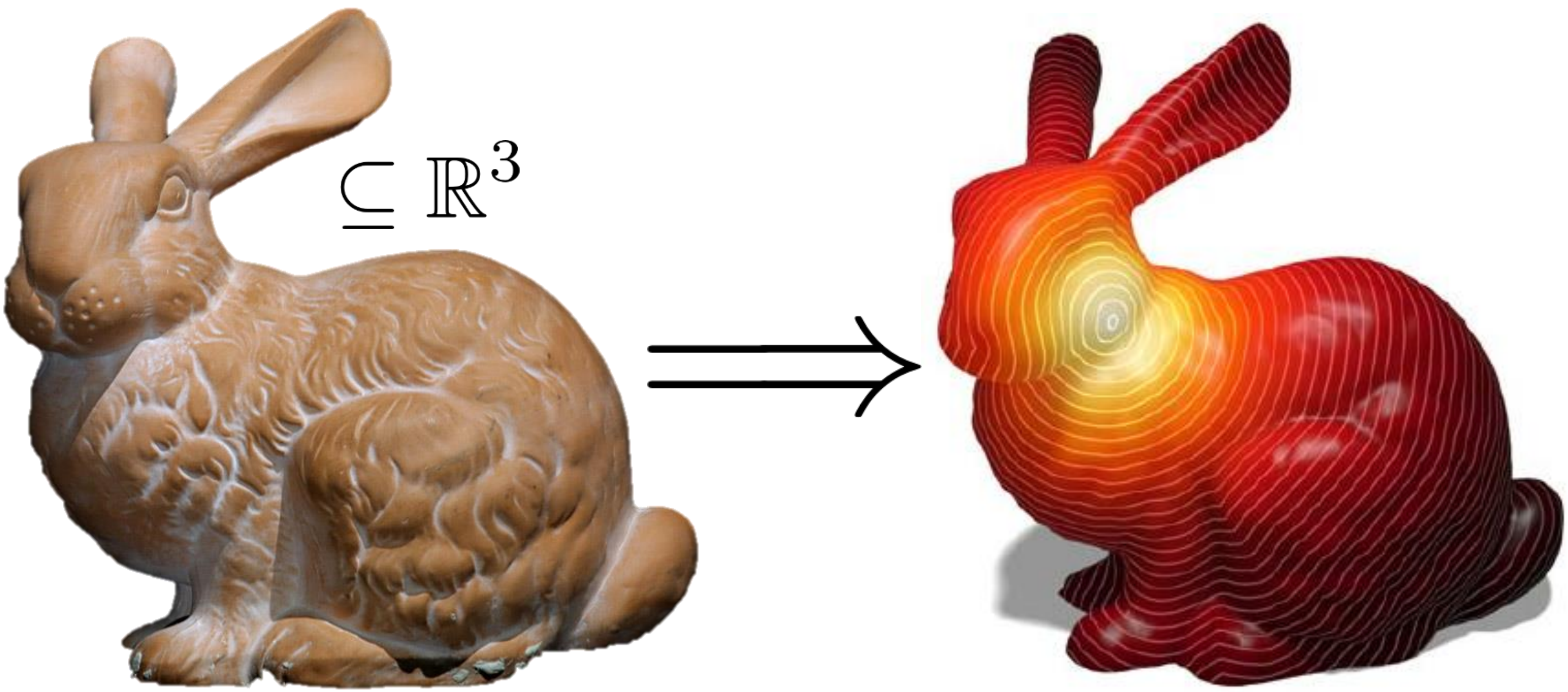
http://thepioneerwoman.com/food-and-friends/chocolate-swiss-roll-cake/

