Measures of distortion for Machine Learning

Leena Chennuru Vankadara

Theory of Machine Learning
Department of Computer Science
University of Tuebingen
&
Department of Computer Science
University of Hamburg

Master thesis

Supervised by Prof.Ulrike von Luxburg

Outline

Background

Consider the following problem:

- Let $(\mathcal{X}, d_x)$ be an arbitrary metric space and let $(\mathbb{R}^d, l_2)$ denote the Euclidean space of dimension d. Determine a value of d such that for any finite dataset $X = \{x_1, x_2, ..., x_n\}$ sampled from $(\mathcal{X}, d_x)$ according to some probability distribution $\mathcal{P}$, there exists a mapping $f : X \to \mathbb{R}^d$ such that the underlying metric is preserved i.e. $\forall i, j \in [n]$, $l_2(f(x_i), f(x_j)) = d_x(x_i, x_j)$.

Background

Consider the following problem:

- Let $(\mathcal{X}, d_x)$ be an arbitrary metric space and let $(\mathbb{R}^d, l_2)$ denote the Euclidean space of dimension d. Determine a value of d such that for any finite dataset $X = \{x_1, x_2, ..., x_n\}$ sampled from $(\mathcal{X}, d_x)$ according to some probability distribution $\mathcal{P}$, there exists a mapping $f : X \to \mathbb{R}^d$ such that the underlying metric is preserved i.e. $\forall i, j \in [n]$, $l_2(f(x_i), f(x_j)) = d_x(x_i, x_j)$.
- Does there **always** exist a solution to this problem ?

Example

Example

Let C_4 denote the 4-cycle and let d_G denote the shortest path metric. (C_4, d_G) can not be isometrically embedded into an Euclidean space no matter how high the dimension.

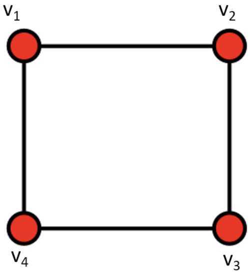

Figure: 4-cycle with the shortest path metric d_G

Example

Example

Let C_4 denote the 4-cycle and let d_G denote the shortest path metric. (C_4, d_G) can not be isometrically embedded into an Euclidean space no matter how high the dimension.

Proof.

$$\|f(v_1) - f(v_2) + f(v_3) - f(v_4)\|_2^2 \geq 0$$

$$l_2(f(v_1), f(v_3))^2 + l_2(f(v_2), f(v_4))^2 \leq l_2(f(v_1), f(v_2))^2 + l_2(f(v_2), f(v_3))^2 + l_2(f(v_3), f(v_4))^2 + l_2(f(v_1), f(v_4))^2$$

d_G on C_4 does not satisfy the same inequality!

□

Relaxed formulation

Consider the following relaxed formulation of the problem:

- Let $(\mathcal{X}, d_x)$ be an arbitrary metric space and let $(\mathbb{R}^d, l_2)$ denote the Euclidean space of dimension d. Determine a value of d such that for any finite dataset $X = \{x_1, x_2, ..., x_n\}$ sampled from $(\mathcal{X}, d_x)$ according to some probability distribution $\mathcal{P}$, there exists a mapping $f : X \to \mathbb{R}^d$ such that the distortion (measured by a meaningful measure of distortion) of the mapping f is bounded.

Relaxed formulation

Consider the following relaxed formulation of the problem:

- Let $(\mathcal{X}, d_x)$ be an arbitrary metric space and let $(\mathbb{R}^d, l_2)$ denote the Euclidean space of dimension d. Determine a value of d such that for any finite dataset $X = \{x_1, x_2, ..., x_n\}$ sampled from $(\mathcal{X}, d_x)$ according to some probability distribution $\mathcal{P}$, there exists a mapping $f : X \to \mathbb{R}^d$ such that the distortion (measured by a meaningful measure of distortion) of the mapping f is bounded.
- Does there **always** exist a solution to this problem ?

