Asymptotic Properties of Nonparametric Estimation on Manifold

Yury Yanovich

Institute for Information Transmission Problems
(Russian Academy of Sciences), Moscow
Higher School of Economics — National Research University, Moscow

Overview

1. Nonparametric Estimation and Manifolds
2. Manifold Learning
3. Data Model and Problem Statement
4. Results

Nonparametric Regression (univariate output)

Unknown function

$$f : \mathbb{R}^p \to \mathbb{R}^1, \quad p \geq 1.$$

Nonparametric Regression (univariate output)

Unknown function

$$f : \mathbb{R}^p \to \mathbb{R}^1, \quad p \geq 1.$$

Given a finite sample

$$\mathbb{X}_N = \{X_1, \ldots, X_N\} \subset \mathbb{M} \subset \mathbb{R}^p,$$

where $\mathbb{M}$ is a compact set,

Nonparametric Regression (univariate output)

Unknown function

$$f : \mathbb{R}^p \to \mathbb{R}^1, \quad p \geq 1.$$

Given a finite sample

$$\mathbb{X}_N = \{X_1, \ldots, X_N\} \subset \mathbb{M} \subset \mathbb{R}^p,$$

where $\mathbb{M}$ is a compact set, and a set of (probably noisy) values

$$\mathbb{Y}_N = \{y_1 = f(X_1), \ldots, y_N = f(X_N)\} \subset \mathbb{R}^1$$

Nonparametric Regression (univariate output)

Unknown function

$$f : \mathbb{R}^p \to \mathbb{R}^1, \quad p \geq 1.$$

Given a finite sample

$$\mathbb{X}_N = \{X_1, \ldots, X_N\} \subset \mathbb{M} \subset \mathbb{R}^p,$$

where $\mathbb{M}$ is a compact set, and a set of (probably noisy) values

$$\mathbb{Y}_N = \{y_1 = f(X_1), \ldots, y_N = f(X_N)\} \subset \mathbb{R}^1$$

estimate f for all $X \in \mathbb{M}$.

Nonparametric Regression (univariate output)

Unknown function

$$f: \mathbb{R}^p \to \mathbb{R}^1, \quad p \geq 1.$$

Given a finite sample

$$\mathbb{X}_N = \{X_1, \dots, X_N\} \subset \mathbb{M} \subset \mathbb{R}^p,$$

where $\mathbb{M}$ is a compact set, and a set of (probably noisy) values

$$\mathbb{Y}_N = \{y_1 = f(X_1), \dots, y_N = f(X_N)\} \subset \mathbb{R}^1$$

estimate f for all $X \in \mathbb{M}$.

Kernel nonparametric estimator

$$\hat{f}(X) = \frac{\sum_{n=1}^{N} K_N(X, X_n) f(X_n)}{\sum_{n=1}^{N} K_N(X, X_n)},$$

where $K_N(X, X')$ is a weight (kernel) function.

Nonparametric Regression-2

Classic result (for example, [Henderson&Parmeter2015]):

- mean absolute error $\sim N^{-\frac{2}{p+4}}$.

Nonparametric Regression-2

Classic result (for example, [Henderson&Parmeter2015]):

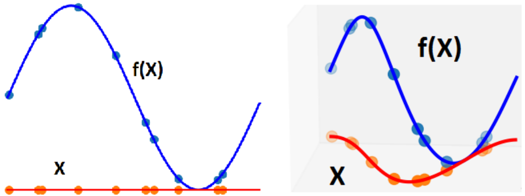
- mean absolute error $\sim N^{-\frac{2}{p+4}}$.

Manifold Assumption [Seung&Lee2000]: data $\mathbb{X}_N$ lie on or near an unknown manifold $\mathbb{M}$ of lower dimensionality $q < p$ embedded in an ambient high-dimensional input space $\mathbb{R}^p$.

Classic result (for example, [Henderson&Parmeter2015]):

- mean absolute error $\sim N^{-\frac{2}{p+4}}$.

Manifold Assumption [Seung&Lee2000]: data $\mathbb{X}_N$ lie on or near an unknown manifold $\mathbb{M}$ of lower dimensionality $q < p$ embedded in an ambient high-dimensional input space $\mathbb{R}^p$.

