e.g. Maximum Likelihood Estimation based on distances).

Some state of the art techniques

PCA: Linear technique based on the estimation of maximal variance directions and thresholding.

kNN Graph: K-Nearest Neighbors Graph based technique, computes $\mathbb{E}\left[L(\mathbf{X})/N^{\alpha}\right]$ where $L(\mathbf{X})$ is a graph length measure, $\alpha = (d' - \gamma)/d'$ $(1 \leq \gamma < d)$, and $\alpha = (d' - \gamma)/d'$; the limit with $N \to \infty$ of this quantity is finite and non-zero only for $d' = d$.

Correlation Dimension: Based on the assumption that the number of samples covered by a sphere with radius r grows proportionally to r^d. An asymptotic smoothed version of this algorithm was proposed by Hein.

Maximum Likelihood Estimation: Based on the maximization of likelihood for the probability distribution of neighboring distances with dependent variable d.

Some state of the art techniques

PCA: Linear technique based on the estimation of maximal variance directions and thresholding.

kNN Graph: K-Nearest Neighbors Graph based technique, computes $\mathbb{E}\left[L(\mathbf{X})/N^{\alpha}\right]$ where $L(\mathbf{X})$ is a graph length measure, $\alpha = (d' - \gamma)/d'$ $(1 \leq \gamma < d)$, and $\alpha = (d' - \gamma)/d'$; the limit with $N \rightarrow \infty$ of this quantity is finite and non-zero only for $d' = d$.

Correlation Dimension: Based on the assumption that the number of samples covered by a sphere with radius r grows proportionally to r^d. An asymptotic smoothed version of this algorithm was proposed by Hein.

Maximum Likelihood Estimation: Based on the maximization of likelihood for the probability distribution of neighboring distances with dependent variable d.

Some state of the art techniques

PCA: Linear technique based on the estimation of maximal variance directions and thresholding.

kNN Graph: K-Nearest Neighbors Graph based technique, computes $\mathbb{E}\left[L(\mathbf{X})/N^{\alpha}\right]$ where $L(\mathbf{X})$ is a graph length measure, $\alpha = (d' - \gamma)/d'$ $(1 \leq \gamma < d)$, and $\alpha = (d' - \gamma)/d'$; the limit with $N \to \infty$ of this quantity is finite and non-zero only for $d' = d$.

Correlation Dimension: Based on the assumption that the number of samples covered by a sphere with radius r grows proportionally to r^d. An asymptotic smoothed version of this algorithm was proposed by Hein.

Maximum Likelihood Estimation: Based on the maximization of likelihood for the probability distribution of neighboring distances with dependent variable d.

Some state of the art techniques

PCA: Linear technique based on the estimation of maximal variance directions and thresholding.

kNN Graph: K-Nearest Neighbors Graph based technique, computes $\mathbb{E}\left[L(\mathbf{X})/N^{\alpha}\right]$ where $L(\mathbf{X})$ is a graph length measure, $\alpha = (d' - \gamma)/d'$ $(1 \leq \gamma < d)$, and $\alpha = (d' - \gamma)/d'$; the limit with $N \to \infty$ of this quantity is finite and non-zero only for $d' = d$.

Correlation Dimension: Based on the assumption that the number of samples covered by a sphere with radius r grows proportionally to r^d. An asymptotic smoothed version of this algorithm was proposed by Hein.

Maximum Likelihood Estimation: Based on the maximization of likelihood for the probability distribution of neighboring distances with dependent variable d.

Some considerations

Statistics about distances

- Statistics are preferable in high dimensional spaces;
- norm compression depends on intrinsic dimensionality;
- the i.d. can be estimated exploiting the norm compression;
- the real pdf is difficult to be estimated, but simulation helps.

Locality

- Can be approximated by the kNN graph;
- consistent local statistics can be defined by means of the normalized k Nearest Neighbors distances;
- given k neighboring points, the closest ones are less affected by the curvature of the manifold $\mathcal{M}$.

Some considerations

Statistics about distances

- Statistics are preferable in high dimensional spaces;
- norm compression depends on intrinsic dimensionality;
- the i.d. can be estimated exploiting the norm compression;
- the real pdf is difficult to be estimated, but simulation helps.

Locality

- Can be approximated by the kNN graph;
- consistent local statistics can be defined by means of the normalized k Nearest Neighbors distances;
- given k neighboring points, the closest ones are less affected by the curvature of the manifold $\mathcal{M}$.

Some considerations

Statistics about distances

- Statistics are preferable in high dimensional spaces;
- norm compression depends on intrinsic dimensionality;
- the i.d. can be estimated exploiting the norm compression;
- the real pdf is difficult to be estimated, but simulation helps.

Locality

- Can be approximated by the kNN graph;
- consistent local statistics can be defined by means of the normalized k Nearest Neighbors distances;
- given k neighboring points, the closest ones are less affected by the curvature of the manifold $\mathcal{M}$.

Some considerations

Statistics about distances

- Statistics are preferable in high dimensional spaces;
- norm compression depends on intrinsic dimensionality;
- the i.d. can be estimated exploiting the norm compression;
- the real pdf is difficult to be estimated, but simulation helps.

Locality

- Can be approximated by the kNN graph;
- consistent local statistics can be defined by means of the normalized k Nearest Neighbors distances;
- given k neighboring points, the closest ones are less affected by the curvature of the manifold $\mathcal{M}$.

Some considerations

Statistics about distances

- Statistics are preferable in high dimensional spaces;
- norm compression depends on intrinsic dimensionality;
- the i.d. can be estimated exploiting the norm compression;
- the real pdf is difficult to be estimated, but simulation helps.

Locality

- Can be approximated by the kNN graph;
- consistent local statistics can be defined by means of the normalized k Nearest Neighbors distances;
- given k neighboring points, the closest ones are less affected by the curvature of the manifold $\mathcal{M}$.

Some considerations

Statistics about distances

- Statistics are preferable in high dimensional spaces;
- norm compression depends on intrinsic dimensionality;
- the i.d. can be estimated exploiting the norm compression;
- the real pdf is difficult to be estimated, but simulation helps.