Notation and definitions

- (X, d_x) - arbitrary finite metric space (original space).
- $(\mathcal{Y}, d_y)$ - homogeneous and translation invariant metric space (target space).
- $f, g : (X, d_x) \rightarrow (\mathcal{Y}, d_y)$ - injective mappings.
- $\mathcal{P}$ - Probability distribution over X.
- $\Pi = \mathcal{P} \times \mathcal{P}$ - product distribution over $X \times X$.
- $\rho_f(u, v) = \frac{d_y(f(u), f(v))}{d_x(u,v)}$ $\forall (u, v) \in \binom{X}{2}$.
- For any $S \subset \binom{X}{2}$,
  $\Phi_{wc}(f_S) = \left(\max_{(u,v)\in S} \rho_f(u, v) \right) \cdot \left(\max_{(u,v)\in S} \frac{1}{\rho_f(u,v)} \right)$.
- $\forall u \in X$, $\mathrm{kNN}(u)$ denotes the set of k nearest neighbours of u.

Measures of distortion

- $\Phi_{wc}(f) := \left(\max_{u,v \in X, u \neq v} \rho_f(u,v) \right) \cdot \left(\max_{u,v \in X, u \neq v} \frac{1}{\rho_f(u,v)} \right)$

Measures of distortion

- $\Phi_{wc}(f) := \left(\max_{u,v \in X, u \neq v} \rho_f(u, v) \right) \cdot \left(\max_{u,v \in X, u \neq v} \frac{1}{\rho_f(u,v)} \right)$
- $\Phi_{avg}(f) := \frac{2}{n(n-1)} \sum_{u,v \in X, u \neq v} \rho_f(u, v)$

Measures of distortion

- $\Phi_{wc}(f) := \left(\max_{u,v\in X, u\neq v} \rho_f(u,v)\right) \cdot \left(\max_{u,v\in X, u\neq v} \frac{1}{\rho_f(u,v)}\right)$
- $\Phi_{avg}(f) := \frac{2}{n(n-1)} \sum_{u,v\in X, u\neq v} \rho_f(u,v)$
- $\Phi_{navg}(f) := \frac{2}{n(n-1)} \sum_{u\neq v\in X} \frac{\rho_f(u,v)}{\alpha}$ with $\alpha = \min_{u,v\in X, u\neq v} \rho_f(u,v)$

Measures of distortion

- $\Phi_{wc}(f) := \left(\max_{u,v\in X, u\neq v} \rho_f(u,v)\right) \cdot \left(\max_{u,v\in X, u\neq v} \frac{1}{\rho_f(u,v)}\right)$
- $\Phi_{avg}(f) := \frac{2}{n(n-1)} \sum_{u,v\in X, u\neq v} \rho_f(u,v)$
- $\Phi_{navg}(f) := \frac{2}{n(n-1)} \sum_{u\neq v\in X} \frac{\rho_f(u,v)}{\alpha}$ with $\alpha = \min_{u,v\in X, u\neq v} \rho_f(u,v)$
- $\Phi_{l_q}(f) := \mathbb{E}_{\Pi}(\rho_f(u,v)^q)^{\frac{1}{q}}$, (with $1 \leq q < \infty$):

Measures of distortion

- $\Phi_{wc}(f) := \left(\max_{u,v \in X, u \neq v} \rho_f(u,v) \right) \cdot \left(\max_{u,v \in X, u \neq v} \frac{1}{\rho_f(u,v)} \right)$
- $\Phi_{avg}(f) := \frac{2}{n(n-1)} \sum_{u,v \in X, u \neq v} \rho_f(u,v)$
- $\Phi_{navg}(f) := \frac{2}{n(n-1)} \sum_{u \neq v \in X} \frac{\rho_f(u,v)}{\alpha}$ with $\alpha = \min_{u,v \in X, u \neq v} \rho_f(u,v)$
- $\Phi_{l_q}(f) := \mathbb{E}_{\Pi}(\rho_f(u,v)^q)^{\frac{1}{q}}$, (with $1 \leq q < \infty$):
- $\Phi_{\epsilon}(f) := \min_{S \subset \binom{X}{2}, |S| \geq (1-\epsilon)\frac{n(n-1)}{2}} \Phi_{wc}(f_S)$, $0 < \epsilon < 1$

