Dimensionality Reduction:

Theoretical Perspective on Practical Measures

NeurIPS19

Nova Ora Fandina
Hebrew University, Israel
fandina@gmail.com

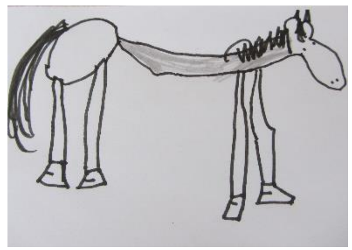
www.cs.huji.ac.il/~fandina

Joint work with

Yair Bartal, Hebrew University, Israel, yair@cs.huji.ac.il
Ofer Neiman, Ben Gurion University, Israel, neimano@cs.bgu.ac.il

Metric Dimensionality Reduction

Given a high-dimensional data set X embed it into a k-dimensional Y



with *small error* on the distances

How to do we measure the error?

How small the error can be?

Most of you probably use dim. red. in your work or research: PCA/MDS/Isomap…

- Visualization, clustering, similarity search, feature extraction and more…
- In ML: often used as a first step before applying further methods

The main message of this talk: *Use the random projection dim. reduction (JL),* as we proved that it has (asymptotically) smallest possible error on distances

What Is Error? In Practice

*A dimensionality reduction method is good if it has a small **average** error*

--- for each pair, compute an error and take the average over all pairs

The most basic and commonly used in practice average-case measures:

Multiplicative error: $dist_f(u,v)$

$$\boxed{\max\{expans_f(u,v), contr_f(u,v)\}}$$

Average distortion

$$\ell_q\text{- dist(f)} = \left(\frac{\sum_{u,v\in X}\left(dist_f(u,v)\right)^q}{\binom{n}{2}}\right)^{1/q}$$

σ-distortion

$$\left|\frac{expans_f(u,v)}{average\ of(expans_f(u,v))} - 1\right|^q$$

Additive error of a pair

$$\boxed{|d_{new}(u,v) - d_{old}(u,v)|}$$

Stress/Energy/REM Measures

normalized average of additive errors

Relative Error Measure

$$\left(\frac{|d_{new}(u,v) - d_{old}(u,v)|}{\min\{d_{new}(u,v), d_{old}(u,v)\}}\right)^q$$

In Practice *"A method is good if it has a small **average** error"*

- Widely used by practitioners to evaluate the quality of a method
- Various ***heuristics*** aim to minimize an average error [PCA/Isomap/MDS]
- No ***rigorous theoretical*** analysis for what can be achieved for each measure

In Theory *"A method is good if it has small* $\max_{u,v \in X} \{dis_f(u,v)\}$ *error"*

- [JL84] Johnosn-Lindenstrauss map

  Project an n-point Euclidean set into a random subspace of $\sim \log n / \epsilon^2$ dimesnions, then the *worst-case. error* is $1+\epsilon$

- No embeddings with small average error values

Bridging the Gap between Theory and Practice questions and results

The first theoretical analysis, almost tight upper and lower bounds for all measures

Theorem 1: Given an n-point Euclidean set, embed it into k-dims. with the JL map (Gaussian entries implementation). Then, with constant probability:

	$1 \le q < \sqrt{k}$	$\sqrt{k} \le q \le k$	$q = k$	$k \le q \le \infty$
ℓ_q-**dist**$(f) =$	$1 + O(1/\sqrt{k})$	$1 + O(q/k)$	$(\log n)^{O(1/k)}$	$n^{O(1/k - 1/q)}$

**Tight for $q \ge \sqrt{k}$*   **Phase transition: choose $k \gg q$*

$\mathbf{Stress_q \backslash Energy_q \backslash REM_q \backslash \sigma\text{-dist}} =$	$O(\sqrt{q/k})$	$1 \le q \le k$

**The bounds are tight for $q \ge 2$*   **Also holds simultaneously for all q*

Bridging the Gap between Theory and Practice questions and results

Approximating the optimal embedding [Computing an optimum is NP-hard]