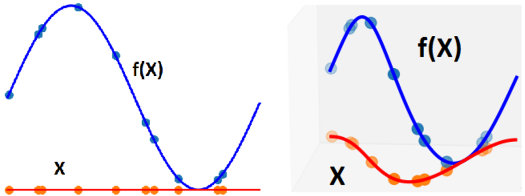


Nonparametric Regression-2

Classic result (for example, [Henderson&Parmeter2015]):

- mean absolute error $\sim N^{-\frac{2}{p+4}}$.

Manifold Assumption [Seung&Lee2000]: data $\mathbb{X}_N$ lie on or near an unknown manifold $\mathbb{M}$ of lower dimensionality $q < p$ embedded in an ambient high-dimensional input space $\mathbb{R}^p$.



Wish: replace $N^{-\frac{2}{p+4}}$ with $N^{-\frac{2}{q+4}}$.

Manifold learning

- **Manifold learning (ML)**: extracting low-dimensional structure of an unknown q-dimensional nonlinear manifold $\mathbb{M}$ from a given dataset $\mathbb{X}_N = \{X_1, X_2, \ldots, X_N\}$ sampled from the $\mathbb{M}$.

Manifold learning

- **Manifold learning (ML)**: extracting low-dimensional structure of an unknown q-dimensional nonlinear manifold $\mathbb{M}$ from a given dataset $\mathbb{X}_N = \{X_1, X_2, \dots, X_N\}$ sampled from the $\mathbb{M}$.
- Given an dataset $\mathbb{X}_N$ to construct an **ML**-solution $\theta = (h, g)$ consisting of

Manifold learning

- **Manifold learning (ML)**: extracting low-dimensional structure of an unknown q-dimensional nonlinear manifold $\mathbb{M}$ from a given dataset $\mathbb{X}_N = \{X_1, X_2, \ldots, X_N\}$ sampled from the $\mathbb{M}$.
- Given an dataset $\mathbb{X}_N$ to construct an **ML**-solution $\theta = (h, g)$ consisting of
  - **Embedding mapping**

    $$h\colon\ \mathbb{M} \to \mathbb{R}^q$$

    and $y = h(X),\ X \in \mathbb{M}$ is a "faithful" representation of X

Manifold learning

- **Manifold learning (ML)**: extracting low-dimensional structure of an unknown q-dimensional nonlinear manifold $\mathbb{M}$ from a given dataset $\mathbb{X}_N = \{X_1, X_2, \ldots, X_N\}$ sampled from the $\mathbb{M}$.
- Given an dataset $\mathbb{X}_N$ to construct an **ML**-solution $\theta = (h, g)$ consisting of
  - **Embedding mapping**

    $$h\colon\ \mathbb{M} \to \mathbb{R}^q$$

    and $y = h(X),\ X \in \mathbb{M}$ is a "faithful" representation of X
  - **Reconstruction mapping** on $\mathbb{Y}_\theta = h(\mathbb{M})$

    $$g\colon \mathbb{Y}_\theta \subset \mathbb{R}^q \to \mathbb{R}^p,$$

Manifold learning

- **Manifold learning (ML)**: extracting low-dimensional structure of an unknown q-dimensional nonlinear manifold $\mathbb{M}$ from a given dataset $\mathbb{X}_N = \{X_1, X_2, \ldots, X_N\}$ sampled from the $\mathbb{M}$.
- Given an dataset $\mathbb{X}_N$ to construct an **ML**-solution $\theta = (h, g)$ consisting of
  - **Embedding mapping**

    $$h\colon\ \mathbb{M} \to \mathbb{R}^q$$

    and $y = h(X),\ X \in \mathbb{M}$ is a "faithful" representation of X
  - **Reconstruction mapping** on $\mathbb{Y}_\theta = h(\mathbb{M})$

    $$g\colon \mathbb{Y}_\theta \subset \mathbb{R}^q \to \mathbb{R}^p,$$

  which ensures the approximate equality

$$X \approx r_\theta(X) \equiv g(h(X)) \quad \text{for all} \quad X \in \mathbb{M}.$$

