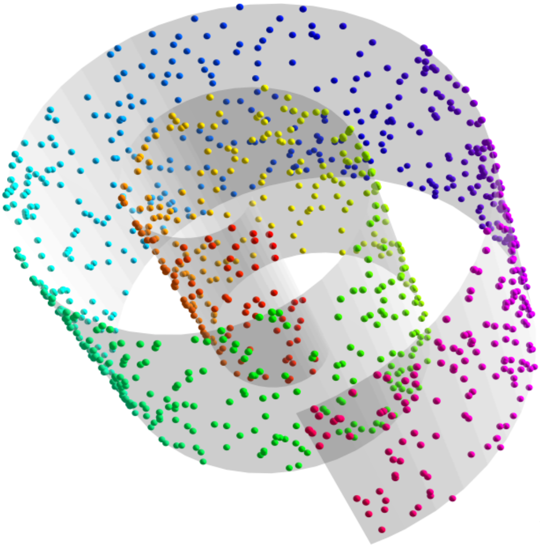
Minimax Rate for Dimension Estimator

Jisu Kim (Carnegie Mellon University)
Alessandro Rinaldo (Carnegie Mellon University)
Larry Wasserman (Carnegie Mellon University)

2015.07.27

Manifold Learning finds an underlying manifold to reduce dimension.



[1]

---

[1] http://www.skybluetrades.net/blog/posts/2011/10/30/machine-learning/

Intrinsic dimension of manifold need to be estimated.

- Most manifold learning algorithms require the intrinsic dimension of the manifold as input.
- Intrinc dimension is rarely known in advance and therefore has to be estimated.

Upper bounds and lower bounds of minimax rate is of interest.

- Various intrinsic dimension estimators have been proposed, but universal theoretical bound have not been obtained.
- Minimax rate is the risk of an estimator that performs best in the worst case, as a function of sample size.
- $$R_n = \inf_{\hat{\dim}_n} \sup_{P \in \mathcal{P}} \mathbb{E}_{P^{(n)}} \left[\ell \left(\hat{\dim}_n(X),\ \dim(P) \right) \right]$$
  - $X = (X_1, \cdots, X_n)$ is drawn from a fixed distribution P, where P is contained in set of distributions $\mathcal{P}$.
  - estimator $\hat{\dim}_n$ is any function of data X.

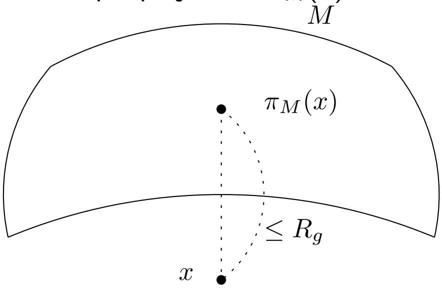
The supporting manifold M is assumed to be bounded.

$$M \subset I := [-K_I, K_I]^m \subset \mathbb{R}^m \text{ with } K_I \in (1, \infty)$$

The curvature is assumed to be bounded to avoid an arbitrarily complicated manifold.

Definition

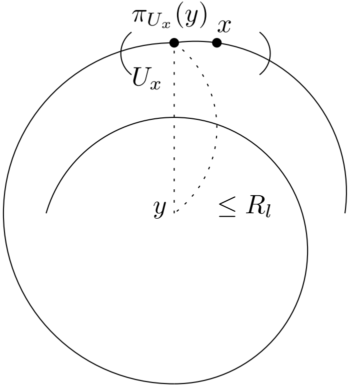
Fix $0 \leq \kappa_g < \infty$. A compact d-dimentional topological manifold M is of global curvature $\leq \kappa_g$, if for all points x in $R_g \left(:= \frac{1}{\kappa_g}\right)$-neighborhood of M has unique projection $\pi_M(x)$ to M.



The curvature is assumed to be bounded to avoid an arbitrarily complicated manifold.

Definition

Fix $0 \leq \kappa_l \leq \kappa_g$. M is of local curvature $\leq \kappa_l$, if for all points in $x \in M$, there exists neighborhood $U_x \subset M$ such that U_x is of global curvature $\leq \kappa_l$.



Density is bounded away from ∞ with respect to the uniform measure.

- Distribution P is absolutely continuous to induced Lebesgue measure vol_M, and $\frac{dP}{dvol_M} \leq K_p$ for fixed K_p.
- This implies that the distribution on the manifold is of essential dimension d.
- $\mathcal{P}^d_{\kappa_l, \kappa_g, K_p}$ denotes set of distributions P that is supported on bounded d-dimensional manifold of global curvature $\leq \kappa_g$, global curvature $\leq \kappa_l$, and whose density is bounded by K_p.