Locality

- Can be approximated by the kNN graph;
- consistent local statistics can be defined by means of the normalized k Nearest Neighbors distances;
- given k neighboring points, the closest ones are less affected by the curvature of the manifold $\mathcal{M}$.

Some considerations

Statistics about distances

- Statistics are preferable in high dimensional spaces;
- norm compression depends on intrinsic dimensionality;
- the i.d. can be estimated exploiting the norm compression;
- the real pdf is difficult to be estimated, but simulation helps.

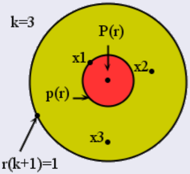
Locality

- Can be approximated by the kNN graph;
- consistent local statistics can be defined by means of the normalized k Nearest Neighbors distances;
- given k neighboring points, the closest ones are less affected by the curvature of the manifold $\mathcal{M}$.

Our approach

Exploited pdf related to distances

- To reduce the bias due to manifold curvature, we extract just the first neighbor distance normalized by the (k+1)-th distance;



- only N distances are available (one per point), but a robust estimator is defined;
- a maximum likelihood solution can be determined.

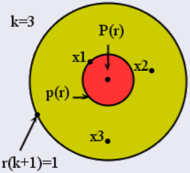
Exploiting the norms compression effect

- real and synthetic pdfs are compared via KL divergence;
- locally uniform distribution is the limit in case of smooth pdf.

Our approach

Exploited pdf related to distances

- To reduce the bias due to manifold curvature, we extract just the first neighbor distance normalized by the (k+1)-th distance;



- only N distances are available (one per point), but a robust estimator is defined;
- a maximum likelihood solution can be determined.

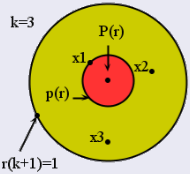
Exploiting the norms compression effect

- real and synthetic pdfs are compared via KL divergence;
- locally uniform distribution is the limit in case of smooth pdf.

Our approach

Exploited pdf related to distances

- To reduce the bias due to manifold curvature, we extract just the first neighbor distance normalized by the (k+1)-th distance;



- only N distances are available (one per point), but a robust estimator is defined;
- a maximum likelihood solution can be determined.

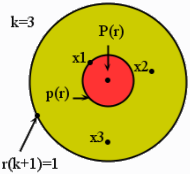
Exploiting the norms compression effect

- real and synthetic pdfs are compared via KL divergence;
- locally uniform distribution is the limit in case of smooth pdf.

Our approach

Exploited pdf related to distances

- To reduce the bias due to manifold curvature, we extract just the first neighbor distance normalized by the (k+1)-th distance;



- only N distances are available (one per point), but a robust estimator is defined;
- a maximum likelihood solution can be determined.

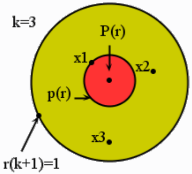
Exploiting the norms compression effect

- real and synthetic pdfs are compared via KL divergence;
- locally uniform distribution is the limit in case of smooth pdf.

Our approach

Exploited pdf related to distances

- To reduce the bias due to manifold curvature, we extract just the first neighbor distance normalized by the (k+1)-th distance;



- only N distances are available (one per point), but a robust estimator is defined;
- a maximum likelihood solution can be determined.

Exploiting the norms compression effect

- real and synthetic pdfs are compared via KL divergence;
- locally uniform distribution is the limit in case of smooth pdf.

Local uniformity

Local pdf

Denoting with $\mathcal{B}_d(\mathbf{0}, 1)$ the unit ball, we define the ϵ-local pdf as:

$$f_\epsilon(\mathbf{z}) = \frac{f(\epsilon\mathbf{z})\chi_{\mathcal{B}_d(\mathbf{0},1)}(\mathbf{z})}{\int_{\mathbf{t}\in\mathcal{B}_d(\mathbf{0},1)} f(\epsilon\mathbf{t})d\mathbf{t}}$$

Theorem 1

Given $\{\epsilon_i\} \to 0^+$, $f_\epsilon(\mathbf{z})$ describes a sequence of pdf having the unit d-dimensional ball as support; such sequence converges uniformly to the uniform distribution $\mathbf{B}_d$ in the ball $\mathcal{B}_d(\mathbf{0}, 1)$.

Local uniformity

Local pdf

Denoting with $\mathcal{B}_d(\mathbf{0},1)$ the unit ball, we define the ϵ-local pdf as:

$$f_\epsilon(\mathbf{z}) = \frac{f(\epsilon\mathbf{z})\chi_{\mathcal{B}_d(\mathbf{0},1)}(\mathbf{z})}{\int_{\mathbf{t}\in\mathcal{B}_d(\mathbf{0},1)} f(\epsilon\mathbf{t})d\mathbf{t}}$$

Theorem 1

Given $\{\epsilon_i\} \to 0^+$, $f_\epsilon(\mathbf{z})$ describes a sequence of pdf having the unit d-dimensional ball as support; such sequence converges uniformly to the uniform distribution $\mathbf{B}_d$ in the ball $\mathcal{B}_d(\mathbf{0},1)$.

A log-likelihood function

First neighbor distance

- Being $V_r = r^d V_1$, the pdf for the first NN distance g is:

$$g(r; k, d) = kdr^{d-1}(1 - r^d)^{k-1}$$

- Given the set $\bar{\mathbf{X}}_{k+1}$ containing the $k+1$ NN of $\mathbf{x}_i$, its normalized minimum neighbor distance is defined as:

$$\rho(\mathbf{x}_i) = \min_{\mathbf{x}_j \in \bar{\mathbf{X}}_{k+1}} \frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{\|\mathbf{x}_i - \hat{\mathbf{x}}\|}, \qquad \hat{\mathbf{x}} = \operatorname*{argmax}_{\mathbf{x} \in \bar{\mathbf{X}}_{k+1}} \|\mathbf{x}_i - \mathbf{x}\|$$

- euclidean distances converge to geodetic ones when $N \to \infty$;
- given the $\mathbf{x}$ smoothly distributed on $\mathcal{M}$, the distribution of ρ converges to $g(r; k, d)$.