Manifold learning

- **Manifold learning (ML)**: extracting low-dimensional structure of an unknown q-dimensional nonlinear manifold $\mathbb{M}$ from a given dataset $\mathbb{X}_N = \{X_1, X_2, \dots, X_N\}$ sampled from the $\mathbb{M}$.
- Given an dataset $\mathbb{X}_N$ to construct an **ML**-solution $\theta = (h, g)$ consisting of
  - **Embedding mapping**

    $$h: \ \mathbb{M} \to \mathbb{R}^q$$

    and $y = h(X), \ X \in \mathbb{M}$ is a "faithful" representation of X
  - **Reconstruction mapping** on $\mathbb{Y}_\theta = h(\mathbb{M})$

    $$g: \mathbb{Y}_\theta \subset \mathbb{R}^q \to \mathbb{R}^p,$$

  which ensures the approximate equality

  $$X \approx r_\theta(X) \equiv g(h(X)) \quad \text{for all} \quad X \in \mathbb{M}.$$

- cost function in the **ML**: $\delta_\theta(X) = |X - g(h(X))|_p$.

Common scheme of the ML algorithms

First step: neighborhoods construction
For each $X_n \in \mathbb{X}_N$ point $X_i \in \mathbb{X}_N$ is in the neighborhood iff $|X_i - X_n|_p < \varepsilon$, where $\varepsilon = \varepsilon(N) > 0$.

Common scheme of the ML algorithms

First step: neighborhoods construction
For each $X_n \in \mathbb{X}_N$ point $X_i \in \mathbb{X}_N$ is in the neighborhood iff $|X_i - X_n|_p < \varepsilon$, where $\varepsilon = \varepsilon(N) > 0$.

Second step: neighborhoods descriptions Chosen descriptions of the neighborhoods (local descriptions of the DM) are computed.

Common scheme of the ML algorithms

First step: neighborhoods construction
For each $X_n \in \mathbb{X}_N$ point $X_i \in \mathbb{X}_N$ is in the neighborhood iff $|X_i - X_n|_p < \varepsilon$, where $\varepsilon = \varepsilon(N) > 0$.

Second step: neighborhoods descriptions Chosen descriptions of the neighborhoods (local descriptions of the DM) are computed. Examples:

- barycentric coordinates of the "central" point X_n with respect to its nearest-neighbors (except itself);
- local PCA basis via local covariance

$$\Sigma(X) = \sum_{n=1}^{N} I(|X - X_n|_p < \varepsilon) \cdot (X_n - X) \cdot (X_n - X)^T.$$

Common scheme of the ML algorithms

Third step: global description

Chosen global description of the DM is computed by solving some convex optimization problems under some normalization constraints. Usually low-dimensional sample features $\mathbb{Y}_N = h(\mathbb{X}_N) = \{y_1, \ldots, y_N\} \subset \mathbb{Y}_h$ are computed by minimization of chosen cost function $L(\mathbb{Y}_N|\mathbb{X}_N)$ over $\mathbb{Y}_N$.

Third step: global description

Chosen global description of the DM is computed by solving some convex optimization problems under some normalization constraints. Usually low-dimensional sample features $\mathbb{Y}_N = h(\mathbb{X}_N) = \{y_1, \ldots, y_N\} \subset \mathbb{Y}_h$ are computed by minimization of chosen cost function $L(\mathbb{Y}_N|\mathbb{X}_N)$ over $\mathbb{Y}_N$. Examples:

$$L_{LLE}(\mathbb{Y}_N|\mathbb{X}_N) = \sum_{n=1}^{N} \|y_n - \sum_{j=1} w_{n,j} y_{n,j}\|_F^2;$$

$$L_{LE}(\mathbb{Y}_N|\mathbb{X}_N) = \sum_{n,j=1}^{N} K(X_n, X_j) \cdot \|y_n - y_j\|_2^2;$$

$$L_{LTSA}(\mathbb{Y}_N|\mathbb{X}_N) =$$

$$= \sum_{j=1}^{N} \|(I_q - Q_{PCA}(X_n) \cdot Q_{PCA}^T(X_n)) \cdot H_n \cdot Y_{(n)}\|_F^2.$$

Fourth step: Out-of-Sample extension
The Feature sample $\mathbb{Y}_N$ gives the values of the Embedding mapping $h(X)$ only at sample points; a finding of low-dimensional features $h(X)$ for Out-of-Sample (OoS) points $X \in \mathbb{M} \setminus \mathbb{X}_N$ is usually called OoS-extension problem.