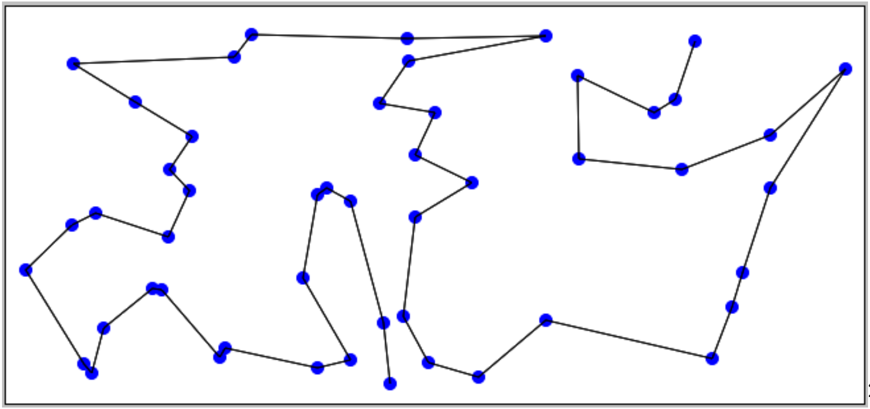
Binary classification and $0-1$ loss are considered.

- $$R_n = \inf_{\hat{\dim}_n} \sup_{P \in \mathcal{P}} \mathbb{E}_{P^{(n)}} \left[\ell \left(\hat{\dim}_n(X),\ \dim(P) \right) \right]$$
- We assume that the manifolds are of two possible dimensions, d_1 and d_2, so considered distribution set is $\mathcal{P} = \mathcal{P}^{d_1}_{\kappa_l,\kappa_g,K_p} \cup \mathcal{P}^{d_2}_{\kappa_l,\kappa_g,K_p}$.
- $0-1$ loss function is considered, so for all $x, y \in \mathbb{R}$, $\ell(x,y) = I(x \neq y)$.

The maximum risk of any chosen estimator provides an upper bound on the minimax rate.

$$
\begin{aligned}
R_n &= \inf_{\hat{\dim}_n} \sup_{P \in \mathcal{P}} \mathbb{E}_{P^{(n)}} \left[\ell \left(\hat{\dim}_n(X), \dim(P) \right) \right] \\
&\leq \underbrace{\sup_{P \in \mathcal{P}} \mathbb{E}_{P^{(n)}} \left[\ell \left(\hat{\dim}_n(X), \dim(P) \right) \right]}_{\text{the maximum risk of any chosen estimator}}
\end{aligned}
$$

TSP(Travelling Salesman Problem) path finds shortest path that visits each points exactly once.


[2]

[2] http://www.heatonresearch.com/fun/tsp/anneal

Our estimator estimates dimension to be d_2 if d_1-squared length of TSP generated by the data is long.

- When intrinsic dimesion is higher, length of TSP path is likely to be longer.
- $$\hat{\dim}_n(X) = d_1 \iff$$
  $$\exists \sigma \in S_n \ s.t \sum_{i=1}^{n-1} \|X_{\sigma(i+1)} - X_{\sigma(i)}\|_{\mathbb{R}^m}^{d_1} \leq C^1_{K_I, d_1, m}\left(1 + \kappa_g^{m-d_1}\right),$$

  where $C^1_{K_I, d_1, m}$ is some constant that depends only on K_I, d_1, and m.

Our estimator has maximum risk of $O\left(n^{-\left(\frac{d_2}{d_1}-1\right)n}\right)$.

- Our estimator makes error with probability at most $O\left(n^{-\left(\frac{d_2}{d_1}-1\right)n}\right)$ if intrinsic dimension is d_2.
- Our estimator is always correct when the intrinsic dimension is d_1.

Our estimator makes error with probability at most $O\left(n^{-\left(\frac{d_2}{d_1}-1\right)n}\right)$ if intrinsic dimension is d_2.

- Based on the following lemma:

Lemma

Let $X_1, \cdots, X_n \sim P \in \mathcal{P}^{d_2}_{\kappa_l, \kappa_g, K_p}$, then

$$P^{(n)}\left[\sum_{i=1}^{n-1} \|X_{i+1} - X_i\|^{d_1} \leq L\right] \leq \frac{\left(C^2_{K_l, K_p, d_1, d_2, m}\right)^{n-1} L^{\frac{d_2}{d_1}(n-1)} \left(1 + \kappa_g^{(m-d_2)(n-1)}\right)}{(n-1)^{\left(\frac{d_2}{d_1}-1\right)(n-1)}(n-1)!},$$

where $C^2_{K_l, K_p, d_1, d_2, m}$ depends only on K_l, K_p, d_1, d_2, m.