A log-likelihood function

First neighbor distance

- Being $V_r = r^d V_1$, the pdf for the first NN distance g is:

$$g(r; k, d) = kdr^{d-1}(1 - r^d)^{k-1}$$

- Given the set $\bar{\mathbf{X}}_{k+1}$ containing the $k+1$ NN of $\mathbf{x}_i$, its normalized minimum neighbor distance is defined as:

$$\rho(\mathbf{x}_i) = \min_{\mathbf{x}_j \in \bar{\mathbf{X}}_{k+1}} \frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{\|\mathbf{x}_i - \hat{\mathbf{x}}\|}, \qquad \hat{\mathbf{x}} = \underset{\mathbf{x} \in \bar{\mathbf{X}}_{k+1}}{\operatorname{argmax}} \|\mathbf{x}_i - \mathbf{x}\|$$

- euclidean distances converge to geodetic ones when $N \to \infty$;
- given the $\mathbf{x}$ smoothly distributed on $\mathcal{M}$, the distribution of ρ converges to $g(r; k, d)$.

A log-likelihood function

First neighbor distance

- Being $V_r = r^d V_1$, the pdf for the first NN distance g is:

$$g(r; k, d) = kdr^{d-1}(1 - r^d)^{k-1}$$

- Given the set $\bar{\mathbf{X}}_{k+1}$ containing the $k+1$ NN of $\mathbf{x}_i$, its normalized minimum neighbor distance is defined as:

$$\rho(\mathbf{x}_i) = \min_{\mathbf{x}_j \in \bar{\mathbf{X}}_{k+1}} \frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{\|\mathbf{x}_i - \hat{\mathbf{x}}\|}, \qquad \hat{\mathbf{x}} = \underset{\mathbf{x} \in \bar{\mathbf{X}}_{k+1}}{\operatorname{argmax}} \|\mathbf{x}_i - \mathbf{x}\|$$

- euclidean distances converge to geodetic ones when $N \to \infty$;
- given the $\mathbf{x}$ smoothly distributed on $\mathcal{M}$, the distribution of ρ converges to $g(r; k, d)$.

A log-likelihood function

First neighbor distance

- Being $V_r = r^d V_1$, the pdf for the first NN distance g is:

$$g(r; k, d) = kdr^{d-1}(1 - r^d)^{k-1}$$

- Given the set $\bar{\mathbf{X}}_{k+1}$ containing the $k+1$ NN of $\mathbf{x}_i$, its normalized minimum neighbor distance is defined as:

$$\rho(\mathbf{x}_i) = \min_{\mathbf{x}_j \in \bar{\mathbf{X}}_{k+1}} \frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{\|\mathbf{x}_i - \hat{\mathbf{x}}\|}, \qquad \hat{\mathbf{x}} = \underset{\mathbf{x} \in \bar{\mathbf{X}}_{k+1}}{\operatorname{argmax}} \|\mathbf{x}_i - \mathbf{x}\|$$

- euclidean distances converge to geodetic ones when $N \to \infty$;
- given the $\mathbf{x}$ smoothly distributed on $\mathcal{M}$, the distribution of ρ converges to $g(r; k, d)$.

Log-likelihood

- Denote with $\tilde{g}(\mathbf{x}_i; k, d)$ the function g applied to $\rho(\mathbf{x}_i)$;
- we compute the log-likelihood $ll(d) = \log(\tilde{g}(\mathbf{x}_i; k, d))$:

$$
\begin{aligned}
ll(d) &= \sum_{\mathbf{x}_i \in \mathbf{X}_N} \log \tilde{g}(\mathbf{x}_i; k, d) = N \log k + N \log d + \\
&\quad (d-1) \sum_{\mathbf{x}_i \in \mathbf{X}_N} \log \rho(\mathbf{x}_i) + (k-1) \sum_{\mathbf{x}_i \in \mathbf{X}_N} \log\left(1 - \rho^d(\mathbf{x}_i)\right)
\end{aligned}
$$

MiND$_{\mathrm{MLk}}$, MiND$_{\mathrm{MLi}}$, MiND$_{\mathrm{ML1}}$

- One estimate for d is obtained solving $\frac{\partial ll}{\partial d} = 0$:

$$
\frac{N}{d} + \sum_{\mathbf{x}_i \in \mathbf{X}_N} \left(\log \rho(\mathbf{x}_i) - (k-1) \frac{\rho^d(\mathbf{x}_i) \log \rho(\mathbf{x}_i)}{1 - \rho^d(\mathbf{x}_i)} \right) = 0
$$

- Notice that choosing $k = 1$, we obtain the MLE algorithm.

Log-likelihood

- Denote with $\tilde{g}(\mathbf{x}_i; k, d)$ the function g applied to $\rho(\mathbf{x}_i)$;
- we compute the log-likelihood $ll(d) = \log(\tilde{g}(\mathbf{x}_i; k, d))$:

$$
\begin{aligned}
ll(d) &= \sum_{\mathbf{x}_i \in \mathbf{X}_N} \log \tilde{g}(\mathbf{x}_i; k, d) = N \log k + N \log d + \\
&\quad (d-1) \sum_{\mathbf{x}_i \in \mathbf{X}_N} \log \rho(\mathbf{x}_i) + (k-1) \sum_{\mathbf{x}_i \in \mathbf{X}_N} \log\left(1 - \rho^d(\mathbf{x}_i)\right)
\end{aligned}
$$

MiND$_{\mathtt{MLk}}$, MiND$_{\mathtt{MLi}}$, MiND$_{\mathtt{ML1}}$

- One estimate for d is obtained solving $\frac{\partial ll}{\partial d} = 0$:

$$
\frac{N}{d} + \sum_{\mathbf{x}_i \in \mathbf{X}_N} \left(\log \rho(\mathbf{x}_i) - (k-1) \frac{\rho^d(\mathbf{x}_i) \log \rho(\mathbf{x}_i)}{1 - \rho^d(\mathbf{x}_i)} \right) = 0
$$

- Notice that choosing $k = 1$, we obtain the MLE algorithm.