Data Model. Manifold

- $\mathbb{M} \subset \mathbb{R}^p$ is an unknown "good" q-dimensional manifold covered with a single (or finite set) map $(g, \mathbb{B})$, i.e. $\mathbb{M} = \{\nu(b) | b \in \mathbb{B} \subset \mathbb{R}^q\}$;

Data Model. Manifold

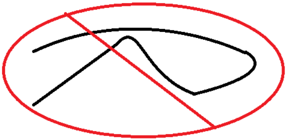
- $\mathbb{M} \subset \mathbb{R}^p$ is an unknown "good" q-dimensional manifold covered with a single (or finite set) map $(g, \mathbb{B})$, i.e. $\mathbb{M} = \{\nu(b) | b \in \mathbb{B} \subset \mathbb{R}^q\}$;
- Closure $\overline{\mathbb{M}}$ of $\mathbb{M}$ is compact;

Data Model. Manifold

- $\mathbb{M} \subset \mathbb{R}^p$ is an unknown "good" q-dimensional manifold covered with a single (or finite set) map $(g, \mathbb{B})$, i.e. $\mathbb{M} = \{\nu(b) | b \in \mathbb{B} \subset \mathbb{R}^q\}$;
- Closure $\overline{\mathbb{M}}$ of $\mathbb{M}$ is compact;
- $\nu \in C^3(\overline{\mathbb{B}})$;

Data Model. Manifold

- $\mathbb{M} \subset \mathbb{R}^p$ is an unknown "good" q-dimensional manifold covered with a single (or finite set) map $(g, \mathbb{B})$, i.e. $\mathbb{M} = \{\nu(b) | b \in \mathbb{B} \subset \mathbb{R}^q\}$;
- Closure $\overline{\mathbb{M}}$ of $\mathbb{M}$ is compact;
- $\nu \in C^3(\overline{\mathbb{B}})$;
- no short circuits



Data Model. Sampling and Function

Sampling

- $\mathbb{X}_N = \{X_1, \ldots, X_N\}$ is an i.i.d. sample from continuous measure $\mu(\mathbb{M})$;

Data Model. Sampling and Function

Sampling

- $\mathbb{X}_N = \{X_1, \ldots, X_N\}$ is an i.i.d. sample from continuous measure $\mu(\mathbb{M})$;
- Density p_μ is bounded from 0 and ∞ on $\mathbb{M}$;

Data Model. Sampling and Function

Sampling

- $\mathbb{X}_N = \{X_1, \ldots, X_N\}$ is an i.i.d. sample from continuous measure $\mu(\mathbb{M})$;
- Density p_μ is bounded from 0 and ∞ on $\mathbb{M}$;
- Values $\mathbb{Y}_N$ are without noise.

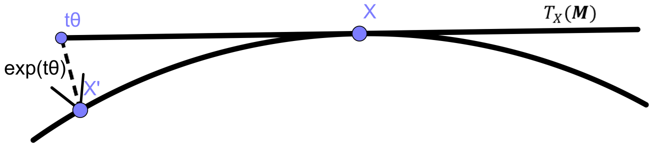
Sampling

- $\mathbb{X}_N = \{X_1, \ldots, X_N\}$ is an i.i.d. sample from continuous measure $\mu(\mathbb{M})$;
- Density p_μ is bounded from 0 and ∞ on $\mathbb{M}$;
- Values $\mathbb{Y}_N$ are without noise.

Function assumption

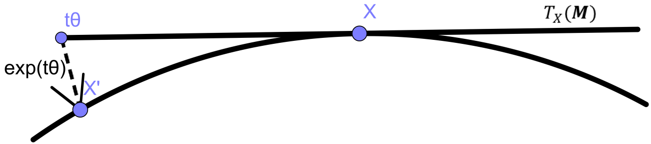
- $f \in C^2(\overline{\mathbb{M}})$ (*).

Data Model. Local Coordinates



$X' = \exp_X(t\theta)$, where $\theta \in T_X(\mathbb{M})$ and $|\theta| = 1$ and $t > 0$.

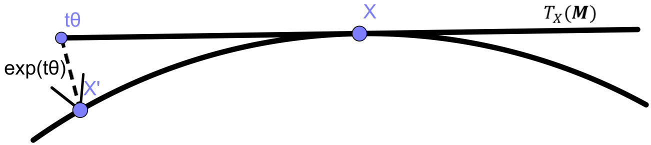
Data Model. Local Coordinates



$X' = \exp_X(t\theta)$, where $\theta \in T_X(\mathbb{M})$ and $|\theta| = 1$ and $t > 0$.

