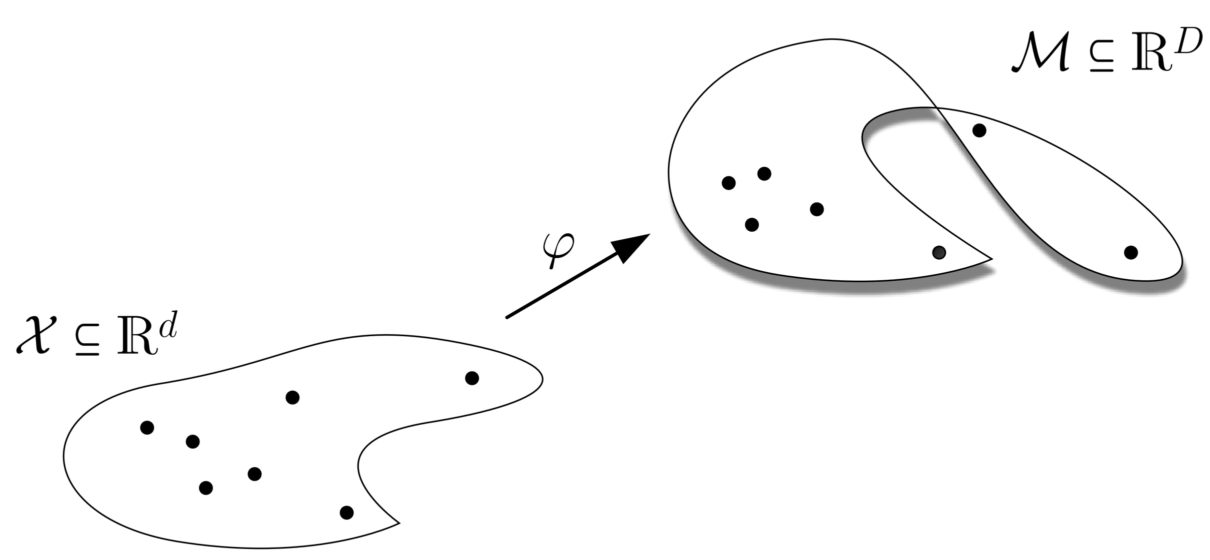
Dimensionality estimation without distances

Matthäus Kleindessner, Ulrike von Luxburg — University of Hamburg

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Intrinsic dimensionality



Manifold assumption: data lies on or close to low-dimensional manifold embedded in a high-dimensional observation space

Formally: $\mathcal{X} \subseteq \mathbb{R}^d$, embedding $\varphi : \mathcal{X} \to \mathcal{M} \subseteq \mathbb{R}^D$, probability density function $f : \mathcal{X} \to \mathbb{R}$ sample points drawn from f are embedded into the observation space $\mathbb{R}^D$ via φ (possibly disturbed by noise), giving the sample $\mathcal{D} = \{x_1, \ldots, x_n\}$

Given some information about $\mathcal{D}$, we want to infer d, the *intrinsic dimension* of the data.

Our estimators

Existing methods: knowledge of distance values $\|x_i - x_j\|_{\mathbb{R}^D}$

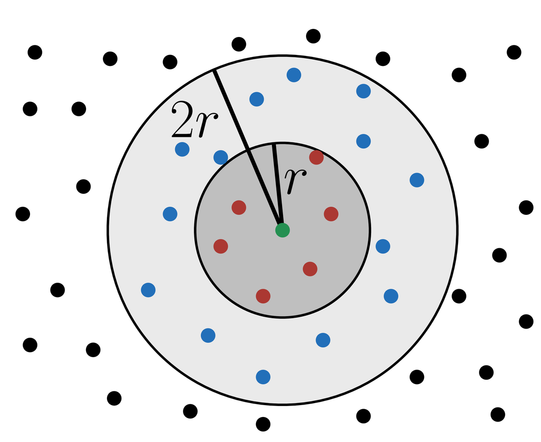
Our methods: directed, UNWEIGHTED kNN-graph on $\mathcal{D}$ — only *ordinal* distance information

Estimator based on doubling property (naive)

doubling property of the Lebesgue measure: $\lambda_d(B(x, 2r)) = 2^d \lambda_d(B(x,r))$