Our estimator is always correct when the intrinsic dimension is d_1.

- Based on following lemma:

Lemma

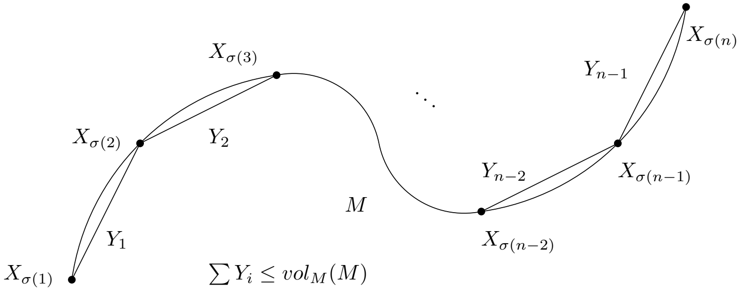
Let M be a d_1-dimensional manifold with global curvature $\leq \kappa_g$ and local curvature $\leq \kappa_l$, and $X_1, \cdots, X_n \in M$. Then there exists $C^3_{K_l,d_1,m}$ which depends only on m, d_1 and K_l, and there exists $\sigma \in S_n$ such that

$$\sum_{i=1}^{n-1} \|X_{\sigma(i+1)} - X_{\sigma(i)}\|_{\mathbb{R}^m}^{d_1} \leq C^3_{K_l,d_1,m}\left(1 + \kappa_g^{m-d_1}\right).$$

Our estimator is always correct when the intrinsic dimension is d_1.

$$\sum_{i=1}^{n-1} \|X_{\sigma(i+1)} - X_{\sigma(i)}\|_{\mathbb{R}^m}^{d_1} \leq C^3_{K_l, d_1, m} \left(1 + \kappa_g^{m-d_1}\right).$$

- When $d_1 = 1$ so that the manifold is a curve, length of TSP path is bounded by length of curve $vol_M(M)$.

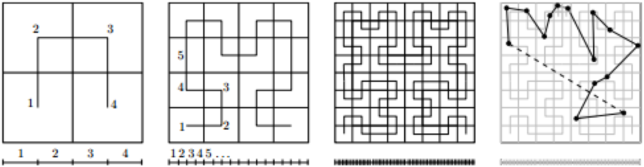


- Global curvature $\leq \kappa_g$ implies $vol_M(M)$ is bounded.

Our estimator is always correct when the intrinsic dimension is d_1.

$$\sum_{i=1}^{n-1} \|X_{\sigma(i+1)} - X_{\sigma(i)}\|_{\mathbb{R}^m}^{d_1} \leq C_{K_I,d_1,m}^3 \left(1 + \kappa_g^{m-d_1}\right).$$

- When $d_1 > 1$, Several conditions implied by regularity conditions combined with Hölder continuity of d_1-dimensional space-filling curve is used.



(a) (b) (c) (d)

Our estimator is always correct when the intrinsic dimension is d_1.

$$\sum_{i=1}^{n-1} \|X_{\sigma(i+1)} - X_{\sigma(i)}\|_{\mathbb{R}^m}^{d_1} \leq C_{K_I,d_1,m}^{3} \left(1 + \kappa_g^{m-d_1}\right).$$

- When $d_1 > 1$, Several conditions implied by regularity conditions combined with Hölder continuity of d_1-dimensional space-filling curve is used.

Lemma

(Space-filling curve) There exists surjective map $\psi_d : \mathbb{R} \to \mathbb{R}^d$ which is Hölder continuous of order $1/d$, i.e.

$$0 \leq \forall s, t \leq 1, \; \|\psi_d(s) - \psi_d(t)\|_{\mathbb{R}^d} \leq 2\sqrt{d+3}|s-t|^{1/d}.$$

Mimimax rate is upper bounded by $O\left(n^{-\left(\frac{d_2}{d_1}-1\right)n}\right)$.