**The $\boldsymbol{first}$ algorithm with theoretical guarantees on the approx. factor*

Theorem 2: For any finite metric space X, for $k \geq 3$ and $2 \leq q < k$, there is a random. poly-time algorithm that embeds X into k-dim. Euclidean space, s.t. with constant probability:

$$l_q\text{-dist}(F) = \left(\mathbf{1} + \boldsymbol{O}\left(\frac{\mathbf{1}}{\sqrt{\boldsymbol{k}}} + \frac{\boldsymbol{q}}{\boldsymbol{k}-\boldsymbol{q}}\right)\right) \cdot \boldsymbol{OPT} \qquad Measure_q(F) = \boldsymbol{O}(\mathbf{1}) \cdot \boldsymbol{OPT} + \boldsymbol{O}\left(\sqrt{\boldsymbol{q}/\boldsymbol{k}}\right)$$

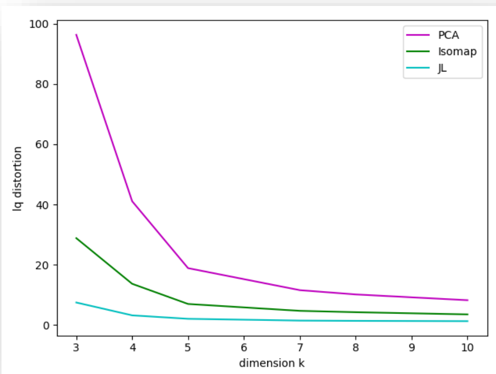
Proof: Convex Programming [computes an optimal embedding into high-dim.]

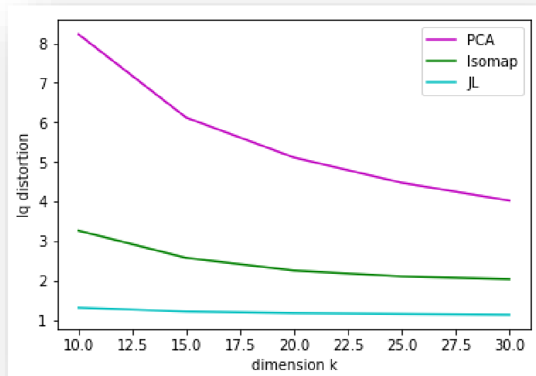
+ JL map [reduces the dimension into k]

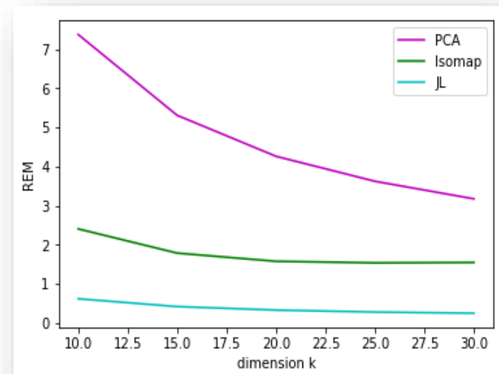
Empirical Experiments

Comparison of the JL based methods to the existing heuristics

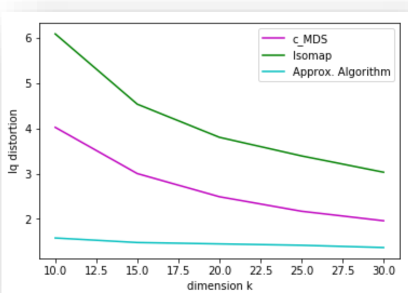
ℓ_q-distortion and REM: superiority of JL to Isomap and PCA $q = 5, k \in [3, 30]$







***Non-Euclidean input space*:** superiority of the JL-based method to Isomap and MDS



JL dramatically outperforms the other methods for all the range of values of *k!*

Take-Home Message: If you were scrolling through the news feed in Facebook thus far, here is the summary slide

- *JL transform is near optimal for (practical) dimensionality reduction*

- *A guidance on how to choose the target dimension k: pick $k \gg q$*

- *Algorithm for approximating an optimal embedding, with theoretical guarantees*

THANKS