Fix a sample point x_i. Balls $B_{SP}(i,1)$ and $B_{SP}(i,2)$ in G (with respect to the shortest path distance) approximately correspond to balls $B(x_i, r)$ and $B(x_i, 2r)$ in $\mathbb{R}^d$ for some small r. Consequently:

$$\frac{k+1}{|B_{SP}(i,2)|} = \frac{|B_{SP}(i,1)|}{|B_{SP}(i,2)|} \approx \frac{n\ f(x_i)\ \lambda_d(B(x_i,r))}{n\ f(x_i)\ \lambda_d(B(x_i,2r))} = \frac{1}{2^d}$$



$$E_{DP}(A) := -\log_2 L_{DP}(A) \text{ with } L_{DP}(A) := \frac{1}{|A|} \sum_{i \in A} \frac{k+1}{|B_{SP}(i,2)|}, \quad A \subseteq \{1, \ldots, n\}$$

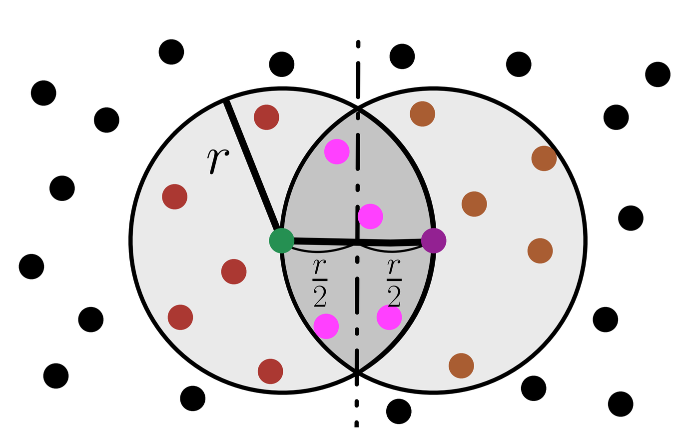
Estimator based on spherical caps (elaborate)

volume of a spherical cap with height $r/2$ of a ball with radius r in $\mathbb{R}^d$:

$$\frac{1}{2} \eta_d r^d I_{3/4}((d+1)/2, 1/2) \qquad (I_x(a,b) \ldots \text{ regularized incomplete beta function})$$

Fix a sample point x_i. If x_j sits at the boundary of $B(x_i, r)$, i.e. $\|x_i - x_j\|_{\mathbb{R}^d} = r$, then:

$$\frac{|B_{SP}(i,1) \cap B_{SP}(j,1)|}{|B_{SP}(i,1)|} \approx \frac{\lambda_d(B(x_i,r) \cap B(x_j,r))}{\lambda_d(B(x_i,r))} = I_{\frac{3}{4}}\left(\frac{d+1}{2}, \frac{1}{2}\right) =: S(d)$$

