Measures of distortion for Machine learning

Leena Chennuru Vankadara, Ulrike von Luxburg

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

MAX-PLANCK-GESELLSCHAFT

Metric embeddings for Machine learning (ML)

Problem setup:

- Let $\mathbb{R}^d$ denote the Euclidean space of dimension d. Given a distance matrix, $D \in \mathbb{R}^{n\times n}$, containing distances, d_{ij}, between points x_i and x_j from an arbitrary metric space $(\mathcal{X})$, find an embedding $f : \mathcal{X} \to \mathbb{R}^d$ such that the distances d_{ij} are respected as well as possible.
- **Distortion measures:** Quantify the error made by such an embedding in distance preservation.

Worstcase distortion: $\Phi_{wc}(f) := \left(\max_{(x_i,x_j)\in\binom{\mathcal{X}}{2}} \frac{\|f(x_i)-f(x_j)\|}{d_{ij}}\right) \cdot \left(\max_{(x_i,x_j)\in\binom{\mathcal{X}}{2}} \frac{d_{ij}}{\|f(x_i)-f(x_j)\|}\right).$

ϵ-distortion $(\forall\, 0 < \epsilon < 1)$: $\Phi_\epsilon(f) := \min_{S\subset\binom{\mathcal{X}}{2}, |S|\geq(1-\epsilon)\binom{n}{2}} \Phi_{wc}(f_S)$, ($f_S \to$ restriction of f to S.)

k-local distortion: $\Phi_{klocal}(f) := \Phi_{wc}(f_S)$ where, $S = \{\{u, v\} \mid u, v \in X, v \in kNN(u)\}$.

Stress function: $Stress(f) := \left(\frac{\sum_{(x_i,x_j)\in\binom{\mathcal{X}}{2}} (d_{ij}-\|f(x_i)-f(x_j)\|)^2}{\sum_{(x_i,x_j)\in\binom{\mathcal{X}}{2}} d_{ij}^2}\right)^{1/2}$.

Which distortion measures are appropriate for ML?

Behavioural discrepancies across distortion measures:

- **Expected behaviour as suggested by a volume argument:** For a fixed embedding dimension, quality of an embedding ↓ as Original dimension ↑. **Observed behavior:** Many distortion measures deviate from expected behavior (See figure below).

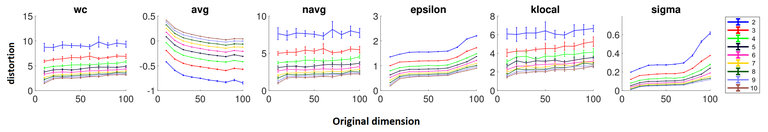


Figure 1: **Color of the curve:** dimension of the embedding space. **Datasets:** Gamma distributed data (a = 1.5, b = 4) of dimensions (10 : 10 : 100). **Embedding algorithm:** Isomap. Results for other distributions and algorithms look similar.

- **Question of interest for Learning theory:** Can we embed a metric space with "nice" underlying geometry into a constant dimensional Euclidean space with $\mathcal{O}(1)$ "**distortion**"?

Property/Distortion measure	wc	avg(l_q)	navg	k-local	ϵ(epsilon)
Constant distortion embeddings	✗	✓	?	✓	✓

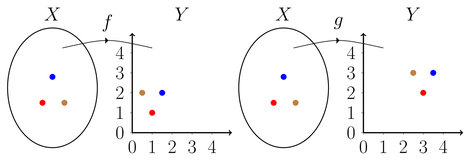
It's unclear which of the results/distortion measures are meaningful in the context of ML.

Desirable properties of distortion measures

$(X, d_X) \to$ arbitrary metric space, $(Y, d_Y) \to$ normed vector space, $\mathcal{P} \to$ probability distribution on X, $\Pi = \mathcal{P} \times \mathcal{P}$, $\mathcal{F} \to$ the space of all injective functions $(f : X \to Y)$. We refer to any function $\Phi : \mathcal{F} \to \mathbb{R}^+$ as a measure of distortion.

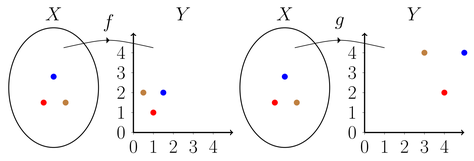
Basic properties:

a) **Translation invariance:**



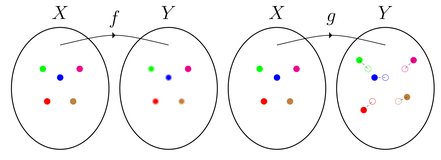
Φ is said to be translation invariant if $\Phi(f) = \Phi(g)$.

b) **Scale invariance:**



Φ is said to be invariant to scaling if $\Phi(f) = \Phi(g)$.

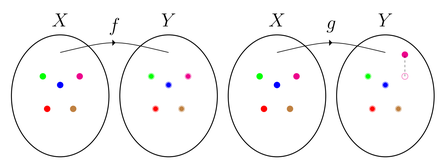
c) **Monotonicity:**



If f preserves distances better than g up to a scale, then $\Phi(g) \geq \Phi(f)$.