Log-likelihood

- Denote with $\tilde{g}(\mathbf{x}_i; k, d)$ the function g applied to $\rho(\mathbf{x}_i)$;
- we compute the log-likelihood $ll(d) = \log(\tilde{g}(\mathbf{x}_i; k, d))$:

$$ll(d) = \sum_{\mathbf{x}_i \in \mathbf{X}_N} \log \tilde{g}(\mathbf{x}_i; k, d) = N \log k + N \log d + (d-1) \sum_{\mathbf{x}_i \in \mathbf{X}_N} \log \rho(\mathbf{x}_i) + (k-1) \sum_{\mathbf{x}_i \in \mathbf{X}_N} \log\left(1 - \rho^d(\mathbf{x}_i)\right)$$

MiND$_{\text{MLk}}$, MiND$_{\text{MLi}}$, MiND$_{\text{ML1}}$

- One estimate for d is obtained solving $\frac{\partial ll}{\partial d} = 0$:

$$\frac{N}{d} + \sum_{\mathbf{x}_i \in \mathbf{X}_N} \left(\log \rho(\mathbf{x}_i) - (k-1) \frac{\rho^d(\mathbf{x}_i) \log \rho(\mathbf{x}_i)}{1 - \rho^d(\mathbf{x}_i)} \right) = 0$$

- Notice that choosing $k = 1$, we obtain the MLE algorithm.

Log-likelihood

- Denote with $\tilde{g}(\mathbf{x}_i; k, d)$ the function g applied to $\rho(\mathbf{x}_i)$;
- we compute the log-likelihood $ll(d) = \log(\tilde{g}(\mathbf{x}_i; k, d))$:

$$ll(d) = \sum_{\mathbf{x}_i \in \mathbf{X}_N} \log \tilde{g}(\mathbf{x}_i; k, d) = N \log k + N \log d + (d-1) \sum_{\mathbf{x}_i \in \mathbf{X}_N} \log \rho(\mathbf{x}_i) + (k-1) \sum_{\mathbf{x}_i \in \mathbf{X}_N} \log\left(1 - \rho^d(\mathbf{x}_i)\right)$$

$\mathtt{MiND}_{\mathtt{MLk}}$, $\mathtt{MiND}_{\mathtt{MLi}}$, $\mathtt{MiND}_{\mathtt{ML1}}$

- One estimate for d is obtained solving $\frac{\partial ll}{\partial d} = 0$:

$$\frac{N}{d} + \sum_{\mathbf{x}_i \in \mathbf{X}_N} \left(\log \rho(\mathbf{x}_i) - (k-1) \frac{\rho^d(\mathbf{x}_i) \log \rho(\mathbf{x}_i)}{1 - \rho^d(\mathbf{x}_i)} \right) = 0$$

- Notice that choosing $k = 1$, we obtain the MLE algorithm.

pdf comparison

- Call $\hat{g}(r;k)$ an estimate of $g(r;k,d)$ computed with $\rho(\mathbf{x}_i)$;
- $\hat{d}$ is obtained maximizing the Kullback-Leibler divergence:

$$\hat{d} = \operatorname*{argmin}_{1 \le d \le D} \int_0^1 \hat{g}(r;k) \log\left(\frac{\hat{g}(r;k)}{g(r;k,d)}\right) dr$$

- we draw N samples from the d-dimensional uniform ball:
$\mathbf{y} = \frac{u^{\frac{1}{d}}}{\|\bar{\mathbf{y}}\|}\bar{\mathbf{y}}, \qquad \bar{\mathbf{y}} \sim \mathcal{N}(\cdot|\mathbf{0}_d, 1), \qquad u \sim U(0,1)$
- we compute ρ over $\mathbf{X}$ and $\mathbf{Y}$ obtaining $\hat{\mathbf{r}}$ and $\check{\mathbf{r}}_d$;
- estimates $\hat{g}$ and $\check{g}_d$ can be computed as follows:

$$\hat{g}(\hat{r}_i;k) = \frac{1/(N-1)}{2\hat{\rho}(\hat{r}_i)} \qquad \check{g}_d(\hat{r}_i;k) = \frac{1/N}{2\check{\rho}_d(\hat{r}_i)}$$

with $\hat{\rho}(\hat{r}_i)$ and $\check{\rho}_d(\hat{r}_i)$ NN distances for $\hat{r}_i$ in $\hat{\mathbf{r}}$ and $\check{\mathbf{r}}_d$.

pdf comparison

- Call $\hat{g}(r;k)$ an estimate of $g(r;k,d)$ computed with $\rho(\mathbf{x}_i)$;
- $\hat{d}$ is obtained maximizing the Kullback-Leibler divergence:

$$\hat{d} = \underset{1 \leq d \leq D}{\operatorname{argmin}} \int_0^1 \hat{g}(r;k) \log \left(\frac{\hat{g}(r;k)}{g(r;k,d)} \right) dr$$

- we draw N samples from the d-dimensional uniform ball:
  $\mathbf{y} = \frac{u^{\frac{1}{d}}}{\|\bar{\mathbf{y}}\|} \bar{\mathbf{y}}, \qquad \bar{\mathbf{y}} \sim \mathcal{N}\left(\cdot | \mathbf{0}_d, 1\right), \qquad u \sim U(0,1)$
- we compute ρ over $\mathbf{X}$ and $\mathbf{Y}$ obtaining $\hat{\mathbf{r}}$ and $\check{\mathbf{r}}_d$;
- estimates $\hat{g}$ and $\check{g}_d$ can be computed as follows:

$$\hat{g}(\hat{r}_i;k) = \frac{1/(N-1)}{2\hat{\rho}(\hat{r}_i)} \qquad \check{g}_d(\hat{r}_i;k) = \frac{1/N}{2\check{\rho}_d(\hat{r}_i)}$$

with $\hat{\rho}(\hat{r}_i)$ and $\check{\rho}_d(\hat{r}_i)$ NN distances for $\hat{r}_i$ in $\hat{\mathbf{r}}$ and $\check{\mathbf{r}}_d$.

pdf comparison

- Call $\hat{g}(r;k)$ an estimate of $g(r;k,d)$ computed with $\rho(\mathbf{x}_i)$;
- $\hat{d}$ is obtained maximizing the Kullback-Leibler divergence:

$$\hat{d} = \underset{1 \le d \le D}{\operatorname{argmin}} \int_0^1 \hat{g}(r;k) \log\left(\frac{\hat{g}(r;k)}{g(r;k,d)}\right) dr$$