Proposition

Let $1 \leq d_1 < d_2 \leq m$. Then

$$\inf_{\hat{\dim}_n} \sup_{P \in \mathcal{P}^{d_1}_{\kappa_l,\kappa_g,K_p,K_v} \cup \mathcal{P}^{d_2}_{\kappa_l,\kappa_g,K_p,K_v}} \mathbb{E}_{P^{(n)}}\left[l\left(\hat{\dim}_n, \dim(P)\right)\right] \leq \left(C^4_{K_l,K_p,d_1,d_2,m}\right)^n \left(1 + \kappa_g^{\left(\frac{d_2}{d_1}m + m - 2d_2\right)n}\right) n^{-\left(\frac{d_2}{d_1}-1\right)n}.$$

for some $C^4_{K_l,K_p,d_1,d_2,m}$ that depends only on K_l, K_p, d_1, d_2, m.

A subset $T \subset [-K_I, K_I]^n$ and set of distributions $\mathcal{P}_1^{d_1}$, $\mathcal{P}_2^{d_2}$ are found so that, whenever $X = (X_1, \cdots, X_n) \in T$, we cannot distinguish two models.

- The lower bound measures how hard it is to tell whether the data come from a d_1 or d_2 -dimensional manifold.
- T, $\mathcal{P}_1^{d_1}$ and $\mathcal{P}_2^{d_2}$ are linked to the lower bound by using Le Cam's lemma.

Le Cam's lemma provides lower bounds based on the minimum of two densities $q_1 \wedge q_2$, where q_1, q_2 are in convex hull of $\mathcal{P}_1^{d_1}$ and convex hull of $\mathcal{P}_2^{d_2}$, respectively.

Lemma

Let $\mathcal{P}$ be a set of probability measures on $(\Omega, \mathcal{F})$, and $\mathcal{P}_1, \mathcal{P}_2 \subset \mathcal{P}$ be such that for all $P \in \mathcal{P}_i$, $\theta(P) = \theta_i$ for $i = 1, 2$, and $X : \Omega \to I^n$ is observations. Let $Q_1 \in conv(\mathcal{P}_1)$ and $Q_2 \in conv(\mathcal{P}_2)$, where $conv(\mathcal{P}_i)$ is convex hull of $\mathcal{P}_i$. Assume that induced measure of X on (Ω, Q_1) and (Ω, Q_2) has density q_1 and q_2 respectively with respect to $(I^n, \mathcal{B}(I^n), \nu)$, so that

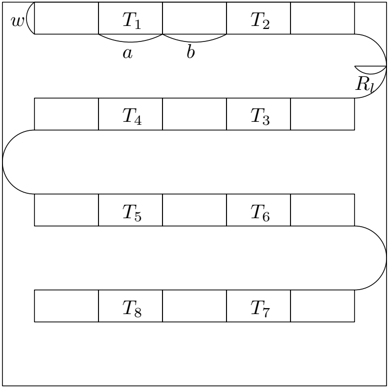
$$Q_1(X \in B) = \int_B q_1(x) d\nu(x) \text{ and } Q_2(X \in B) = \int_B q_2(x) d\nu(x).$$

Then

$$\inf_{\hat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_P[d(\hat{\theta}, \theta(P))] \geq \frac{d(\theta_1, \theta_2)}{4} \int [q_1(x) \wedge q_2(x)] d\nu(x).$$

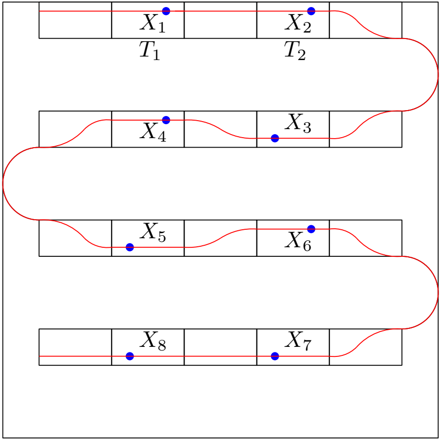
T is constructed so that for any $x = (x_1, \cdots, x_n) \in T$, there exists a d_1-dimensional manifold that satisfies regularity conditions and passes through $x_1, \cdots, x_n$.

- T_i's are cylinder sets in $[-K_I, K_I]^{d_2}$, and then T is constructed as $T = S_n \prod_{i=1}^{n} T_i$, where the permutation group S_n acts on $\prod_{i=1}^{n} T_i$ as a coordinate change.



T is constructed so that for any $x = (x_1, \cdots, x_n) \in T$, there exists a d_1-dimensional manifold that satisfies regularity conditions and passes through $x_1, \cdots, x_n$.