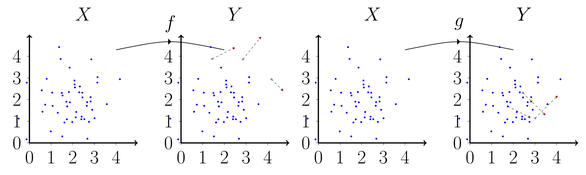
Advanced properties:

d) **Robustness to outliers in data and distances:**



The influence of a single data point or a distance value should be small $(\Phi(f) \approx \Phi(g))$.

e) **Incorporation of probability distribution:**



Distortions in higher denisty regions should be costlier than distortions in lower density regions $(\Phi(f) \leq \Phi(g))$.

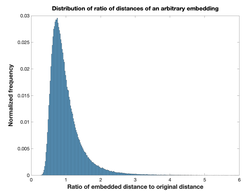
Nice to have property:

f) **Constant distortion embeddings:** If the underlying geometry of the metric space is "nice" (e.g, doubling metrics), then it would be nice to guarantee the existence of an embedding in constant dimensional Euclidean space with bounded (chosen) distortion.

σ-distortion

For any $u \neq v \in X$, let $\rho_f(u, v) = d_Y(f(u), f(v))/d_X(u, v)$ and let $\widetilde{\rho_f}(u, v) := \rho_f(u, v)/\mathcal{Z}$, where $\mathcal{Z} = \sum_{(u,v)\in\binom{X}{2}} \rho_f(u, v)/\binom{n}{2}$. The σ-distortion is then defined as $\mathbb{E}_\Pi(\widetilde{\rho}_f(u, v) - 1)^2$.

- Sharper concentration around 1 $\implies$ higher quality of an embedding.
- σ-distortion measures the variance of this distribution up to a scale.



Theoretical results

Property/Distortion measure	σ(sigma)	wc	avg(l_q)	navg	k-local	ϵ(epsilon)
Translation invariance	✓	✓	✓	✓	✓	✓
Monotonicity	✓	✓	✗	✓	✓	✓
Scale invariance	✓	✓	✗	✓	✓	✓
Robustness to outliers	✓	✗	✓	✗	✗	✓
Robustness to noise	✓	✗	✗	✗	✗	✓
Incorporation of probability	✓	✗	✗	✗	✗	✗
Constant distortion embeddings	✓	✗	✓	?	✓	✓

Experiments

Data generation process:

- For a dataset of dimension D, sample each coordinate independently from a specified 1-dimensional distribution. Gaussian distribution, Gamma distribution, Beta distribution, Gaussian mixture distribution, Laakso Space with many different parameter settings have been used.
- Embeddings - generated by Isomap, Maximum Variance Unfolding, Multidimensional Scaling, PCA , Probabilistic PCA, and Structure preserving embedding.

Distortion vs classification accuracy:

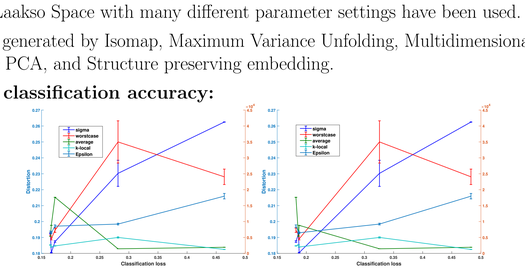


Figure 2: Classification error vs. distortion, for kNN (left) and Kernel SVM (right). **Dataset**: Mixture of gaussians in $\mathbb{R}^2$ with additive gaussian noise in $\mathbb{R}^{20}$. **Embedding algorithms:** PCA, GPLVM, Isomap, MVU, SPE. Each curve corresponds to a distortion measure as indicated in the legend. The distortions are scaled appropriately for visualization.

Distortion vs variance of noise:

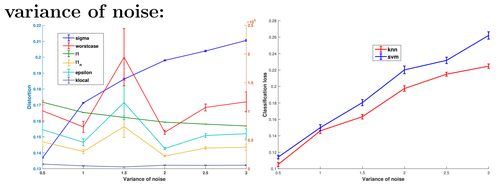


Figure 3: **Left:** Variance of noise vs distortion measures. The distortions are scaled appropriately for visualization. **Right:** Variance of noise vs. classification error. **Datasets:** Mixture of Gaussian data in $\mathbb{R}^2$ with additive Gaussian noise in $\mathbb{R}^{20}$ of increasing variance. **Embedding algorithm:** Isomap. The behavior corresponding to the other embedding algorithms is similar.

- σ-distortion acts as a better representative of the quality of an embedding since it satisfies all the aforementioned properties (Scale invariance, Robustness to outliers, etc).

Takeaway

- There are practical as well as theoretical behavioral discrepancies across various distortion measures.
- Need for a systematic study of desirable properties of measures of distortion for ML.
- Many existing distortion measures behave undesirably both in theory as well as in practice.
- σ-distortion appears to better resonate with the quality of an embedding for ML settings.
- Need for a more general systematic study of such unsupervised evaluation criterion.