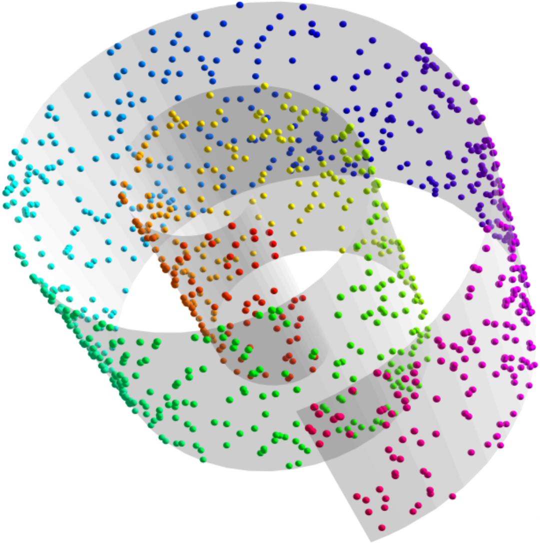
Minimax Rate for Estimating the Dimension of a Manifold

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

2015.11.14

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 a Manifold needs to be Estimated.

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

Minimax Rate is of Interest.

- Minimax rate is the risk of an estimator that performs best in the worst case, as a function of the sample size.
- $$R_n = \inf_{\hat{\dim}_n} \sup_{P \in \mathcal{P}} \underbrace{\mathbb{E}_{P^{(n)}} \left[\ell \left(\hat{\dim}_n(X),\ \dim(P) \right) \right]}_{\text{risk of an estimator}}$$
  - $X = (X_1, \cdots, X_n)$ 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 the data X.

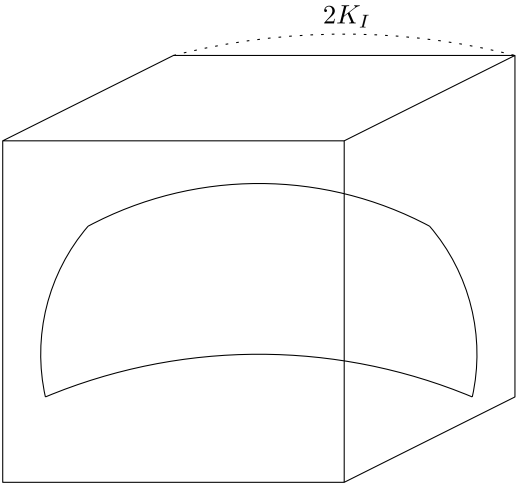
Minimax Rate is of Interest.

- Minimax rate is the risk of an estimator that performs best in the worst case, as a function of the sample size.
- $$R_n = \inf_{\hat{\dim}_n} \underbrace{\sup_{P \in \mathcal{P}} \mathbb{E}_{P^{(n)}} \left[\ell \left(\hat{\dim}_n(X),\ \dim(P) \right) \right]}_{\text{risk of an estimator in the worst case}}$$
  - $X = (X_1, \cdots, X_n)$ 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 the data X.

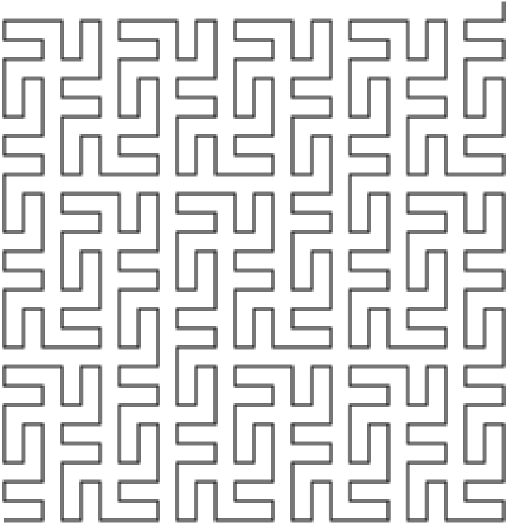
Minimax Rate is of Interest.

- Minimax rate is the risk of an estimator that performs best in the worst case, as a function of the sample size.
- $$R_n = \underbrace{\inf_{\hat{\dim}_n} \sup_{P \in \mathcal{P}} \mathbb{E}_{P^{(n)}} \left[\ell \left(\hat{\dim}_n(X),\ \dim(P) \right) \right]}_{\text{risk of an estimator that performs best in the worst case}}$$
  - $X = (X_1, \cdots, X_n)$ 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 the data X.

Regularity Conditions on Measures and Conditions on Distributions and Supporting Manifolds are Assumed.



Regularity Conditions on Measures and Conditions on Distributions and Supporting Manifolds are Assumed.

 [2]

[2] https://people.csail.mit.edu/jaffer/Geometry/PSFC

Regularity Conditions on Measures and Conditions on Distributions and Supporting Manifolds are Assumed.

$$\frac{dP}{dvol_M} \leq K_p$$

Regularity Conditions on Measures and Conditions on Distributions and Supporting Manifolds are Assumed.

- $\mathcal{P}^d$ denotes set of distributions P with bounded support, bounded curvature, and bounded density.

Binary Classification and $0 - 1$ loss are Considered.