- we draw N samples from the d-dimensional uniform ball:

$$\mathbf{y} = \frac{u^{\frac{1}{d}}}{\|\bar{\mathbf{y}}\|}\bar{\mathbf{y}}, \qquad \bar{\mathbf{y}} \sim \mathcal{N}\left(\cdot|\mathbf{0}_d, 1\right), \qquad u \sim U(0,1)$$

- we compute ρ over $\mathbf{X}$ and $\mathbf{Y}$ obtaining $\hat{\mathbf{r}}$ and $\check{\mathbf{r}}_d$;
- estimates $\hat{g}$ and $\check{g}_d$ can be computed as follows:

$$\hat{g}(\hat{r}_i;k) = \frac{1/(N-1)}{2\hat{\rho}(\hat{r}_i)} \qquad \check{g}_d(\hat{r}_i;k) = \frac{1/N}{2\check{\rho}_d(\hat{r}_i)}$$

with $\hat{\rho}(\hat{r}_i)$ and $\check{\rho}_d(\hat{r}_i)$ NN distances for $\hat{r}_i$ in $\hat{\mathbf{r}}$ and $\check{\mathbf{r}}_d$.

pdf comparison

- Call $\hat{g}(r;k)$ an estimate of $g(r;k,d)$ computed with $\rho(\mathbf{x}_i)$;
- $\hat{d}$ is obtained maximizing the Kullback-Leibler divergence:

$$\hat{d} = \underset{1 \le d \le D}{\operatorname{argmin}} \int_0^1 \hat{g}(r;k) \log\left(\frac{\hat{g}(r;k)}{g(r;k,d)}\right) dr$$

- we draw N samples from the d-dimensional uniform ball:

$$\mathbf{y} = \frac{u^{\frac{1}{d}}}{\|\bar{\mathbf{y}}\|}\bar{\mathbf{y}}, \qquad \bar{\mathbf{y}} \sim \mathcal{N}\left(\cdot|\mathbf{0}_d, 1\right), \qquad u \sim U(0,1)$$

- we compute ρ over $\mathbf{X}$ and $\mathbf{Y}$ obtaining $\hat{\mathbf{r}}$ and $\check{\mathbf{r}}_d$;
- estimates $\hat{g}$ and $\check{g}_d$ can be computed as follows:

$$\hat{g}(\hat{r}_i;k) = \frac{1/(N-1)}{2\hat{\rho}(\hat{r}_i)} \qquad \check{g}_d(\hat{r}_i;k) = \frac{1/N}{2\check{\rho}_d(\hat{r}_i)}$$

with $\hat{\rho}(\hat{r}_i)$ and $\check{\rho}_d(\hat{r}_i)$ NN distances for $\hat{r}_i$ in $\hat{\mathbf{r}}$ and $\check{\mathbf{r}}_d$.

pdf comparison

- Call $\hat{g}(r;k)$ an estimate of $g(r;k,d)$ computed with $\rho(\mathbf{x}_i)$;
- $\hat{d}$ is obtained maximizing the Kullback-Leibler divergence:

$$\hat{d} = \underset{1\leq d\leq D}{\operatorname{argmin}} \int_0^1 \hat{g}(r;k)\log\left(\frac{\hat{g}(r;k)}{g(r;k,d)}\right)dr$$

- we draw N samples from the d-dimensional uniform ball:

$$\mathbf{y} = \frac{u^{\frac{1}{d}}}{\|\bar{\mathbf{y}}\|}\bar{\mathbf{y}}, \qquad \bar{\mathbf{y}} \sim \mathcal{N}(\cdot|\mathbf{0}_d, 1), \qquad u \sim U(0,1)$$

- we compute ρ over $\mathbf{X}$ and $\mathbf{Y}$ obtaining $\hat{\mathbf{r}}$ and $\check{\mathbf{r}}_d$;
- estimates $\hat{g}$ and $\check{g}_d$ can be computed as follows:

$$\hat{g}(\hat{r}_i;k) = \frac{1/(N-1)}{2\hat{\rho}(\hat{r}_i)} \qquad \check{g}_d(\hat{r}_i;k) = \frac{1/N}{2\check{\rho}_d(\hat{r}_i)}$$

with $\hat{\rho}(\hat{r}_i)$ and $\check{\rho}_d(\hat{r}_i)$ NN distances for $\hat{r}_i$ in $\hat{\mathbf{r}}$ and $\check{\mathbf{r}}_d$.

MiND$_{\text{KL}}$

- We estimate the KL div by means of the Wang's algorithm[a];
- The estimate of $KL(\hat{g}, \check{g}_d)$ becomes:

$$\hat{KL}(\hat{g}, \check{g}_d) = \frac{1}{N} \sum_{i=1}^{N} \log \frac{\hat{g}(\hat{r}_i; k)}{\check{g}_d(\hat{r}_i; k)}$$

- Using this KL approximation, d can be estimated as:

$$\hat{d} = \underset{d \in \{1..D\}}{\operatorname{argmin}} \left(\log \frac{N}{N-1} + \frac{1}{N} \sum_{i=1}^{N} \log \frac{\hat{\rho}(\hat{r}_i)}{\check{\rho}_d(\hat{r}_i)} \right)$$

- The proposed estimator is consistent, that is $\lim_{N \to \infty} \hat{d} = d$.

[a] "A nearest-neighbor approach to estimating divergence between continuous random vector"

MiND$_{KL}$

- We estimate the KL div by means of the Wang's algorithm[a];
- The estimate of $KL(\hat{g}, \check{g}_d)$ becomes:

$$\hat{KL}(\hat{g}, \check{g}_d) = \frac{1}{N} \sum_{i=1}^{N} \log \frac{\hat{g}(\hat{r}_i; k)}{\check{g}_d(\hat{r}_i; k)}$$

- Using this KL approximation, d can be estimated as:

$$\hat{d} = \operatorname*{argmin}_{d \in \{1..D\}} \left(\log \frac{N}{N-1} + \frac{1}{N} \sum_{i=1}^{N} \log \frac{\hat{\rho}(\hat{r}_i)}{\check{\rho}_d(\hat{r}_i)} \right)$$

- The proposed estimator is consistent, that is $\lim_{N \to \infty} \hat{d} = d$.