Inverse Distance Problems

Justin Solomon

MIT, Spring 2019

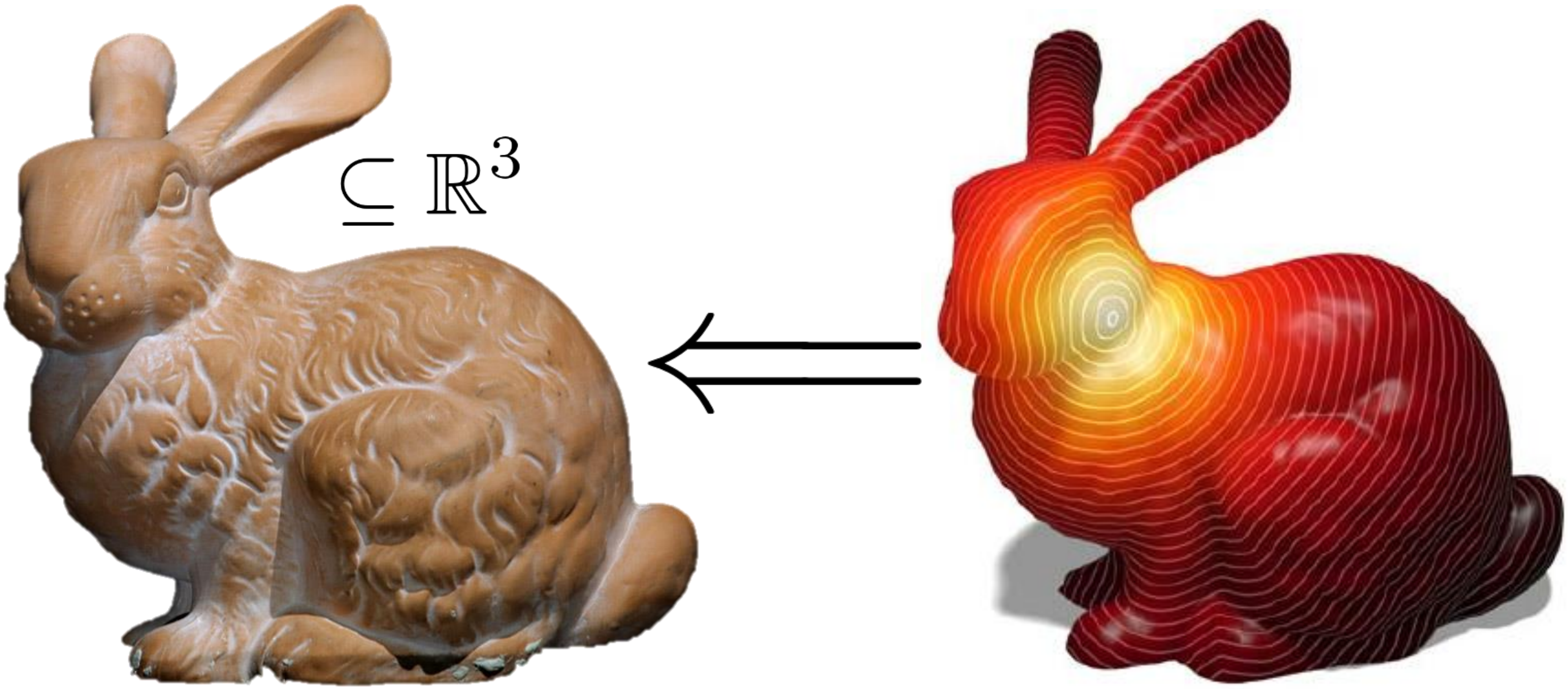
Last Time



Embedding **Geodesic distance**

Right bunny from "Geodesics in Heat" (Crane et al.)

Today



Embedding **Geodesic distance**

Right bunny from "Geodesics in Heat" (Crane et al.)

Many Names

- Dimensionality reduction
- Embedding
- Parameterization
- Manifold learning

…

Basic Task

Given pairwise distances extract an embedding.

Is it always possible?
What dimensionality?

Metric Space

Ordered pair (*M*, *d*) where *M* is a set and $d: M \times M \to \mathbb{R}$ satisfies

$$d(x,y) \geq 0$$

$$d(x,y) = 0 \iff x = y$$

$$d(x,y) = d(y,x)$$

$$d(x,z) \leq d(x,y) + d(y,z)$$

$$\forall x, y, z \in M$$

Many Examples of Metric Spaces

$$\mathbb{R}^n, d(x, y) := \|x - y\|_p$$

$$S \subset \mathbb{R}^3, d(x, y) := \text{geodesic}$$

$$C^\infty(\mathbb{R}), d(f, g)^2 := \int_{\mathbb{R}} (f(x) - g(x))^2 \, dx$$

Isometry [ahy-**som**-i-tree]**:**

A map between metric spaces that preserves pairwise distances.

Q:

Can you **always embed** a metric space isometrically in $\mathbb{R}^n$?

Q:

Can you always embed a finite metric space isometrically in $\mathbb{R}^n$?

Disappointing Example

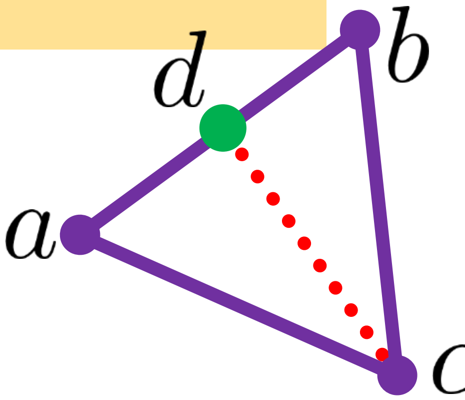
$$X := \{a, b, c, d\}$$

$$d(a, d) = d(b, d) = 1$$

$$d(a, b) = d(a, c) = d(b, c) = 2$$

$$d(c, d) = 1.5$$

Cannot be embedded in Euclidean space!



https://chiasme.wordpress.com/2013/10/07/when-does-a-finite-metric-space-embed-isometrically-into-an-euclidean-space/

Contrasting Example

$$\ell_\infty(\mathbb{R}^n) := (\mathbb{R}^n, \|\cdot\|_\infty)$$
$$\|\mathbf{x}\|_\infty := \max_k |\mathbf{x}_k|$$

Proposition. Every finite metric space embeds isometrically into $\ell_\infty(\mathbb{R}^n)$ for some n.

Extends to infinite-dimensional spaces!

Fréchet Embedding

Definition 7.3 (Fréchet embedding). *Suppose* (M, d) *is a metric space that* $S_1, \ldots, S_r \subseteq M$. *We define the* Fréchet embedding *of* M *with respect to* $\{S_1, \ldots, S_r\}$ *to be the map* $\phi : M \to \mathbb{R}^r$ *given by*

$$\phi(x) := (d(x, S_1), d(x, S_2), \ldots, d(x, S_r)), \tag{7.2}$$

where $d(x, S) := \min_{y \in S} d(x, y)$.

Approximate Embedding

$$\text{expansion}(f) := \max_{x,y} \frac{\mu(f(x), f(y))}{\rho(x, y)}$$

$$\text{contraction}(f) := \max_{x,y} \frac{\rho(x, y)}{\mu(f(x), f(y))}$$

$$\text{distortion}(f) := \text{expansion}(f) \times \text{contraction}(f)$$

http://www.cs.toronto.edu/~avner/teaching/S6-2414/LN1.pdf

Well-Known Result

Proposition 7.2 (Bourgain's Theorem). *Suppose (M, d) is a metric space consisting of n points, that is, $|M| = n$. Then, for $p \geq 1$, M embeds into $\ell_p(\mathbb{R}^m)$ with $O(\log n)$ distortion, where $m = O(\log^2 n)$.*

Matousek improved the distortion bound to $\log n / p$ [14].