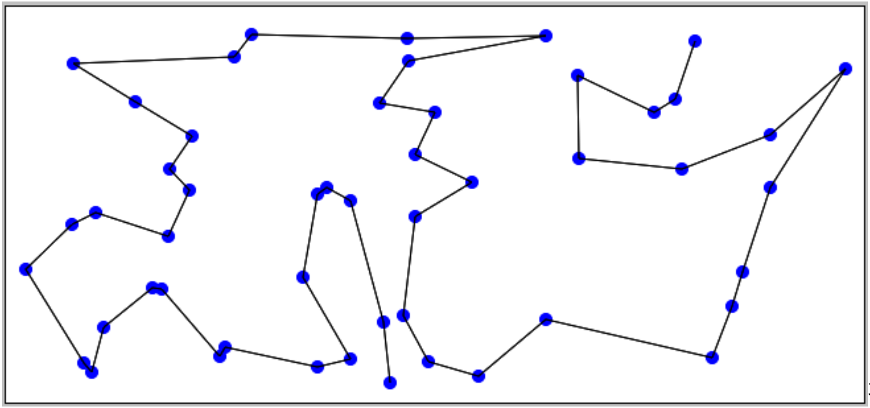
$$R_n = \inf_{\hat{\dim}_n} \sup_{P \in \mathcal{P}} \mathbb{E}_{P^{(n)}} \left[\ell \left(\hat{\dim}_n(X),\ \dim(P) \right) \right]$$

- We assume that the manifolds are of two possible dimensions, d_1 and d_2, so the considered distribution set is $\mathcal{P} = \mathcal{P}^{d_1} \cup \mathcal{P}^{d_2}$.
- $0 - 1$ 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.


[3]

---

[3] 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.
- $$\hat{\dim}_n(X) = d_1 \iff$$
  $$\exists \sigma \in S_n \ s.t \sum_{i=1}^{n-1} \|X_{\sigma(i+1)} - X_{\sigma(i)}\|_{\mathbb{R}^m}^{d_1} \leq C,$$

  where C is some constant that depends only on regularity conditions.

Mimimax Rate is Upper Bounded by $O\left(n^{-\left(\frac{d_2}{d_1}-1\right)n}\right)$.

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

Proposition

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

$$\inf_{\hat{\dim}_n} \sup_{P \in \mathcal{P}^{d_1} \cup \mathcal{P}^{d_2}} \mathbb{E}_{P^{(n)}}\left[l\left(\hat{\dim}_n, \dim(P)\right)\right] \lesssim n^{-\left(\frac{d_2}{d_1}-1\right)n}.$$

Le Cam's Lemma provides Lower Bounds based on the Minimum of Two Densities $q_1 \wedge q_2$, where q_i are in Convex Hull of $\mathcal{P}^{d_i}$.

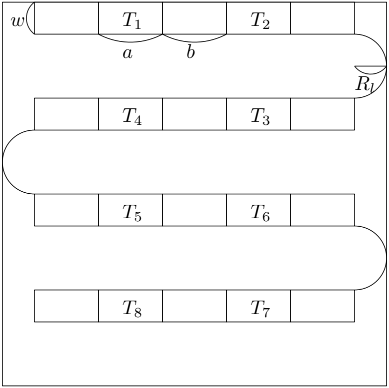
Lemma

Let $\mathcal{P}$ be a set of probability measures, and $\mathcal{P}^{d_1}, \mathcal{P}^{d_2} \subset \mathcal{P}$ be such that for all $P \in \mathcal{P}^{d_i}$, $\theta(P) = \theta_i$ for $i = 1, 2$. For any $Q_i \in co(\mathcal{P}_i)$, let q_i be density of Q_i with respect to measure ν. Then

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

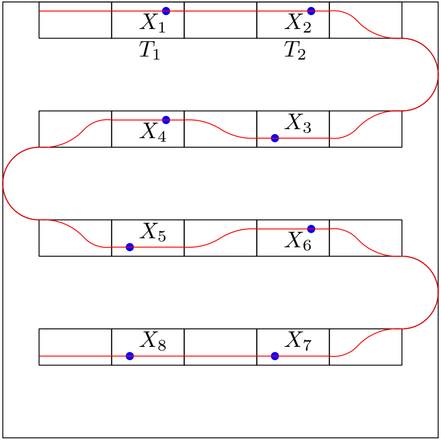
T is Constructed so that for any $x = (x_1, \cdots, x_n) \in T$, there exists a d_1-dimensional Manifold belonging to the Model and Passing through $x_1, \cdots, x_n$.

- 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 belonging to the Model and Passing through $x_1, \cdots, x_n$.

- Given $x_1, \cdots, x_n \in T$ (blue points), a manifold in the model (red line) passes through $x_1, \cdots, x_n$.



$\mathcal{P}^{d_1}$ 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 $\mathcal{P}^{d_2}$ is a Singleton Set Consisting of the Uniform Distirbution on $[-K_I, K_I]^{d_2}$.

Mimimax Rate is Lower Bounded by $\Omega\left(n^{-2(d_2-d_1)n}\right)$.

- The lower bound below is from Le Cam's lemma with the constructed $\mathcal{P}_1^{d_1}$ and $\mathcal{P}_2^{d_2}$.

Proposition

$$\inf_{\hat{\dim}} \sup_{P \in \mathcal{P}^{d_1} \cup \mathcal{P}^{d_2}} \mathbb{E}_{P^{(n)}}[l(\hat{\dim}_n, \dim(P))] \gtrsim n^{-2(d_2-d_1)n}.$$

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 $\mathcal{P} = \bigcup_{d=1}^{m} \mathcal{P}^d$.
- $0 - 1$ 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

$$n^{-2n} \lesssim \inf_{\hat{\dim}_n} \sup_{P \in \mathcal{P}} \mathbb{E}_{P^{(n)}}\left[l\left(\hat{\dim}_n, \dim(P)\right)\right] \lesssim n^{-\frac{1}{m-1}n}.$$

Thank you!