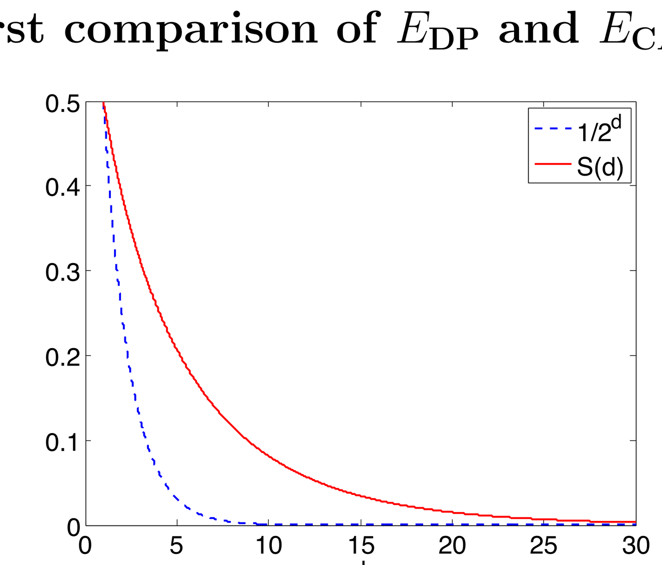


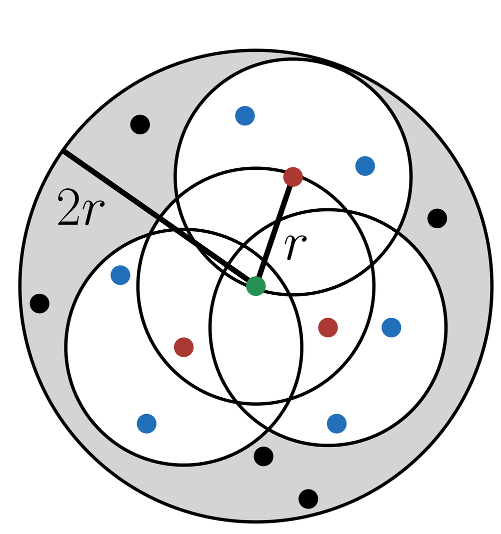
$|B_{SP}(i,1) \cap B_{SP}(j,1)|$ decreases as $\|x_i - x_j\|_{\mathbb{R}^d}$ increases. Consequently:

$$L_{CAP}(i) := \frac{\min_{j \in \{1,\ldots,n\}: i \to j \text{ in the graph}} |B_{SP}(i,1) \cap B_{SP}(j,1)|}{k+1} \approx S(d)$$

$$E_{CAP}(A) := S^{-1}(L_{CAP}(A)) \text{ with } L_{CAP}(A) := \frac{1}{|A|} \sum_{i \in A} L_{CAP}(i), \quad A \subseteq \{1, \ldots, n\}$$

First comparison of E_{DP} and E_{CAP}





Two observations suggest that E_{CAP} might perform better than E_{DP}. Left: The function S is more well-behaved in terms of inversion than $T(d) = 1/2^d$. Right: E_{DP} might be biased: the union of the small balls approximates the large ball, but ignores a substantial part close to its boundary (shaded area).

Theoretical results: consistency of our estimators

Regularity assumptions:

- domain $\mathcal{X} \subseteq \mathbb{R}^d$ is compact and has boundary of measure 0, i.e. $\lambda_d(\partial\mathcal{X}) = 0$
- boundary $\partial\mathcal{X}$ is nice in the sense that there exist constants $\alpha, \varepsilon_0 > 0$ such that

$$\lambda_d(B(x,\varepsilon) \cap \mathcal{X}) \geq \alpha \cdot \lambda_d(B(x,\varepsilon)), \quad x \in \mathcal{X}, \varepsilon < \varepsilon_0$$

- density $f : \mathcal{X} \to \mathbb{R}$ satisfies $0 < f_{min} \leq f(x) \leq f_{max} < \infty$, $x \in \mathcal{X}$, and is Lipschitz continuous with constant L, i.e. $|f(x) - f(y)| \leq L\|x - y\|$, $x, y \in \mathcal{X}$

Theorem (Consistency of E_{DP} and E_{CAP})

Let the regularity assumptions hold. Let $\mathcal{D} = \{x_1, \ldots, x_n\} \subseteq \mathcal{X}$ be an i.i.d. sample from f and G be the directed, unweighted kNN-graph on $\mathcal{D}$. Given G as input and a vertex $i \in \{1, \ldots, n\}$ chosen uniformly at random, both $E_{DP}(\{i\})$ and $E_{CAP}(\{i\})$ converge to the true dimension d in probability as $n \to \infty$ if $k = k(n)$ satisfies $k \in o(n)$, $\log n \in o(k)$, and there exists $k' = k'(n)$ with $k' \in o(k)$ and $\log n \in o(k')$.