$m := 576 \log n)$
for $j = 1$ **to** $\log n$ **do** /* levels of density */
  for $i = 1$ **to** m **do** /* repeat for high probability */
    choose set S_{ij} by sampling each node in X
    independently with probability 2^{-j}
  end
end
$f_{ij}(x) := d(x, S_{ij})$
$f(x) := \bigoplus_{j=1}^{\log n} \bigoplus_{i=1}^{m} f_{ij}(x)$

Uses Fréchet embedding

http://ttic.uchicago.edu/~harry/teaching/pdf/lecture3.pdf

Euclidean Problem

Given:

$$P_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2^2, P \in \mathbb{R}^{n \times n}$$

Reconstruct:

$$\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^m$$

Alternative notation:

$$X \in \mathbb{R}^{m \times n}$$

Gram Matrix [gram **mey**-triks]**:**

A matrix of inner products

$$X^\top X$$

Classical Multidimensional Scaling

1. Double centering: $G := -\frac{1}{2} J^\top P J$
   Centering matrix $J := I_{n \times n} - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$

2. Find m largest eigenvalues/eigenvectors
   $G = Q \Lambda Q^\top$

3. $\overline{X} = \sqrt{\Lambda} Q^\top$

"MDS"

Extension: Landmark MDS

Torgerson, Warren S. (1958). *Theory & Methods of Scaling*.

Visualization

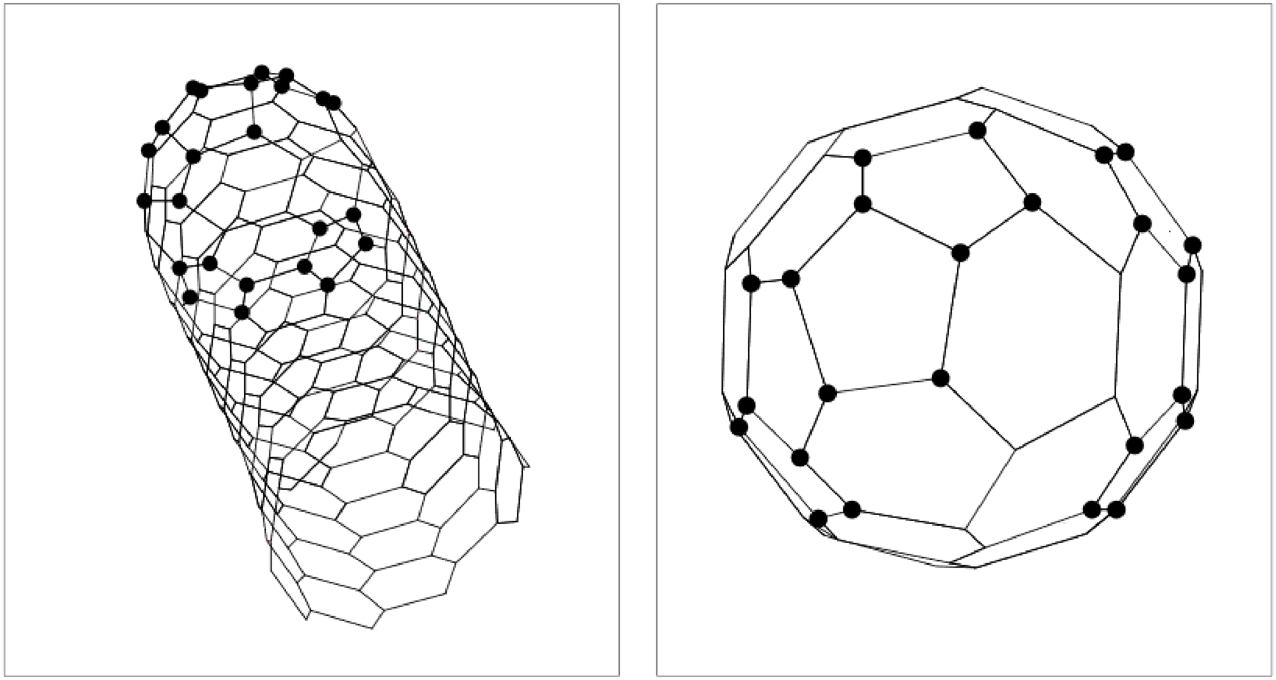


Figure 10: *Nanotube Embedding. One of Asimov's graphs for a nanotube is rendered with MDS in 3-D (Stress=0.06). The nodes represent carbon atoms, the lines represent chemical bonds. The right hand frame shows the cap of the tube only. The highlighted points show some of the pentagons that are necessary for forming the cap.*

http://www.stat.yale.edu/~lc436/papers/JCGS-mds.pdf

Stress Majorization

$$\min_{X} \sum_{ij} \left(D_{0ij} - \|\mathbf{x}_i - \mathbf{x}_j\|_2\right)^2$$

Nonconvex!

SMACOF:

Scaling by Majorizing a Complicated Function

de Leeuw, J. (1977), "Applications of convex analysis to multidimensional scaling" *Recent developments in statistics*, 133–145.

SMACOF Potential Terms

$$\min_X \sum_{ij} \left(D_{0ij} - \|\mathbf{x}_i - \mathbf{x}_j\|_2\right)^2$$

$$\sum_{ij} (D_{0ij})^2 = \text{const.}$$

$$\sum_{ij} \|\mathbf{x}_i - \mathbf{x}_j\|_2^2 = \text{tr}(XVX^\top), \text{ where } V = 2nJ$$

$$-2\sum_{ij} D_{0ij}\|\mathbf{x}_i - \mathbf{x}_j\|_2 = -2\text{tr}(XB(X)X^\top)$$

$$\text{where } B_{ij}(X) := \begin{cases} -\frac{2D_{0ij}}{\|\mathbf{x}_i - \mathbf{x}_j\|_2} & \text{if } \mathbf{x}_i \neq \mathbf{x}_j, i \neq j \\ 0 & \text{if } \mathbf{x}_i = \mathbf{x}_j, i \neq j \\ -\sum_{j \neq i} B_{ij} & \text{if } i = j \end{cases}$$

SMACOF Lemma

$$\sum_{ij} (D_{0ij})^2 = \text{const.}$$

$$\sum_{ij} \|\mathbf{x}_i - \mathbf{x}_j\|_2^2 = \text{tr}(XVX^\top), \text{ where } V = 2nJ$$

$$-2\sum_{ij} D_{0ij} \|\mathbf{x}_i - \mathbf{x}_j\|_2 = -2\text{tr}(XB(X)X^\top)$$

$$\text{where } B_{ij}(X) := \begin{cases} -\frac{2D_{0ij}}{\|\mathbf{x}_i - \mathbf{x}_j\|_2} & \text{if } \mathbf{x}_i \neq \mathbf{x}_j, i \neq j \\ 0 & \text{if } \mathbf{x}_i = \mathbf{x}_j, i \neq j \\ -\sum_{j \neq i} B_{ij} & \text{if } i = j \end{cases}$$

Lemma. Define

$$\tau(X, Z) := \text{const.} + \text{tr}(XVX^\top) - 2\text{tr}(XB(Z)Z^\top)$$

Then,

$$\tau(X, X) \leq \tau(X, Z) \; \forall Z$$

with equality exactly when $X \propto Z$.