Measures of distortion

- $\Phi_{wc}(f) := \left(\max_{u,v \in X, u \neq v} \rho_f(u,v) \right) \cdot \left(\max_{u,v \in X, u \neq v} \frac{1}{\rho_f(u,v)} \right)$
- $\Phi_{avg}(f) := \frac{2}{n(n-1)} \sum_{u,v \in X, u \neq v} \rho_f(u,v)$
- $\Phi_{navg}(f) := \frac{2}{n(n-1)} \sum_{u \neq v \in X} \frac{\rho_f(u,v)}{\alpha}$ with $\alpha = \min_{u,v \in X, u \neq v} \rho_f(u,v)$
- $\Phi_{l_q}(f) := \mathbb{E}_{\Pi}(\rho_f(u,v)^q)^{\frac{1}{q}}$, (with $1 \leq q < \infty$):
- $\Phi_{\epsilon}(f) := \min_{S \subset \binom{X}{2}, |S| \geq (1-\epsilon)\frac{n(n-1)}{2}} \Phi_{wc}(f_S)$, $0 < \epsilon < 1$
- $\Phi_{klocal}(f) := \left(\max_{u \in X, v \in kNN(u)} \rho_f(u,v) \right) \cdot \left(\max_{u \in X, v \in kNN(u)} \frac{1}{\rho_f(u,v)} \right)$

Worstcase distortion

Consider the following problem:

- Let $(\mathcal{X}, d_x)$ be an arbitrary metric space and let $(\mathbb{R}^d, l_2)$ denote the Euclidean space of dimension d. Determine a value of d such that for any finite dataset $X = \{x_1, x_2, ..., x_n\}$ sampled from $(\mathcal{X}, d_x)$ according to some probability distribution $\mathcal{P}$, there exists a mapping $f : X \to \mathbb{R}^d$ such that the worstcase distortion (Φ_{wc}) of the mapping f is bounded.

Worstcase distortion

Consider the following problem:

- Let $(\mathcal{X}, d_x)$ be an arbitrary metric space and let $(\mathbb{R}^d, l_2)$ denote the Euclidean space of dimension d. Determine a value of d such that for any finite dataset $X = \{x_1, x_2, ..., x_n\}$ sampled from $(\mathcal{X}, d_x)$ according to some probability distribution $\mathcal{P}$, there exists a mapping $f : X \to \mathbb{R}^d$ such that the worstcase distortion (Φ_{wc}) of the mapping f is bounded.
- Does there **always** exist a solution to this problem ?

Impossibility results for wc distortion

Theorem (Bourgain 1985, Johnson and Lindenstrauss 1984)

Any finite metric space can be embedded into Euclidean space with dimension $O(\log n)$ and wc distortion $O(\log n)$

Impossibility results for wc distortion

Theorem (Bourgain 1985, Johnson and Lindenstrauss 1984)

Any finite metric space can be embedded into Euclidean space with dimension $O(\log n)$ and wc distortion $O(\log n)$

Theorem (Larsen and Nelson 2016)

For any integers $n, k \geq 2$ and for $1/(\min\{n,k\})^{0.4999} < \epsilon \leq 1$, there exists an n point subset of $\mathbb{R}^k$ such that any embedding in $(\mathbb{R}^d, l_2)$ that has wc distortion $1+\epsilon$ requires that $d = \Omega(\frac{\log(\epsilon^2 n)}{\epsilon^2})$.

Impossibility results for wc distortion

Theorem (Bourgain 1985, Johnson and Lindenstrauss 1984)

Any finite metric space can be embedded into Euclidean space with dimension $O(\log n)$ and wc distortion $O(\log n)$

Theorem (Larsen and Nelson 2016)

For any integers $n, k \geq 2$ and for $1/(\min\{n,k\})^{0.4999} < \epsilon \leq 1$, there exists an n point subset of $\mathbb{R}^k$ such that any embedding in $(\mathbb{R}^d, l_2)$ that has wc distortion $1+\epsilon$ requires that $d = \Omega(\frac{\log(\epsilon^2 n)}{\epsilon^2})$.