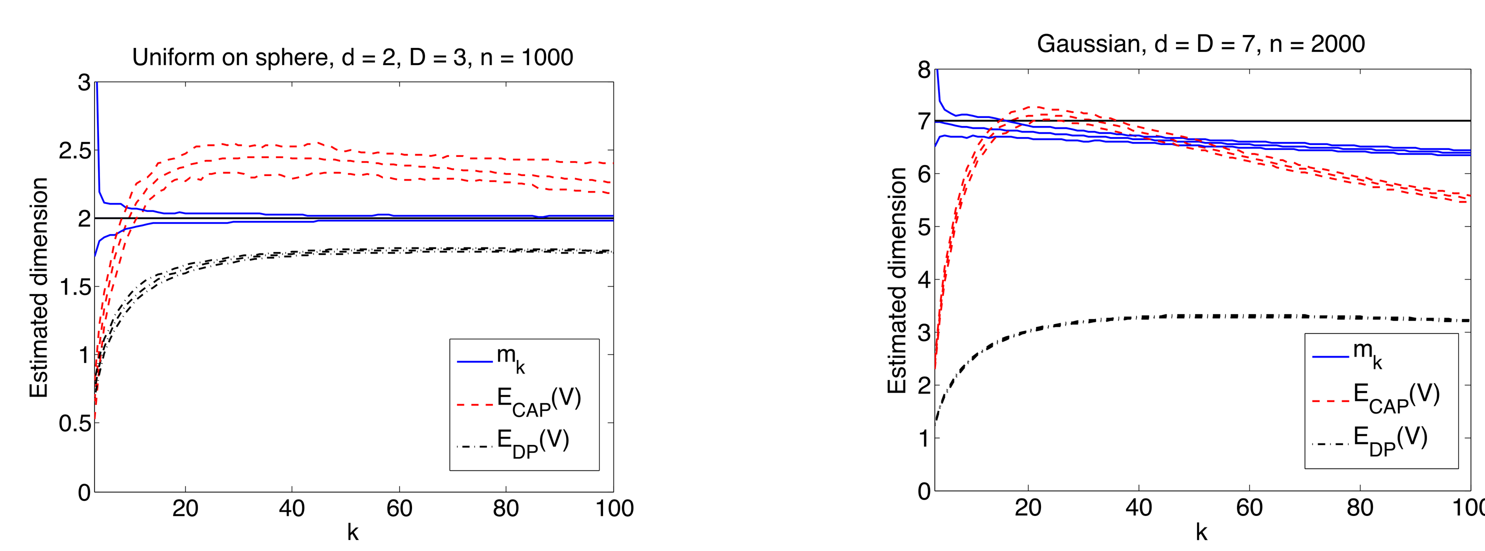
How can we choose k and k'? For example: $k = (\log n)^{1+\tau}$ and $k' = (\log n)^{1+\tau/2}$ for some $\tau > 0$

Experiments

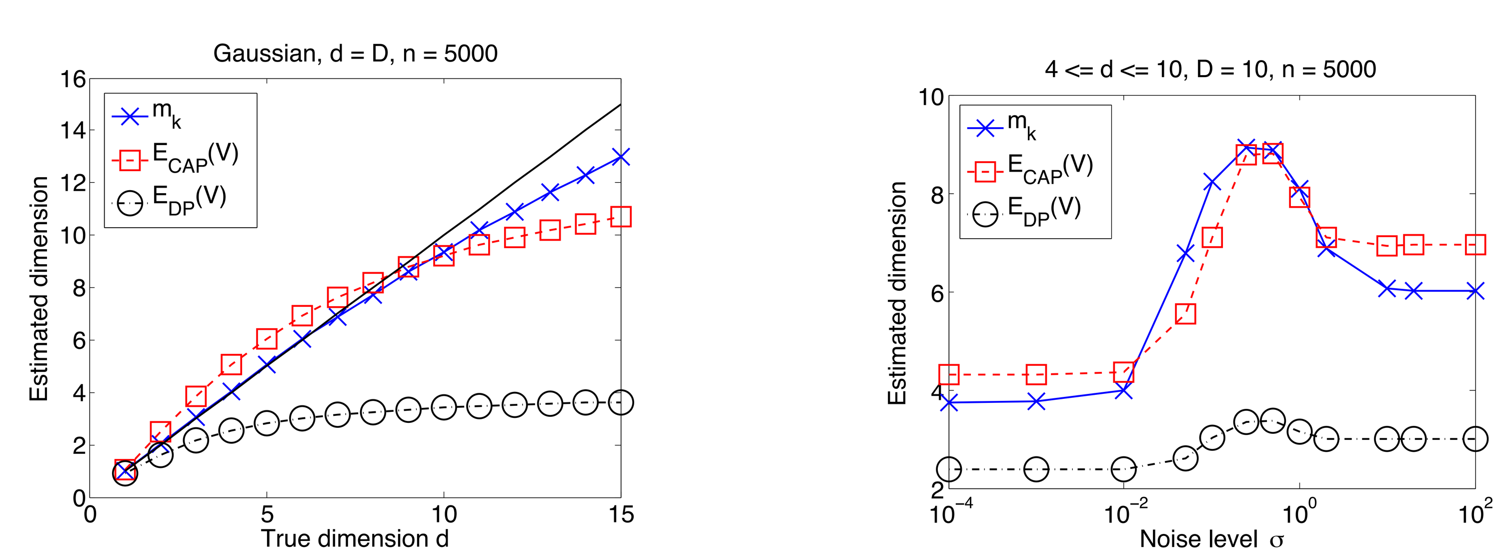
Comparison with estimators from the literature