Proof on board using Cauchy-Schwarz.

See *Modern Multidimensional Scaling* (Borg, Groenen)

SMACOF: Single Step

$$X^{k+1} \leftarrow \min_X \tau(X, X^k)$$

$$\tau(X, Z) := \text{const.} + \text{tr}(XVX^\top) - 2\text{tr}(XB(Z)Z^\top)$$

$$\implies 0 = \nabla_X[\tau(X, X^k)]$$

$$= 2XV - 2X^k B(X^k)$$

$$\implies X^{k+1} = X^k B(X^k) \left(I_{n\times n} - \frac{\mathbf{1}\mathbf{1}^\top}{n} \right)$$

Majorization-Minimization (MM) algorithm

Objective convergence:
$\tau(X^{k+1}, X^{k+1}) \leq \tau(X^k, X^k)$

Visualization

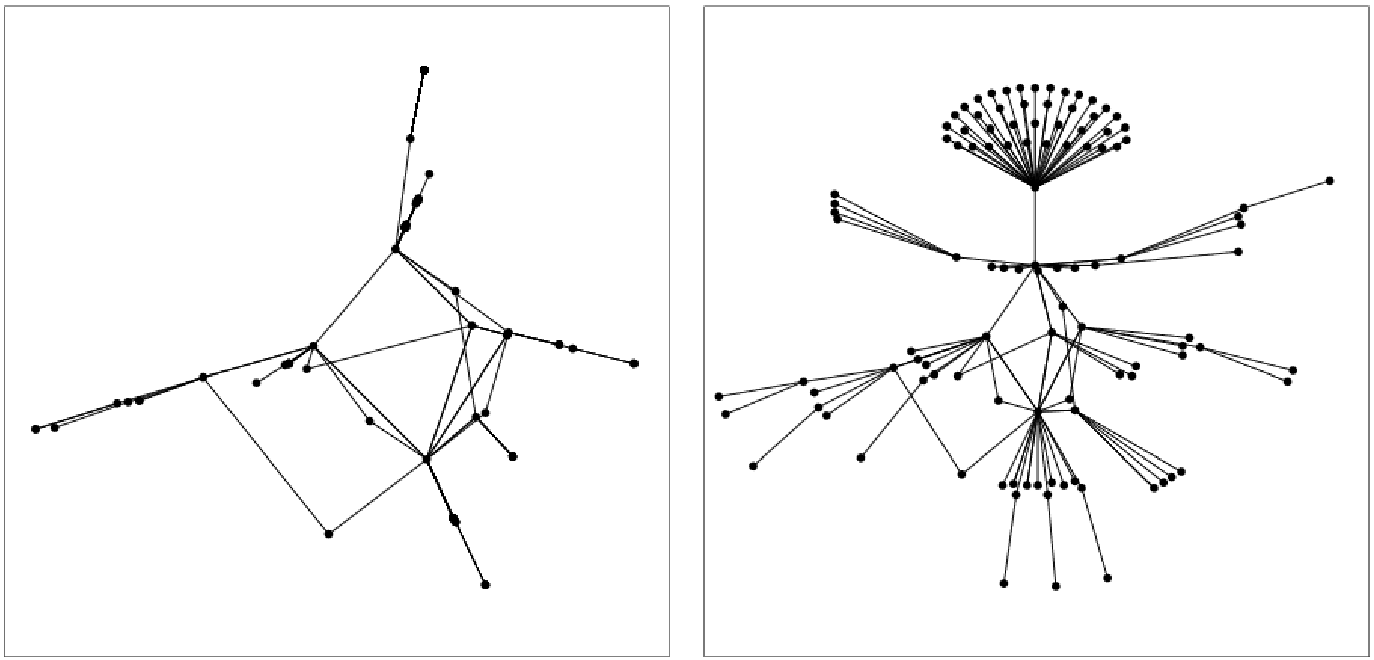


Figure 9: *A Telephone Call Graph, Layed Out in 2-D. Left: classical scaling (Stress=0.34); right: distance scaling (Stress=0.23). The nodes represent telephone numbers, the edges represent the existence of a call between two telephone numbers in a given time period.*

http://www.stat.yale.edu/~lc436/papers/JCGS-mds.pdf

Recent SMACOF Application

DOI: 10.1111/cgf.12558
EUROGRAPHICS 2015 / O. Sorkine-Hornung and M. Wimmer (Guest Editors)
Volume 34 (2015), Number 2