[a] "A nearest-neighbor approach to estimating divergence between continuous random vector"

$\texttt{MiND}_{\texttt{KL}}$

- We estimate the KL div by means of the Wang's algorithm[a];
- The estimate of $KL(\hat{g}, \check{g}_d)$ becomes:

$$\hat{KL}(\hat{g}, \check{g}_d) = \frac{1}{N} \sum_{i=1}^{N} \log \frac{\hat{g}(\hat{r}_i; k)}{\check{g}_d(\hat{r}_i; k)}$$

- Using this KL approximation, d can be estimated as:

$$\hat{d} = \underset{d \in \{1..D\}}{\operatorname{argmin}} \left(\log \frac{N}{N-1} + \frac{1}{N} \sum_{i=1}^{N} \log \frac{\hat{\rho}(\hat{r}_i)}{\check{\rho}_d(\hat{r}_i)} \right)$$

- The proposed estimator is consistent, that is $\lim_{N \to \infty} \hat{d} = d$.

[a] "A nearest-neighbor approach to estimating divergence between continuous random vector"

MiND$_{\mathrm{KL}}$

- We estimate the KL div by means of the Wang's algorithm[a];
- The estimate of $KL(\hat{g}, \check{g}_d)$ becomes:

$$\hat{KL}(\hat{g}, \check{g}_d) = \frac{1}{N} \sum_{i=1}^{N} \log \frac{\hat{g}(\hat{r}_i; k)}{\check{g}_d(\hat{r}_i; k)}$$

- Using this KL approximation, d can be estimated as:

$$\hat{d} = \underset{d \in \{1..D\}}{\operatorname{argmin}} \left(\log \frac{N}{N-1} + \frac{1}{N} \sum_{i=1}^{N} \log \frac{\hat{\rho}(\hat{r}_i)}{\check{\rho}_d(\hat{r}_i)} \right)$$

- The proposed estimator is consistent, that is $\lim_{N \to \infty} \hat{d} = d$.

[a] "A nearest-neighbor approach to estimating divergence between continuous random vector"

- Tests were performed on both synthetic and real datasets;
- the Hein's generator[a] was used for the synthetic datasets;
- the real datasets are ISOMAP, MNIST, and Santa Fe.

[a] "Intrinsic dimensionality estimation of submanifolds in Euclidean space"

Dataset	Name	d	D	Description
Syntethic	$\mathcal{M}_1$	10	11	Uniformly sampled sphere linearly embedded.
	$\mathcal{M}_2$	3	5	Affine space.
	$\mathcal{M}_3$	4	6	Concentrated figure, confusable with a 3d one.
	$\mathcal{M}_4$	4	8	Non-linear manifold.
	$\mathcal{M}_5$	2	3	2-d Helix
	$\mathcal{M}_6$	6	36	Non-linear manifold.
	$\mathcal{M}_7$	2	3	Swiss-Roll.
	$\mathcal{M}_8$	12	72	Non-linear manifold.
	$\mathcal{M}_9$	20	20	Affine space.
	$\mathcal{M}_{10a}$	10	11	Uniformly sampled hypercube.
	$\mathcal{M}_{10b}$	17	18	Uniformly sampled hypercube.
	$\mathcal{M}_{10c}$	24	25	Uniformly sampled hypercube.
	$\mathcal{M}_{11}$	2	3	Möebius band 10-times twisted.
	$\mathcal{M}_{12}$	20	20	Isotropic multivariate Gaussian.
	$\mathcal{M}_{13}$	1	13	Curve.
Real	$\mathcal{M}_{\mathtt{Faces}}$	3	4096	ISOMAP face dataset.
	$\mathcal{M}_{\mathtt{MNIST1}}$	8 — 11	784	MNIST database (digit 1).
	$\mathcal{M}_{\mathtt{SantaFe}}$	9	50	Santa Fe dataset (version $D2$).

- Tests were performed on both synthetic and real datasets;
- the Hein's generator[a] was used for the synthetic datasets;
- the real datasets are ISOMAP, MNIST, and Santa Fe.

[a] "Intrinsic dimensionality estimation of submanifolds in Euclidean space"

Dataset	Name	d	D	Description
Syntethic	$\mathcal{M}_1$	10	11	Uniformly sampled sphere linearly embedded.
	$\mathcal{M}_2$	3	5	Affine space.
	$\mathcal{M}_3$	4	6	Concentrated figure, confusable with a 3*d* one.
	$\mathcal{M}_4$	4	8	Non-linear manifold.
	$\mathcal{M}_5$	2	3	2-d Helix
	$\mathcal{M}_6$	6	36	Non-linear manifold.
	$\mathcal{M}_7$	2	3	Swiss-Roll.
	$\mathcal{M}_8$	12	72	Non-linear manifold.
	$\mathcal{M}_9$	20	20	Affine space.
	$\mathcal{M}_{10a}$	10	11	Uniformly sampled hypercube.
	$\mathcal{M}_{10b}$	17	18	Uniformly sampled hypercube.
	$\mathcal{M}_{10c}$	24	25	Uniformly sampled hypercube.
	$\mathcal{M}_{11}$	2	3	Möebius band 10-times twisted.
	$\mathcal{M}_{12}$	20	20	Isotropic multivariate Gaussian.
	$\mathcal{M}_{13}$	1	13	Curve.
Real	$\mathcal{M}_{\texttt{Faces}}$	3	4096	ISOMAP face dataset.
	$\mathcal{M}_{\texttt{MNIST1}}$	8 — 11	784	MNIST database (digit 1).
	$\mathcal{M}_{\texttt{SantaFe}}$	9	50	Santa Fe dataset (version *D2*).

- Tests were performed on both synthetic and real datasets;
- the Hein's generator[a] was used for the synthetic datasets;
- the real datasets are ISOMAP, MNIST, and Santa Fe.