Theorem (Linial, London, and Rabinovich 1995)

Any embedding of an n vertex constant degree expander into Euclidean space requires that $\Phi_{wc} = \Omega(\log n)$.

Other measures of distortion

Theorem (Abraham, Bartal, and Neiman 2006)

For any arbitrary finite metric space there exists an embedding into $(\mathbb{R}^d, l_2)$ with average distortion $O(1)$, where $d = O(\log n)$. The worstcase distortion of this embedding is $O(\log n)$.

- Its probably unsurprising since we imposed no restrictions on the underlying metric space. For instance, if we consider worstcase distortion as our measure of distortion, we are essentially expecting bi-lipschitz equivalence between any discrete metric space and Euclidean space.

Growth restricted metrics

- Metric spaces of bounded intrinsic dimension.
- Dispels the volume argument.
- Doubling property is preserved under bi-lipschitz maps.
- Its a necessary condition for bi-lipschitz equivalence to Euclidean space.

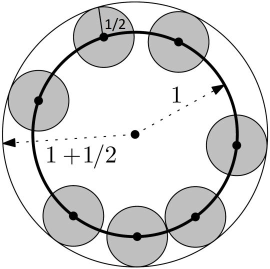

Figure: Equilateral space and the volume argument.

Doubling metrics

Theorem (Gupta, Krauthgamer, and Lee 2003)

There exists a family of metrics (L_k, d_G) which are uniformly doubling such that the minimum distortion required for an embedding f of (L_k, d_G) into any Euclidean space requires that the worst case distortion incurred by f is $\Omega(\sqrt{\log |L_k|})$.

Theorem (Abraham, Bartal, and Neiman 2006)

For any arbitrary finite doubling space (X, d_X) with cardinality n and doubling dimension λ, there exists an embedding $f : (X, d_X) \to (\mathbb{R}^d, l_2)$, where $d = O(\lambda \log \lambda)$ such that $\Phi_{avg} = O(1)$.

Summary so far and further questions...

- It is possible to achieve constant dimensional embeddings of uniformly doubling spaces with bounded average distortion but it is not possible to find such embeddings which can achieve bounded worstcase distortion.
- Similar results exist for other measures of distortion.
- Which of these results/measures of distortion are *meaningful* in the context of ML ?
- Rephrasing this, what are some of the properties that a measure of distortion needs to satisfy in order to be deemed as *meaningful* in the context of ML ?

Properties of distortion measures

Guiding principle

A good embedding preserves **most** distances as well as possible while better preserving the distances between pairs of points which could be critical for a given task.

- Basic properties
  - For any distortion function, irrespective of the context.
- Advances properties
  - Necessary characteristics of distortion measures in the context of ML.

Basic properties

Definition (Scale invariance)

Let $f : (X, d_X) \rightarrow (Y, d_Y)$ and $g : (X, d_X) \rightarrow (Y, d_Y)$ be two injective mappings. A distortion measure Φ is said to be scale invariant if $\exists \alpha \in \mathbb{R}, \forall u \in X, f(u) = \alpha g(u) \implies \Phi(f) = \Phi(g)$.

Definition (Translation invariance)

A measure of distortion Φ is said to be translation invariant if $\exists \alpha \in \mathbb{R}, \forall u \in X, f(u) = g(u) + \alpha; \implies \Phi(f) = \Phi(g)$.