- Given $x_1, \cdots, x_n \in T$ (blue points), manifold of global curvature $\leq \kappa_g$ and local curvature $\leq \kappa_l$ (red line) passes through $x_1, \cdots, x_n$.



$\mathcal{P}_1^{d_1}$ is constructed as set of distributions that are supported on manifolds that passes through $x_1, \cdots, x_n$ for $x = (x_1, \cdots, x_n) \in T$, and $\mathcal{P}_2^{d_2}$ is a singleton set consisting of the uniform distirbution on $[-K_I, K_I]^{d_2}$.

If $X \in T$, it is hard to determine whether X is sampled from distribution P in either $\mathcal{P}_1^{d_1}$ or $\mathcal{P}_2^{d_2}$.

- There exists $Q_1 \in conv(\mathcal{P}_1^{d_1})$ and $Q_2 \in conv(\mathcal{P}_2^{d_2})$ such that $q_1(x) \geq Cq_2(x)$ for every $x \in T$ with $C < 1$.
- Then $q_1(x) \wedge q_2(x) \geq Cq_2(x)$ if $x \in T$, so $C \int_T q_2(x)dx$ can serve as lower bound of minimax rate.
- Based on following claim:

Claim

Let $T = S_n \prod_{i=1}^{n} T_i$. Then for all $x \in \operatorname{int} T$, there exists $r_x > 0$ such that for all $r < r_x$,

$$Q_1\left(\prod_{i=1}^{n} B_{\|\|_{\mathbb{R}^{d_2},\infty}}(x_i, r)\right) \geq \frac{2^{(1-d_1)n}}{\kappa_l^{(d_2-d_1)n} K_l^{(d_2-d_1)n}} Q_2\left(\prod_{i=1}^{n} B_{\|\|_{\mathbb{R}^{d_2},\infty}}(x_i, r)\right).$$

Mimimax rate is lower bounded by $\Omega\left(n^{-2(d_2-d_1)n}\right)$.

- Lower bound below is now combination of Le Cam's lemma, constructions of T, $\mathcal{P}_1^{d_1}$, $\mathcal{P}_2^{d_2}$, and claim.

Proposition

Suppose $R_l < K_l$, then

$$\inf_{\hat{\dim}} \sup_{P \in \mathcal{P}^{d_1}_{\kappa_l,\kappa_g,K_p,K_v} \cup \mathcal{P}^{d_2}_{\kappa_l,\kappa_g,K_p,K_v}} \mathbb{E}_{P^{(n)}}[l(\hat{\dim}_n, \dim(P))]$$
$$\geq \left(C^5_{d_1,d_2,K_l}\right)^n \kappa_l^{-(d_2-d_1)n} \min\left\{\kappa_l^{2(d_2-d_1)+1} n^{-2}, 1\right\}^{(d_2-d_1)n},$$

for some constant $C^5_{d_1,d_2,K_l}$ that depends only on d_1, d_2, and K_l.

Multinary classification and $0-1$ loss are considered.

- $$R_n = \inf_{\hat{\dim}_n} \sup_{P \in \mathcal{P}} \mathbb{E}_{P^{(n)}} \left[\ell \left(\hat{\dim}_n(X),\ \dim(P) \right) \right]$$
- Now the manifolds are of any dimensions between 1 and m, so considered distribution set is $\mathcal{P} = \bigcup_{d=1}^{m} \mathcal{P}^{d}_{\kappa_l, \kappa_g, K_p}$.
- $0-1$ loss function is considered, so for all $x, y \in \mathbb{R}$, $\ell(x, y) = I(x = y)$.

Mimimax rate is upper bounded by $O\left(n^{-\frac{1}{m-1}n}\right)$, and lower bounded by $\Omega\left(n^{-2n}\right)$.

Proposition

Suppose $R_I < K_I$, then

$$\begin{aligned}&\left(C^7_{K_I}\right)^n \kappa_I^{-n} \min\left\{\kappa_I^3 n^{-2}, 1\right\}^n \\ &\leq \inf_{\hat{\dim}_n} \sup_{P\in\mathcal{P}} \mathbb{E}_{P^{(n)}}\left[I\left(\hat{\dim}_n, \dim(P)\right)\right] \\ &\leq \left(C^6_{K_I,K_p,m}\right)^n \left(1+\kappa_g^{(m^2-m)n}\right) n^{-\frac{1}{m-1}n}.\end{aligned}$$

for some $C^6_{K_I,K_p,m}$ that depends only on K_I, K_p, m, and for some $C^7_{K_I}$ *that depends only on K_I.*

Thank you!