[a] "Intrinsic dimensionality estimation of submanifolds in Euclidean space"

Dataset	Name	d	D	Description
Syntethic	$\mathcal{M}_1$	10	11	Uniformly sampled sphere linearly embedded.
	$\mathcal{M}_2$	3	5	Affine space.
	$\mathcal{M}_3$	4	6	Concentrated figure, confusable with a 3d one.
	$\mathcal{M}_4$	4	8	Non-linear manifold.
	$\mathcal{M}_5$	2	3	2-d Helix
	$\mathcal{M}_6$	6	36	Non-linear manifold.
	$\mathcal{M}_7$	2	3	Swiss-Roll.
	$\mathcal{M}_8$	12	72	Non-linear manifold.
	$\mathcal{M}_9$	20	20	Affine space.
	$\mathcal{M}_{10a}$	10	11	Uniformly sampled hypercube.
	$\mathcal{M}_{10b}$	17	18	Uniformly sampled hypercube.
	$\mathcal{M}_{10c}$	24	25	Uniformly sampled hypercube.
	$\mathcal{M}_{11}$	2	3	Möebius band 10-times twisted.
	$\mathcal{M}_{12}$	20	20	Isotropic multivariate Gaussian.
	$\mathcal{M}_{13}$	1	13	Curve.
Real	$\mathcal{M}_{\texttt{Faces}}$	3	4096	ISOMAP face dataset.
	$\mathcal{M}_{\texttt{MNIST1}}$	8 — 11	784	MNIST database (digit 1).
	$\mathcal{M}_{\texttt{SantaFe}}$	9	50	Santa Fe dataset (version $D2$).

Experimental Setting

Algorithms comparison

- State-of-the-art techniques and our algorithms were tested;
- The following parameters were used for testing:

Method	Synthetic	Real
PCA	*Threshold* $= 0.025$	*Threshold* $= 0.0025$
CD	*None*	*None*
MLE	$k_1 = 6\ k_2 = 20$	$k_1 = 3\ k_2 = 8$
kNNG$_1$	$k_1 = 6,\ k_2 = 20,\ \gamma = 1,\ M = 1,\ N = 10$	$k_1 = 3,\ k_2 = 8,\ \gamma = 1,\ M = 1,\ N = 10$
kNNG$_2$	$k_1 = 6,\ k_2 = 20,\ \gamma = 1,\ M = 10,\ N = 1$	$k_1 = 3,\ k_2 = 8,\ \gamma = 1,\ M = 10,\ N = 1$
MiND$_{\mathrm{ML1}}$	$k = 1$	$k = 1$
MiND$_{\mathrm{MLk}}$	$k = 10$	$k = 5$
MiND$_{\mathrm{MLi}}$	$k = 10$	$k = 5$
MiND$_{\mathrm{KL}}$	$k = 10$	$k = 5$

- For comparison we computed the Mean Percentage Error:

$$\mathrm{MPE} = \frac{100}{\#\mathcal{M}} \sum_{\mathcal{M}} \frac{|\hat{d}_{\mathcal{M}} - d_{\mathcal{M}}|}{d_{\mathcal{M}}}$$

Experimental Setting

Algorithms comparison

- State-of-the-art techniques and our algorithms were tested;
- The following parameters were used for testing:

Method	Synthetic	Real
PCA	*Threshold* $= 0.025$	*Threshold* $= 0.0025$
CD	*None*	*None*
MLE	$k_1 = 6\ k_2 = 20$	$k_1 = 3\ k_2 = 8$
kNNG$_1$	$k_1 = 6,\ k_2 = 20,\ \gamma = 1,\ M = 1,\ N = 10$	$k_1 = 3,\ k_2 = 8,\ \gamma = 1,\ M = 1,\ N = 10$
kNNG$_2$	$k_1 = 6,\ k_2 = 20,\ \gamma = 1,\ M = 10,\ N = 1$	$k_1 = 3,\ k_2 = 8,\ \gamma = 1,\ M = 10,\ N = 1$
MiND$_{\text{ML1}}$	$k = 1$	$k = 1$
MiND$_{\text{MLk}}$	$k = 10$	$k = 5$
MiND$_{\text{MLi}}$	$k = 10$	$k = 5$
MiND$_{\text{KL}}$	$k = 10$	$k = 5$

- For comparison we computed the Mean Percentage Error:

$$\mathrm{MPE} = \frac{100}{\#\mathcal{M}} \sum_{\mathcal{M}} \frac{|\hat{d}_{\mathcal{M}} - d_{\mathcal{M}}|}{d_{\mathcal{M}}}$$