Kernel assumptions

- $K_\varepsilon(X, X') = K(X, \theta, \frac{|X - X'|}{\varepsilon})$, where $K(X, \theta, z)$ is a fixed nonnegative bounded function with bounded derivative, which is symmetric with respect to θ and z;

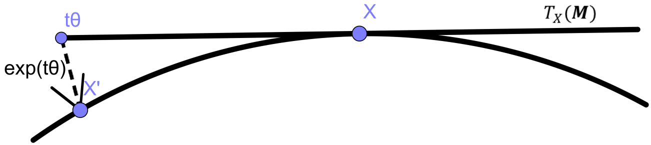


$X' = \exp_X(t\theta)$, where $\theta \in T_X(\mathbb{M})$ and $|\theta| = 1$ and $t > 0$.

Kernel assumptions

- $K_\varepsilon(X, X') = K(X, \theta, \frac{|X - X'|}{\varepsilon})$, where $K(X, \theta, z)$ is a fixed nonnegative bounded function with bounded derivative, which is symmetric with respect to θ and z;
- Function $K(X, \theta, z)$ has finite support with respect to z: $K(X, \theta, z) = 0$ for $z \geq 1$;

Data Model. Local Coordinates

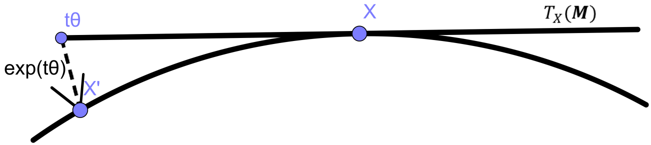


$X' = \exp_X(t\theta)$, where $\theta \in T_X(\mathbb{M})$ and $|\theta| = 1$ and $t > 0$.

Kernel assumptions

- $K_\varepsilon(X, X') = K(X, \theta, \frac{|X-X'|}{\varepsilon})$, where $K(X, \theta, z)$ is a fixed nonnegative bounded function with bounded derivative, which is symmetric with respect to θ and z;
- Function $K(X, \theta, z)$ has finite support with respect to z: $K(X, \theta, z) = 0$ for $z \geq 1$;
- Function $K(X, \theta, z)$ is Lipshitz with respect to z with constant $\tilde{C}_K$;

Data Model. Local Coordinates



$X' = \exp_X(t\theta)$, where $\theta \in T_X(\mathbb{M})$ and $|\theta| = 1$ and $t > 0$.

Kernel assumptions

- $K_\varepsilon(X, X') = K(X, \theta, \frac{|X - X'|}{\varepsilon})$, where $K(X, \theta, z)$ is a fixed nonnegative bounded function with bounded derivative, which is symmetric with respect to θ and z;
- Function $K(X, \theta, z)$ has finite support with respect to z: $K(X, \theta, z) = 0$ for $z \geq 1$;
- Function $K(X, \theta, z)$ is Lipshitz with respect to z with constant $\tilde{C}_K$;
- $\inf_{X \in \mathbb{M}} \int_{S_q} \int_{1/2}^{1} K(X, \theta, z) z^{q-1} d\theta dt \geq C_{K,1/2} > 0$.

Data Model. Parameter

For the neighborhood parameter $\varepsilon = \varepsilon(N) > 0$ it is assumed that

- For $N \to \infty$: $\varepsilon \to 0$;
- For $N \to \infty$: $N \cdot \varepsilon^q \to \infty$;
- For $N \to \infty$: $N \cdot \varepsilon^{q+4} \to 0$.

Problem Statement for Manifold Learning

Replace $f(X') \mapsto F(X'|X)$.

Replace $f(X') \mapsto F(X'|X)$.
In local coordinates

$$F(X'|X) = \phi(X, \theta) \cdot t^d + O(t^{d+1}),$$

where $d \in \mathbb{N}$ is the main order of $F(X'|X)$ (and d doesn't depend on $X \in \mathbb{M}$).