				Our estimators (kNN-graph)		Standard estimators (distance information)		
	n	**Distribution / Dataset**	d	$E_{CAP}(V)$	$E_{DP}(V)$	MLE [1]	***CorrDim*** [2]	***RegDim*** [3]
	Artificial datasets (results averaged over 100 runs, $\pm STD$)							
1	1000	uniform on a helix in $\mathbb{R}^3$	1	1.00 (±0.05)	0.88 (±0.01)	1.00 (±0.01)	1.00 (±0.11)	0.99 (±0.01)
2	1000	Swiss roll in $\mathbb{R}^3$	2	2.14 (±0.05)	1.44 (±0.01)	1.94 (±0.02)	1.99 (±0.23)	1.87 (±0.04)
3	1000	$N_5(0, I)$	5	5.33 (±0.07)	2.47 (±0.01)	5.00 (±0.04)	4.91 (±0.56)	4.86 (±0.05)
4	1000	uniform on sphere $\mathbb{S}^7 \subseteq \mathbb{R}^8$	7	5.88 (±0.06)	2.82 (±0.01)	6.53 (±0.07)	6.85 (±0.66)	6.23 (±0.09)
5	5000	uniform on sphere $\mathbb{S}^7 \subseteq \mathbb{R}^8$	7	6.85 (±0.03)	3.21 (±0.00)	6.72 (±0.03)	6.95 (±0.98)	6.46 (±0.04)
6	1000	uniform on $[0,1]^{12}$	12	7.74 (±0.08)	3.04 (±0.01)	9.32 (±0.10)	10.66 (±1.18)	8.78 (±0.10)
7	5000	uniform on $[0,1]^{12}$	12	9.24 (±0.04)	3.50 (±0.00)	9.76 (±0.05)	10.83 (±1.49)	9.26 (±0.05)
	Real datasets (D = dimension of observation space)							
8	698	Isomap faces, $D = 4096 = 64^2$	?	3.04	1.73	3.99	3.53	4.22
9	481	Hands, $D = 245760 = 480 \times 512$	?	1.27	0.95	2.88	3.92	2.56
10	7141	MNIST digit 3, $D = 784 = 28^2$	?	8.92	3.21	15.95	14.17	14.75
11	6824	MNIST digit 4, $D = 784 = 28^2$	?	8.13	3.07	14.44	9.54	13.16
12	6313	MNIST digit 5, $D = 784 = 28^2$	?	8.40	3.12	15.55	18	14.28

Our estimators in detail



The estimators $\hat{m}_k$ from [1], $E_{DP}(V)$ and $E_{CAP}(V)$ as a function of k.



Left: Estimated dimensions as a function of the true dimension d (solid black line). 5000 points from $N_d(0, I)$, $k = 20$. Right: Estimates as a function of the noise level σ. 5000 points from $U(0,1)^4 \times N_6(0, \sigma I)$, $k = 20$.

References

[1] E. Levina and P. Bickel. Maximum likelihood estimation of intrinsic dimension. In *Neural Information Processing Systems (NIPS)*, 2005.

[2] P. Grassberger and I. Procaccia. Measuring the strangeness of strange attractors. *Physica*, 9:189–208, 1983.

[3] K. Pettis, T. Bailey, A. Jain, and R. Dubes. An intrinsic dimensionality estimator from near-neighbor information. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 1(1):25–37, 1979.