Basic properties

Definition (**Monotonicity**)

Let $f : (X, d_X) \to (Y, d_Y)$ and $g : (X, d_X) \to (Y, d_Y)$ be embeddings. Define $\alpha(f) = (\frac{2}{n(n-1)}) \sum_{u \neq v \in X} \rho_f(u, v)$ as the scaling constant of f. Then a measure of distortion Φ is said to be monotonic if $\forall u, v \in X$:

$$\left(\left(\frac{\rho_f(u,v)}{\alpha(f)} \leq \frac{\rho_g(u,v)}{\alpha(g)} \leq 1\right) \text{ or } \left(\frac{\rho_f(u,v)}{\alpha(f)} \geq \frac{\rho_g(u,v)}{\alpha(g)} \geq 1\right)\right)$$

$$\implies \Phi(f) \geq \Phi(g)$$

Advanced properties - Robustness to outliers

Definition (Robustness to outliers in data)

Let $I : (X, d_X) \to (X, d_X)$ be an isometry. Fix arbitrary $x_0, x^* \in X$ and $\beta > 0$. For any $n \in \mathbb{N}$, let $X_n = \{x_1, x_2, ..., x_n\} \subset X \setminus B(x_0, \beta)$. Let $f_n : X_n \cup \{x_0\} \to X$ such that

$$f_n(x) = \begin{cases} x^*, & \text{if } x = x_0. \\ x, & \text{otherwise.} \end{cases} \quad (1)$$

We say that a measure of distortion Φ is not robust to outliers if $\lim_{n\to\infty} \Phi(f_n) \neq \lim_{n\to\infty} \Phi(I_n)$, where I_n denotes the restriction of the mapping I to $X_n \cup \{x_0\}$.

- Can be extended to distorting a constant order of the points.

Advanced properties - Robustness to outliers

Definition (Robustness to outliers in distances)

Let $I : (X, d_X) \to (X, d_X)$ be an isometry. Let $X_D = \{x_1, x_2, ..., \} \subset X$. Let $f : X_D \to X$ be an injective mapping such that there exists a $K \in \mathbb{N}$ such that $|G| < K$, where $G = \{(u, v) \in X_D \times X_D : d_X(f(u), f(v)) \neq d_X(u, v)\}$. For any $n \in \mathbb{N}$, let f_n and I_n denote the restriction of the mappings f and I, respectively, to $X_n = \{x_1, x_2, ..., x_n\} \subset X_D$. We say that a measure of distortion Φ is not robust to outliers if $\lim_{n\to\infty} \Phi(f_n) \neq \lim_{n\to\infty} \Phi(I_n)$.

Advanced Properties

Definition (Incorporation of a probability measure)

Let (X, d_X) be an arbitrary metric space. Let $X_n = \{x_1, x_2, ..., x_n\}$ be a finite subset of X. Let P_n denote a probability distribution on X_n. Fix arbitrary $x^*, y^* \in X_n$ such that $P_n(x^*) > P_n(y^*)$. Let $x^{'}, y^{'} \in X$ such that $\forall i \in [n], d_X(x_i, x^{'}) = \alpha_i d_X(x_i, x^*)$ and $d_X(x_i, y^{'}) = \alpha_i d_X(x_i, y^*)$. Let $f, g : X_n \to X$ be two embeddings such that:

$$f(x) = \begin{cases} x^{'}, & \text{if } x = x^*. \\ x, & \text{otherwise.} \end{cases} \quad , \quad g(x) = \begin{cases} y^{'}, & \text{if } x = y^*. \\ x, & \text{otherwise.} \end{cases}$$

Then a measure of distortion Φ is said to incorporate the probability distribution P_n if $\Phi(f) > \Phi(g)$.

Properties of existing distortion measures

Theorem (Properties of existing distortion measures)

For all choices of the parameters $1 \leq q < \infty$, $0 < \epsilon < 1$, $1 \leq k \leq n$, the following statements are true:

(a) Φ_{wc}, Φ_{avg}, Φ_{navg}, Φ_{l_q}, Φ_{ϵ} *and* Φ_{klocal} *satisfy the property of* **translation invariance**.