Problem Statement for Manifold Learning

Replace $f(X') \mapsto F(X'|X)$.
In local coordinates

$$F(X'|X) = \phi(X, \theta) \cdot t^d + O(t^{d+1}),$$

where $d \in \mathbb{N}$ is the main order of $F(X'|X)$ (and d doesn't depend on $X \in \mathbb{M}$).
Find asymptotic properties of the kernel nonparametric estimator

$$F_N(X) = \frac{\sum_{n=1}^{N} K_N(X, X_n) F(X_n|X)}{\sum_{n=1}^{N} K_N(X, X_n)}$$

Problem Statement for Manifold Learning

Replace $f(X') \mapsto F(X'|X)$.
In local coordinates

$$F(X'|X) = \phi(X, \theta) \cdot t^d + O(t^{d+1}),$$

where $d \in \mathbb{N}$ is the main order of $F(X'|X)$ (and d doesn't depend on $X \in \mathbb{M}$).
Find asymptotic properties of the kernel nonparametric estimator

$$F_N(X) = \frac{\sum_{n=1}^{N} K_N(X, X_n) F(X_n|X)}{\sum_{n=1}^{N} K_N(X, X_n)}$$

and of its average

$$F_N = \frac{1}{N} \sum_{n=1}^{N} F_N(X_n).$$

Limits

$$\overline{F}(X) = \frac{\int_{S^{q-1}} \rho_{E,m}(X,\theta)\phi(X,\theta)d\theta}{\int_{S^{q-1}} \rho_{E,0}(X,\theta)d\theta},$$

Limits

$$\overline{F}(X) = \frac{\int_{S^{q-1}} \rho_{E,m}(X,\theta)\phi(X,\theta)d\theta}{\int_{S^{q-1}} \rho_{E,0}(X,\theta)d\theta},$$

and

$$\overline{F} = \int_{\mathbb{M}} \overline{F}(X)d\mu(X),$$

Limits

$$\overline{F}(X) = \frac{\int_{S^{q-1}} \rho_{E,m}(X,\theta)\phi(X,\theta)d\theta}{\int_{S^{q-1}} \rho_{E,0}(X,\theta)d\theta},$$

and

$$\overline{F} = \int_{\mathbb{M}} \overline{F}(X)d\mu(X),$$

where S^{q-1} is sphere in q-dimensional space,

$$\rho_{E,m}(X,\theta) = \int_0^1 K(X,\theta,t)\, t^{q+m-1}dt.$$

Limits for $F(X'|X) = f(X')$

In local coordinates

$$f(X') = f(X) \cdot t^0 + O(t).$$

$$\overline{F}(X) = \frac{\int_{S^{q-1}} \rho_{E,m}(X,\theta)\phi(X,\theta)d\theta}{\int_{S^{q-1}} \rho_{E,0}(X,\theta)d\theta} = f(X),$$

and

$$\overline{F} = \int_{\mathbb{M}} f(X) d\mu(X).$$

Results-1

Statement 1 ($F_N(X)$ consistency). For each $X \in \mathbb{M}$ as $N \to \infty$

$$\mathbb{E}F_N(X) \to \overline{F}(X);$$
$$N\varepsilon^q \cdot \mathtt{Var}F_N(X) \to d(X);$$
$$F_N(X) \to^p \overline{F}(X),$$

where $d(X)$ is an explicit function, $\to^p$ is convergence in probability.

Statement 1 ($F_N(X)$ consistency). For each $X \in \mathbb{M}$ as $N \to \infty$

$$
\begin{aligned}
\mathbb{E} F_N(X) &\to \overline{F}(X); \\
N\varepsilon^q \cdot \mathtt{Var} F_N(X) &\to d(X); \\
F_N(X) &\to^p \overline{F}(X),
\end{aligned}
$$

where $d(X)$ is an explicit function, $\to^p$ is convergence in probability.

Theorem 1 (central limit for $F_N(X)$). For each $X \in \mathbb{M}$ as $N \to \infty$

$$\sqrt{N\varepsilon^q} \cdot \big(F_N(X) - \overline{F}(X)\big) \to^D N(m(X), d(X)),$$

where $m(X)$ and $d(X)$ are explicit functions, $m(X) \equiv 0$ for d:2, $\to^D$ is convergence in distribution.

Theorem 2 (large deviation probability for $F_N(X)$). For each ε-bounded from manifold boundary point $X \in \mathbb{M}_\varepsilon$, and for each $N > N_0$ and $\varepsilon < \varepsilon_0$ and $z \in [0, 1]$

$$P\left(\left|F_N(X) - \overline{F}(X)\right| \geq z + \varepsilon \cdot I(d \neq 2) \cdot C_{LD1} + \varepsilon^2 \cdot C_{LD2}\right) \leq 4 \cdot \exp\left(-\frac{z^2 \cdot N\varepsilon^q}{\sigma^2(X)}\right),$$

where $N_0, \varepsilon_0, C_{LD1}, C_{LD2}$ are positive constants, $\sigma^2(X)$ is bounded function.