Shape-from-Operator: Recovering Shapes from Intrinsic Operators

Davide Boscaini, Davide Eynard, Drosos Kourounis, and Michael M. Bronstein

Università della Svizzera Italiana (USI), Lugano, Switzerland

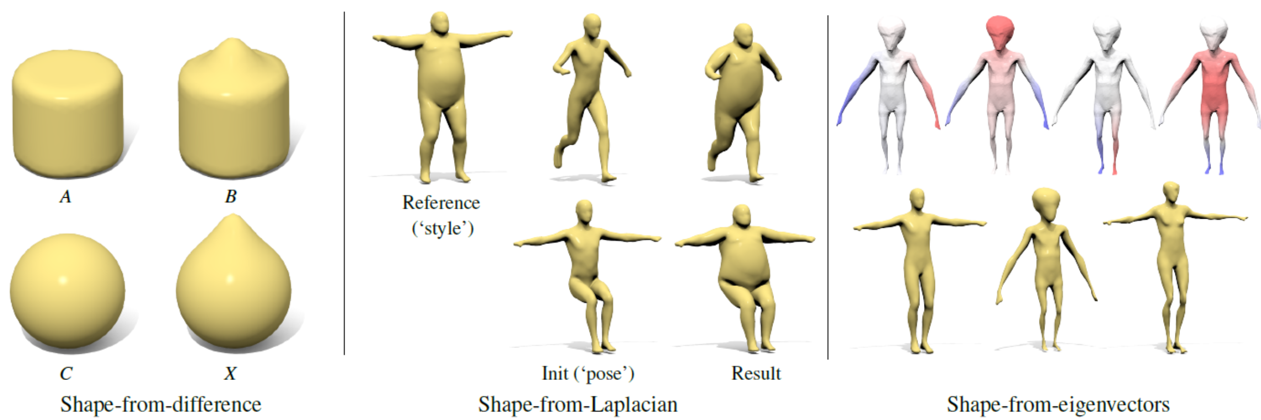
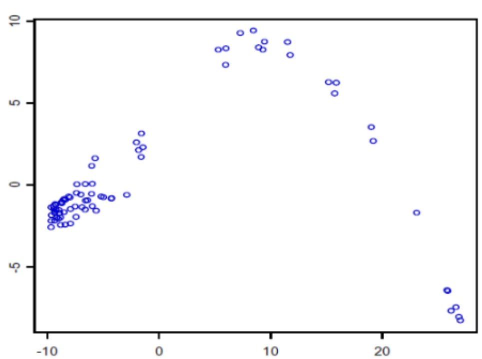


Figure 1: *Examples of three different shape-from-operator problems considered in the paper. Left: shape analogy synthesis as shape-from-difference operator problem (shape X is synthesized such that the intrinsic difference operator between C,X is as close as possible to the difference between A,B). Center: style transfer as shape-from-Laplacian problem. The Laplacian of the*

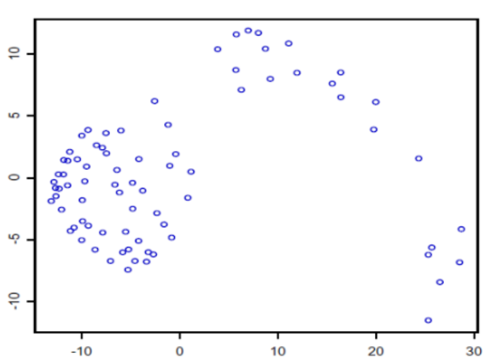
Related Method

$$\min_{X} \sum_{ij} \frac{\left(D_{0ij} - \|\mathbf{x}_i - \mathbf{x}_j\|_2\right)^2}{D_{0ij}}$$

Cares more about preserving small distances



Classical MDS



Sammon

"Sammon mapping"

Sammon (1969). "A nonlinear mapping for data structure analysis." IEEE Transactions on Computers 18.

http://www.stat.pitt.edu/sungkyu/course/2221Fall13/lec8_mds_combined.pdf

Intrinsic-to-Extrinsic: Theory

Theorem 7.1 (Whitney embedding theorem). *Any smooth, real k-dimensional manifold maps smoothly into* $\mathbb{R}^{2k}$.

Theorem 7.2 (Nash–Kuiper embedding theorem, simplified). *Any k-dimensional Riemannian manifold admits an isometric, differentiable embedding into* $\mathbb{R}^{2k}$.