(b) Φ_{wc}, Φ_{navg}, Φ_{ϵ}, Φ_{klocal} *satisfy the properties of* **scale invariance** *and* **monotonicity.** Φ_{avg} *and* Φ_{l_q} *fail to satisfy these properties.*

(c) Φ_{ϵ}, Φ_{avg}, Φ_{l_q} *satisfy the property of* **robustness to outliers**. *The measures* Φ_{wc}, Φ_{navg}, Φ_{klocal} *fail to do so.*

(d) *The distortion measures* Φ_{wc}, Φ_{avg}, Φ_{navg}, Φ_{l_q}, Φ_{ϵ}, Φ_{klocal} *fail to* **incorporate a probability distribution** *defined on the data space.*

σ-distortion and properties

Definition (σ-distortion)

Let X_n be a finite subset of X. Given a distribution P_n over X_n, let $\Pi = P_n \times P_n$ denote the distribution on the product space $X_n \times X_n$. For any embedding f, let $\widetilde{\rho}_f(u, v)$ denote the normalized ratio of distances given by $\rho_f(u, v)/\sum_{(u,v)\in X\times X, u\neq v} \rho_f(u, v)$. The σ-distortion is then defined as

$$\mathbb{E}_{\Pi}(\widetilde{\rho}_f(u, v) - 1)^2. \tag{2}$$

σ-distortion and properties

Definition (σ-distortion)

Let X_n be a finite subset of X. Given a distribution P_n over X_n, let $\Pi = P_n \times P_n$ denote the distribution on the product space $X_n \times X_n$. For any embedding f, let $\widetilde{\rho}_f(u, v)$ denote the normalized ratio of distances given by $\rho_f(u, v) / \sum_{(u,v) \in X \times X, u \neq v} \rho_f(u, v)$. The σ-distortion is then defined as

$$\mathbb{E}_{\Pi}(\widetilde{\rho}_f(u, v) - 1)^2. \tag{2}$$

Theorem (Basic and advanced properties of σ-distortion)

The σ- distortion (a) is **invariant to scale and translations** *(b) satisfies the property of* **monotonicity***. (c) is* **robust to outliers in data** *and* **outliers in distances** *(d)* **incorporates a probability distribution** *into its evaluation.*

Euclidean representation with bounded σ-distortion

Consider the following problem:

- Let $(\mathcal{X}, d_x)$ be an arbitrary metric space and let $(\mathbb{R}^d, l_2)$ denote the Euclidean space of dimension d. Determine a value of d such that for any finite dataset $X = \{x_1, x_2, ..., x_n\}$ sampled from $(\mathcal{X}, d_x)$ according to some probability distribution $\mathcal{P}$, there exists a mapping $f : X \to \mathbb{R}^d$ such that the σ-distortion (Φ_σ) of the mapping f is bounded.

Euclidean representation with bounded σ-distortion

Consider the following problem:

- Let $(\mathcal{X}, d_x)$ be an arbitrary metric space and let $(\mathbb{R}^d, l_2)$ denote the Euclidean space of dimension d. Determine a value of d such that for any finite dataset $X = \{x_1, x_2, ..., x_n\}$ sampled from $(\mathcal{X}, d_x)$ according to some probability distribution $\mathcal{P}$, there exists a mapping $f : X \to \mathbb{R}^d$ such that the σ-distortion (Φ_σ) of the mapping f is bounded.
- Does there **always** exist a solution to this problem ?

Euclidean representation with bounded σ-distortion

Theorem (General metric spaces: constant distortion, log n dimensions)

Given any finite sample $X_n = \{x_1, x_2, ..., x_n\}$ generated by a probability distribution $\mathcal{P}$ on a metric space (X, d_X), for any $1 \leq p < \infty$ there exists an embedding $f : (X_n) \to l_p^D$, where $D = O(\log n)$ with $\sigma-$distortion $= O(1)$.

Euclidean representation with bounded σ-distortion

Theorem (General metric spaces: constant distortion, $\log n$ dimensions)

Given any finite sample $X_n = \{x_1, x_2, ..., x_n\}$ generated by a probability distribution $\mathcal{P}$ on a metric space (X, d_X), for any $1 \leq p < \infty$ there exists an embedding $f : (X_n) \to l_p^D$, where $D = O(\log n)$ with $\sigma-$distortion $= O(1)$.