Results

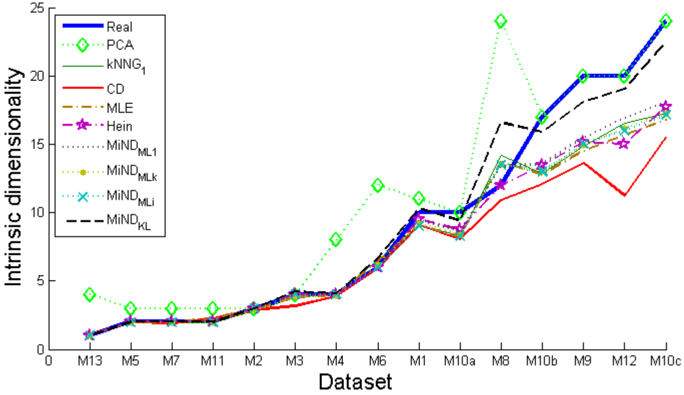
Synthetic datasets

Dataset	d	PCA	kNNG_1	kNNG_2	CD	MLE	Hein	$\mathrm{MiND}_{\mathrm{ML1}}$	$\mathrm{MiND}_{\mathrm{MLk}}$	$\mathrm{MiND}_{\mathrm{MLi}}$	$\mathrm{MiND}_{\mathrm{KL}}$
$\mathcal{M}_{13}$	*1*	4.00	**1.00**	1.01	1.07	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
$\mathcal{M}_{5}$	*2*	3.00	1.96	**2.00**	1.98	1.96	**2.00**	1.97	1.97	**2.00**	**2.00**
$\mathcal{M}_{7}$	*2*	3.00	1.93	1.98	1.94	1.97	**2.00**	1.98	1.96	**2.00**	**2.00**
$\mathcal{M}_{11}$	*2*	3.00	1.96	2.01	2.23	2.30	**2.00**	1.97	1.97	**2.00**	**2.00**
$\mathcal{M}_{2}$	*3*	**3.00**	2.85	2.93	2.88	2.87	**3.00**	2.93	2.88	**3.00**	**3.00**
$\mathcal{M}_{3}$	*4*	**4.00**	3.80	4.22	3.16	3.82	**4.00**	3.89	3.84	**4.00**	4.25
$\mathcal{M}_{4}$	*4*	8.00	4.08	4.06	3.85	3.98	**4.00**	3.95	3.93	**4.00**	4.10
$\mathcal{M}_{6}$	*6*	12.00	6.53	13.99	5.91	6.45	5.95	5.91	6.17	**6.00**	6.65
$\mathcal{M}_{1}$	*10*	11.00	9.07	9.39	9.09	9.06	9.50	9.41	9.23	9.00	**10.30**
$\mathcal{M}_{10a}$	*10*	**10.00**	8.35	9.00	8.04	8.22	8.75	8.68	8.38	8.25	9.40
$\mathcal{M}_{8}$	*12*	24.00	14.19	8.29	10.91	13.69	**12.00**	13.35	13.53	13.50	16.60
$\mathcal{M}_{10b}$	*17*	**17.00**	12.85	15.58	12.09	12.77	13.45	13.59	13.02	13.00	15.90
$\mathcal{M}_{9}$	*20*	**20.00**	14.87	17.07	13.60	14.54	15.15	15.49	14.90	15.00	18.10
$\mathcal{M}_{12}$	*20*	**20.00**	16.50	14.58	11.24	15.67	15.00	16.91	16.19	16.00	19.05
$\mathcal{M}_{10c}$	*24*	**24.00**	17.26	23.68	15.48	16.80	17.70	18.10	17.24	17.15	22.50
MPE		50.67	11.20	16.23	15.38	12.03	7.65	8.32	10.02	9.14	**6.26**

Real datasets

Dataset	d	PCA	kNNG_1	kNNG_2	CD	MLE	Hein	$\mathrm{MiND}_{\mathrm{ML1}}$	$\mathrm{MiND}_{\mathrm{MLk}}$	$\mathrm{MiND}_{\mathrm{MLi}}$	$\mathrm{MiND}_{\mathrm{KL}}$
$\mathcal{M}_{\mathrm{Faces}}$	*3*	21.00	3.60	4.32	3.37	4.05	**3.00**	3.52	3.59	4.00	3.90
$\mathcal{M}_{\mathrm{MNIST1}}$	*8-11*	11.80	10.37	9.58	6.96	10.29	8.00	11.33	10.02	**9.45**	11.00
$\mathcal{M}_{\mathrm{Santa\ Fe}}$	*9*	18.00	7.28	7.43	4.39	7.16	6.00	6.31	6.78	7.00	**7.60**

Results


Real
PCA
kNNG1
CD
MLE
Hein
MiNDML1
MiNDMLk
MiNDMLi
MiNDKL
Intrinsic dimensionality
Dataset
0
5
10
15
20
25
M13
M5
M7
M11
M2
M3
M4
M6
M1
M10a
M8
M10b
M9
M12
M10c

Conclusions

- To estimate the i.d. is a difficult task in case of small sample size, high dimension, and non-linearly embedded manifolds;
- statistic-based techniques are largely adopted for this purpose;
- we propose novel algorithms for the estimation of the i.d.;
- our algorithms are robust to the choice of k and to the high dimensionality of the datasets.

Future Works

- Relax the assumption of smoothness for the pdf f;
- define a local estimator, useful for multi-manifold learning problems having different intrinsic dimensions.

Conclusions

- To estimate the i.d. is a difficult task in case of small sample size, high dimension, and non-linearly embedded manifolds;
- statistic-based techniques are largely adopted for this purpose;
- we propose novel algorithms for the estimation of the i.d.;
- our algorithms are robust to the choice of k and to the high dimensionality of the datasets.

Future Works

- Relax the assumption of smoothness for the pdf f;
- define a local estimator, useful for multi-manifold learning problems having different intrinsic dimensions.

Conclusions

- To estimate the i.d. is a difficult task in case of small sample size, high dimension, and non-linearly embedded manifolds;
- statistic-based techniques are largely adopted for this purpose;
- we propose novel algorithms for the estimation of the i.d.;
- our algorithms are robust to the choice of k and to the high dimensionality of the datasets.

Future Works

- Relax the assumption of smoothness for the pdf f;
- define a local estimator, useful for multi-manifold learning problems having different intrinsic dimensions.

Conclusions

- To estimate the i.d. is a difficult task in case of small sample size, high dimension, and non-linearly embedded manifolds;
- statistic-based techniques are largely adopted for this purpose;
- we propose novel algorithms for the estimation of the i.d.;
- our algorithms are robust to the choice of k and to the high dimensionality of the datasets.

Future Works

- Relax the assumption of smoothness for the pdf f;
- define a local estimator, useful for multi-manifold learning problems having different intrinsic dimensions.

Conclusions

- To estimate the i.d. is a difficult task in case of small sample size, high dimension, and non-linearly embedded manifolds;
- statistic-based techniques are largely adopted for this purpose;
- we propose novel algorithms for the estimation of the i.d.;
- our algorithms are robust to the choice of k and to the high dimensionality of the datasets.

Future Works

- Relax the assumption of smoothness for the `pdf` f;
- define a local estimator, useful for multi-manifold learning problems having different intrinsic dimensions.

Conclusions

- To estimate the i.d. is a difficult task in case of small sample size, high dimension, and non-linearly embedded manifolds;
- statistic-based techniques are largely adopted for this purpose;
- we propose novel algorithms for the estimation of the i.d.;
- our algorithms are robust to the choice of k and to the high dimensionality of the datasets.

Future Works

- Relax the assumption of smoothness for the pdf f;
- define a local estimator, useful for multi-manifold learning problems having different intrinsic dimensions.

Any questions?