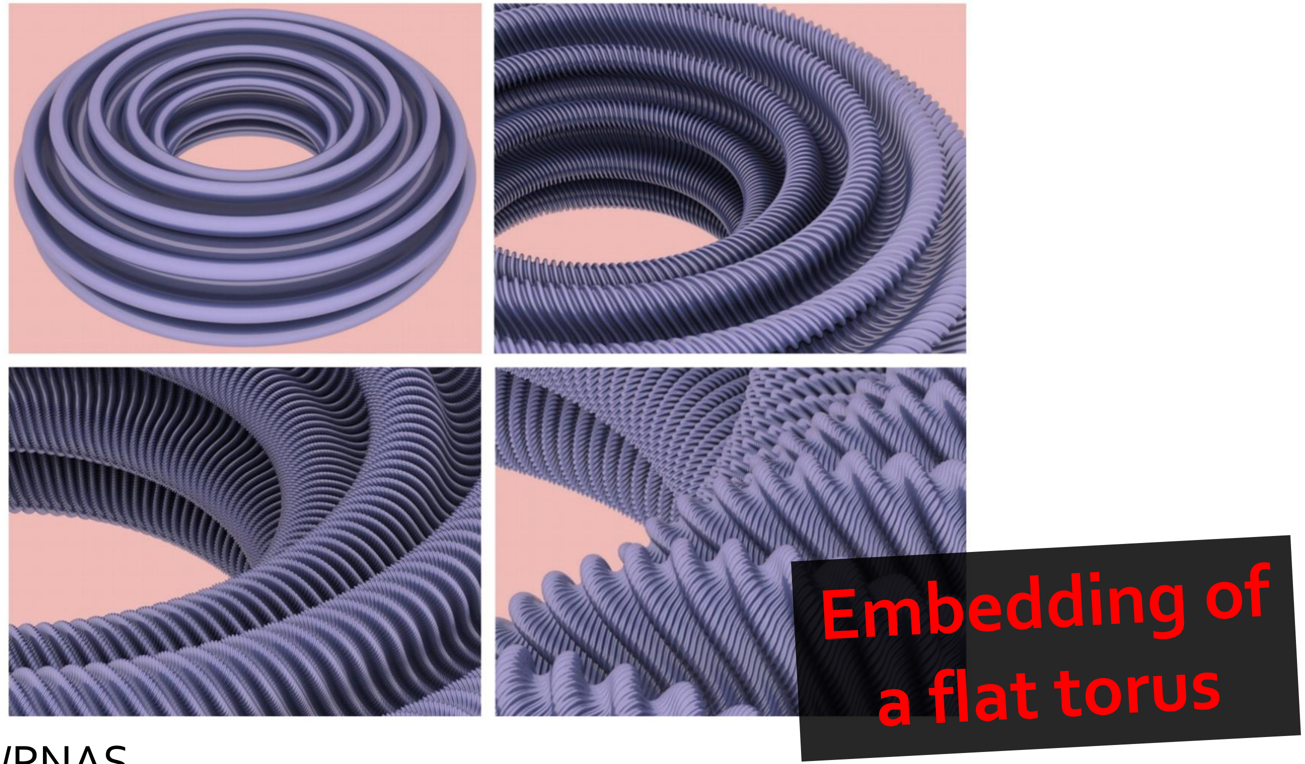


Image: HEVEA Project/PNAS

Intrinsic-to-Extrinsic: ISOMAP

- **Construct neighborhood graph**

  ***k*-nearest neighbor graph or *ε*-neighborhood graph**

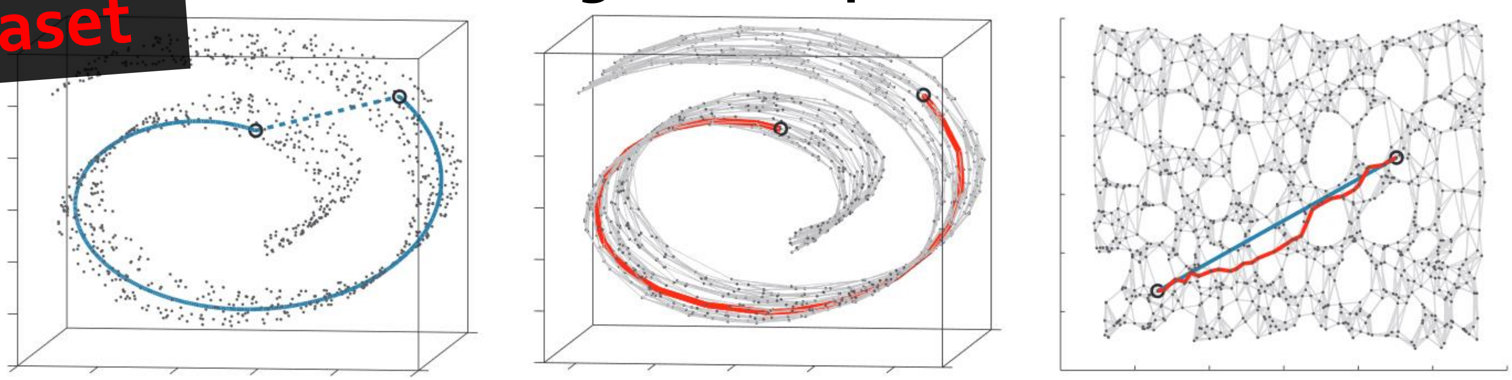
- **Compute shortest-path distances**

  Floyd-Warshall algorithm or Dijkstra

- **Classical MDS**

  Eigenvalue problem

Swiss roll dataset



Tenenbaum, de Silva, Langford.
"A Global Geometric Framework for Nonlinear Dimensionality Reduction." Science (2000).

Floyd-Warshall Algorithm

```
let dist be a |V| × |V| array of minimum distances initialized to ∞ (infinity)
for each vertex v
   dist[v][v] ← 0
for each edge (u,v)
   dist[u][v] ← w(u,v)  // the weight of the edge (u,v)
for k from 1 to |V|
   for i from 1 to |V|
      for j from 1 to |V|
         if dist[i][j] > dist[i][k] + dist[k][j]
             dist[i][j] ← dist[i][k] + dist[k][j]
         end if
```

https://en.wikipedia.org/wiki/Floyd%E2%80%93Warshall_algorithm

Landmark ISOMAP

- **Construct neighborhood graph**

  k-nearest neighbor graph or *ε*-neighborhood graph

- **Compute some shortest-path distances**

  Dijkstra: $O(kn\,N \log N)$, *n* landmarks, *N* points

- **MDS on landmarks**

  Smaller $n \times n$ problem

- **Closed-form embedding formula**

  $\delta(x)$ vector of squared distances from *x* to landmarks

$$\text{Embedding}(x)_i = -\frac{1}{2}\frac{v_i^\top}{\sqrt{\lambda_i}}\left(\delta(x) - \delta_{\text{average}}\right)$$

Landmark MDS

Locally Linear Embedding (LLE)

- **Construct neighborhood graph**

k-nearest neighbor graph or *ε*-neighborhood graph

- **Analysis step: Compute weights W_{ij}**

$$\min_{\omega^1,\ldots,\omega^k} \left\|\mathbf{x}_i - \sum_j \omega^j \mathbf{n}_j\right\|_2$$
$$\text{subject to} \quad \sum_j \omega^j = 1$$

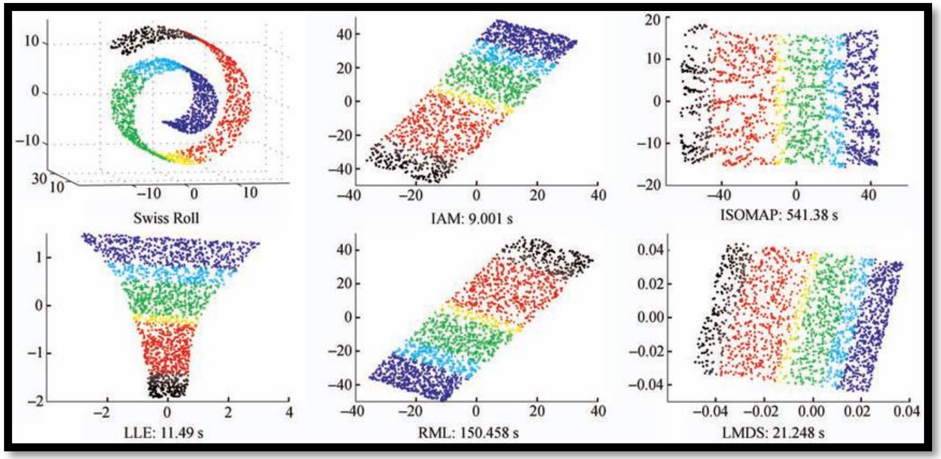
- **Embedding step: Minimum eigenvalue problem**

$$\begin{aligned} \min_Y \quad & \|Y - YW^\top\|^2_{\text{Fro}} \\ \text{subject to} \quad & YY^\top = I_{p\times p} \\ & Y\mathbf{1} = \mathbf{0} \end{aligned}$$