Theorem (Doubling metric spaces: constant distortion, constant dimensions)

Given any finite sample $X_n = \{x_1, x_2, ..., x_n\}$ generated by a probability distribution $\mathcal{P}$ on a doubling metric space (X, d_X), for any $1 \leq p < \infty$ there exists an embedding $f : (X_n) \to l_p^D$, where $D = O(1)$ with $\sigma-$distortion $= O(1)$.

Experiments

Experiments

Distortion vs Embedding dimension

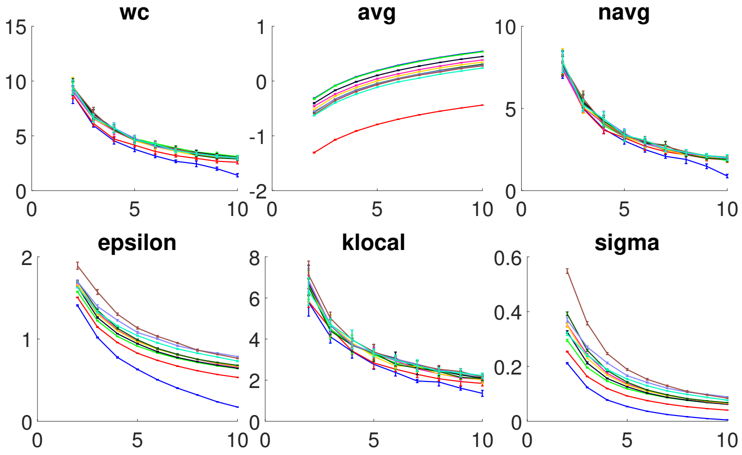

Figure: x-axis-dimension of the embedding space. Y-axis -distortion. Curves represent embeddings of normally distributed data in dimensions 10:10:100 generated by Isomap.

Distortion vs Original dimension

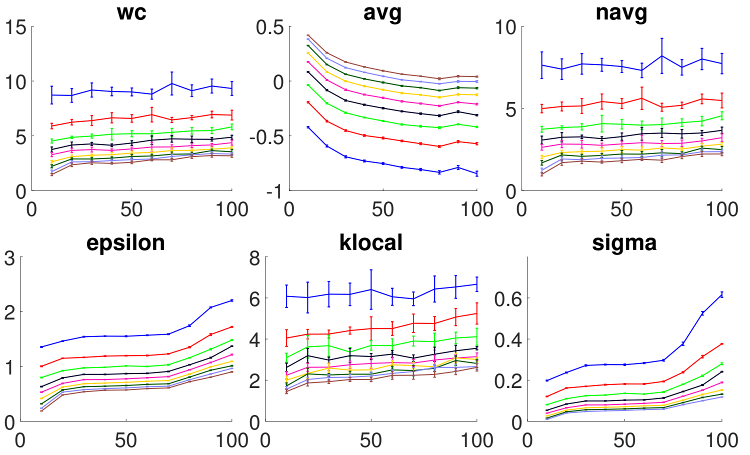

Figure: x-axis-dimension of the original space. Y-axis -distortion. Curves represent embeddings of gamma distributed data(a = 1.5, b = 4) from dimensions 10:10:100 generated by Isomap into a fixed dimension.

Effect of Noise

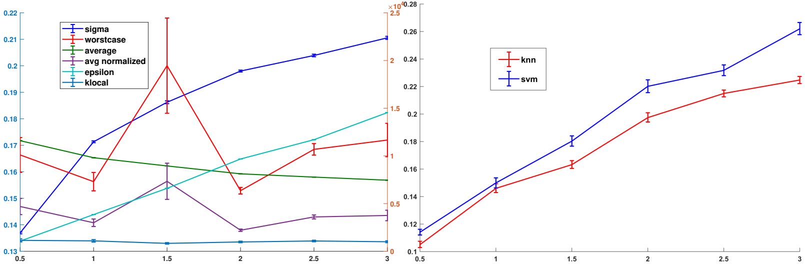

Figure: **Left**:Variance of noise(x-axis) vs distortion measures(y-axis).
Right: Variance of noise(x-axis) vs classification error(y-axis).