Note for Results-2

The optimal rate in **Theorem 2** is ε^2 for

$$\begin{cases} z^2 = \frac{1}{N\varepsilon^q} \\ z = \varepsilon^2 \end{cases}$$

and it equals to $N^{-\frac{2}{q+4}}$.
Such rate is equal to the rate for classical case (where $p \mapsto q$).

Theorem 3 (uniform large deviation probability for $F_N(X)$). For each $z \in [0, 1]$, for $N > N_0$, $\varepsilon < \varepsilon_0$ for each ε-bounded from manifold boundary point $X \in \mathbb{M}_\varepsilon$

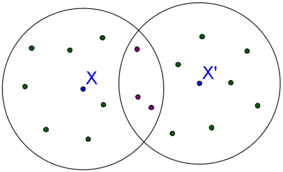
$$P\left(\sup_{X \in \mathbb{M}_\varepsilon} \left|F_N(X) - \overline{F}(X)\right| \geq z + \varepsilon \cdot I(d \not\vdots 2) \cdot C_{LD1} + \varepsilon^2 \cdot \tilde{C}_{LD2}\right) \leq$$

$$\leq 4 \cdot \left(\frac{2a\sqrt{p}}{\varepsilon^3}\right)^p \cdot \exp\left(-\frac{z^2 \cdot N\varepsilon^q}{C_\sigma}\right)$$

$$+ \left(\frac{6a\sqrt{p}}{\varepsilon}\right)^p \cdot \exp\left(-1/16 \cdot N\varepsilon^q V_q p_{\min}^2 / (9 p_{\max})\right),$$

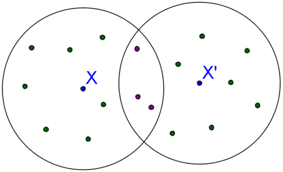
where $a, N_0, \varepsilon_0, C_{LD1}, \tilde{C}_{LD2}, p_{\min}, p_{\max}$ are positive constants.

Results-4



Statement 2 (F_N consistency). For $N \to \infty$:

$$F_N = \frac{1}{N} \sum_{n=1}^{N} F_N(X_n) \to^p \overline{F} = \int_{\mathbb{M}} \overline{F}(X) d\mu(X).$$



Statement 2 (F_N consistency). For $N \to \infty$:

$$F_N = \frac{1}{N}\sum_{n=1}^{N} F_N(X_n) \to^p \overline{F} = \int_{\mathbb{M}} \overline{F}(X)d\mu(X).$$

Theorem 4 (large deviation probability for F_N). Exist positive constants $N_{U,0}$, $C_{U,1}$, $C_{U,2}$, $C_{U,p}$ such that for $z \in [0,1]$ and $N > N_{U,0}$:

$$P\left(\left|F_N - \overline{F}\right| \geq z + \varepsilon \cdot I(d \not\vdots 2) \cdot C_{U,1} + \varepsilon^2 \cdot C_{U,2}\right) \leq \exp\left(-z^2 \cdot N\varepsilon^q \cdot C_{U,p}\right).$$

Relations with Other Papers

1. [Singer&Wu2012] local consistency and large deviation for local covariance (without explicit form for limit function)

Relations with Other Papers

1. [Singer&Wu2012] local consistency and large deviation for local covariance (without explicit form for limit function)
2. [Gine&Koltchinskii2006] asymptotic expansion for squared gradient (not nonparametric, but sample mean)

Relations with Other Papers

1. [Singer&Wu2012] local consistency and large deviation for local covariance (without explicit form for limit function)
2. [Gine&Koltchinskii2006] asymptotic expansion for squared gradient (not nonparametric, but sample mean)
3. [Belkin et.al. 2004, 2006, 2010] consistency for Laplace-Beltrami operator.

Conclusions

- specific for Manifold Learning nonparametric estimation on the manifold and its average are considered
- asymptotic expansion and uniform large deviation results are obtained for the considered nonparametric estimates
- the results could be used for the manifold learning algorithms analysis (already are used to get properties of Manifold Learning optimization procedures [Yanovich2017]).