Derive on board

Comparison: ISOMAP vs. LLE

ISOMAP	LLE
Global distances	Local averaging
k-NN graph distances	*k*-NN graph weighting
Largest eigenvectors	Smallest eigenvectors
Dense matrix	Sparse matrix



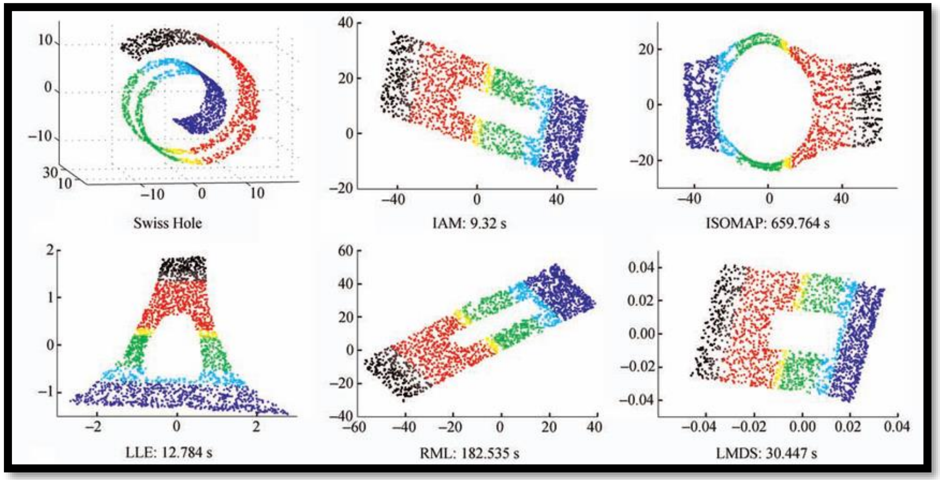


Image from "Incremental Alignment Manifold Learning." Han et al. JCST 26.1 (2011).

Other option:

Diffusion Maps

- Construct similarity matrix

Example: $K(x,y) := e^{-\|x-y\|^2/\varepsilon}$

- Normalize rows

$$M := D^{-1}K$$

- Embed from *k* largest eigenvectors

$$(\lambda_1\psi_1, \lambda_2\psi_2, \ldots, \lambda_k\psi_k)$$

(more later)

Coifman, R.R.; S. Lafon. (2006). "Diffusion maps." *Applied and Computational Harmonic Analysis*. 21: 5–30.

Embedding from Geodesic Distance

On reconstruction of non-rigid shapes with intrinsic regularization

Yohai S. Devir Guy Rosman Alexander M. Bronstein Michael M. Bronstein
Ron Kimmel
{yd|rosman|bron|mbron|ron}@cs.technion.ac.il
Department of Computer Science
Technion - Israel Institute of Technology

Abstract

Shape-from-X *is a generic type of inverse problems in computer vision, in which a shape is reconstructed from some measurements. A specially challenging setting of this problem is the case in which the reconstructed shapes are non-rigid. In this paper, we propose a framework for intrinsic regularization of such problems. The assumption is that we have the geometric structure of a shape which is intrinsically (up to bending) similar to the one we would like to reconstruct. For that goal, we formulate a variation with respect to vertex coordinates of a triangulated mesh approximating the continuous shape. The numerical core of the proposed method is based on differentiating the fast marching update step for geodesic distance computation.*

1. Introduction

In many tasks, both in human and computer vision, one tries to deduce the shape of an object given an observa-

many other problems, in which an object is reconstructed based on some measurement, are known as *shape reconstruction problems*. They are a subset of what is called *inverse problems*. Most such inverse problems are underdetermined, in the sense that measuring different objects may yield similar measurements. Thus, in the above illustration, the essence of the shadow theater is that it is hard to distinguish between shadows cast by an animal and shadows cast by hands. Therefore prior knowledge about the unknown object is needed.

Of particular interest are reconstruction problems involving non-rigid shapes. The world surrounding us is full with objects such as live bodies, paper products, plants, clothes etc., which may be deformed to different postures. These objects may be deformed to an infinite number of different postures. While bending, though, objects tends to preserve their internal geometric structure. Two objects differing by a bending are said to be *intrinsically similar*. In many cases, while we do not know the measured object, we have a prior on its intrinsic geometry. For example, in the shadow theater, though we do not know which exact posture of the hand

Take-Away

Huge zoo
of embedding techniques.

Each with different theoretical properties: Try them all!

But what if the distance matrix is incomplete or noisy?