Distortion vs Classification accuracy

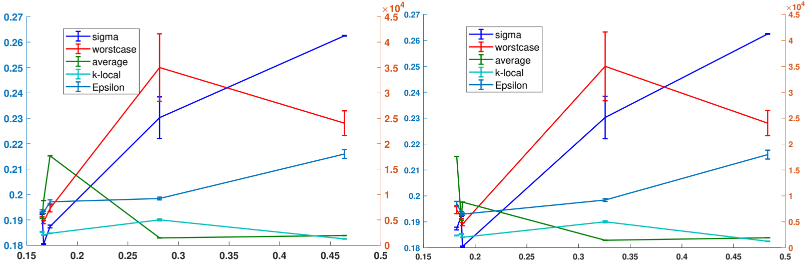

Figure: **Left**:SVM Classification error(x-axis) vs distortion measures(y-axis), **Right**: Knn Classification error(x-axis) vs distortion measures(y-axis).

Main contributions

- Initiated a systematic study of the desirable properties of distortion measures in the context of ML.

Main contributions

- Initiated a systematic study of the desirable properties of distortion measures in the context of ML.
- Showed that the existing measures of distortion are not well suited for our purpose.

Main contributions

- Initiated a systematic study of the desirable properties of distortion measures in the context of ML.
- Showed that the existing measures of distortion are not well suited for our purpose.
- Proposed a new measure of distortion which arguably is more suitable.

Main contributions

- Initiated a systematic study of the desirable properties of distortion measures in the context of ML.
- Showed that the existing measures of distortion are not well suited for our purpose.
- Proposed a new measure of distortion which arguably is more suitable.
- Showed that σ-distortion satisfies all the necessary properties.

Main contributions

- Initiated a systematic study of the desirable properties of distortion measures in the context of ML.
- Showed that the existing measures of distortion are not well suited for our purpose.
- Proposed a new measure of distortion which arguably is more suitable.
- Showed that σ-distortion satisfies all the necessary properties.
- Proved that given any doubling space, we can always find a Euclidean space such that any finite subset of the doubling space can be embedded into the Euclidean space with bounded σ-distortion.

Main contributions

- Initiated a systematic study of the desirable properties of distortion measures in the context of ML.
- Showed that the existing measures of distortion are not well suited for our purpose.
- Proposed a new measure of distortion which arguably is more suitable.
- Showed that σ-distortion satisfies all the necessary properties.
- Proved that given any doubling space, we can always find a Euclidean space such that any finite subset of the doubling space can be embedded into the Euclidean space with bounded σ-distortion.
- Provided with experimental evidence to support our theoretical results.

For Further Reading I

- Ittai Abraham, Yair Bartal, and Ofer Neiman. "Advances in metric embedding theory". In: *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*. ACM. 2006, pp. 271–286.
- Jean Bourgain. "On Lipschitz embedding of finite metric spaces in Hilbert space". In: *Israel Journal of Mathematics* 52.1 (1985), pp. 46–52.
- Anupam Gupta, Robert Krauthgamer, and James R Lee. "Bounded geometries, fractals, and low-distortion embeddings". In: *Foundations of Computer Science, 2003. Proceedings. 44th Annual IEEE Symposium on*. IEEE. 2003, pp. 534–543.

For Further Reading II

William B Johnson and Joram Lindenstrauss. "Extensions of Lipschitz mappings into a Hilbert space". In: *Contemporary mathematics* 26.189-206 (1984), p. 1.

Nathan Linial, Eran London, and Yuri Rabinovich. "The geometry of graphs and some of its algorithmic applications". In: *Combinatorica* 15.2 (1995), pp. 215–245.

Kasper Green Larsen and Jelani Nelson. "Optimality of the Johnson-Lindenstrauss Lemma". In: *arXiv preprint arXiv:1609.02094* (2016).