More General: Metric Nearness

$$\min_{X \in \mathcal{M}_{N\times N}} \|X - D\|_{\text{Fro}}^2$$

TRIANGLE_FIXING(D, ϵ)
Input: Input dissimilarity matrix D, tolerance ϵ
Output: $M = \text{argmin}_{X \in \mathscr{M}_N} \|X - D\|_2$.
for $1 \le i < j < k \le n$
  $(z_{ijk}, z_{jki}, z_{kij}) \leftarrow 0$
for $1 \le i < j \le n$
  $e_{ij} \leftarrow 0$
$\delta \leftarrow 1 + \epsilon$
while $(\delta > \epsilon)$ {convergence test}
  foreach triangle (i, j, k)
    $b \leftarrow d_{ki} + d_{jk} - d_{ij}$
    $\mu \leftarrow \frac{1}{3}(e_{ij} - e_{jk} - e_{ki} - b)$
    $\theta \leftarrow \min\{-\mu, z_{ijk}\}$ {Stay within half-space of constraint}
    $e_{ij} \leftarrow e_{ij} - \theta$, $e_{jk} \leftarrow e_{jk} + \theta$, $e_{ki} \leftarrow e_{ki} + \theta$
    $z_{ijk} \leftarrow z_{ijk} - \theta$ {Update correction term}
  end *foreach*
  $\delta \leftarrow$ sum of changes in the e
end *while*
return $M = D + E$

In other words, the vector e is projected orthogonally onto the constraint set $\{e' : e'_{ij} - e'_{jk} - e'_{ki} \le b_{ijk}\}$. This is tantamount to solving

$$\min_{e'} \ \tfrac{1}{2}\left[(e'_{ij} - e_{ij})^2 + (e'_{jk} - e_{jk})^2 + (e'_{ki} - e_{ki})^2)\right], \quad \text{subject to} \quad e'_{ij} - e'_{jk} - e'_{ki} = b_{ijk}. \tag{3.2}$$

It is easy to check that the solution is given by

$$e'_{ij} \leftarrow e_{ij} - \mu_{ijk}, \qquad e'_{jk} \leftarrow e_{jk} + \mu_{ijk}, \qquad \text{and} \qquad e'_{ki} \leftarrow e_{ki} + \mu_{ijk}, \tag{3.3}$$

where $\mu_{ijk} = \frac{1}{3}(e_{ij} - e_{jk} - e_{ki} - b_{ijk}) > 0$.

Iterative projection

Dhillon, Sra, Tropp. "Triangle Fixing Algorithms for the Metric Nearness Problem." NIPS.

Euclidean Matrix Completion

$$\min_{G} \|H \circ (\mathcal{D}(G) - D_{\text{input}})\|_{\text{Fro}}^2$$

$$\text{s.t. } G \succeq 0$$

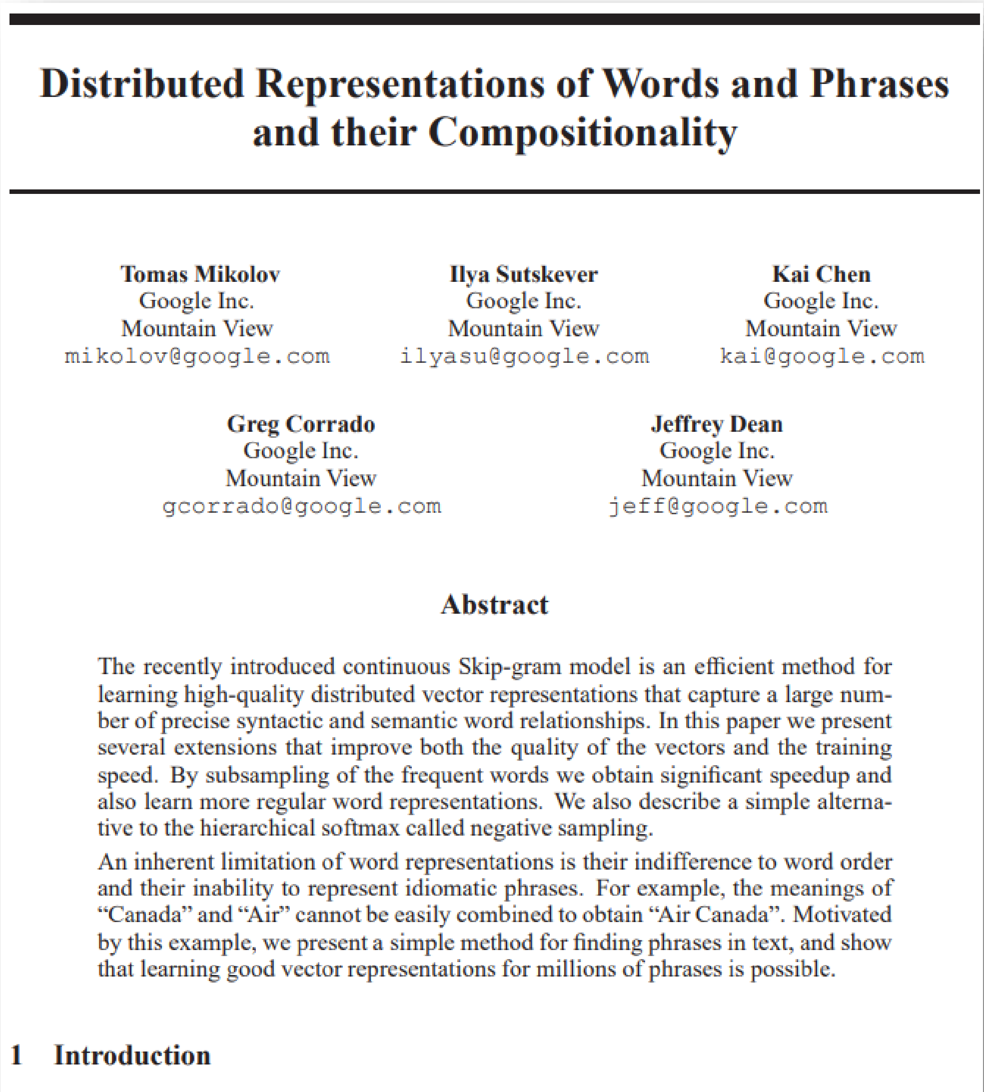
Convex program

Alfakih, Khandani, and Wolkowicz. "Solving Euclidean distance matrix completion problems via semidefinite programming." Comput. Optim. Appl., 12 (1999).

Maximum Variance Unfolding

(on the board)

Network Embedding



Distributed Representations of Words and Phrases and their Compositionality

Tomas Mikolov
Google Inc.
Mountain View
mikolov@google.com

Ilya Sutskever
Google Inc.
Mountain View
ilyasu@google.com

Kai Chen
Google Inc.
Mountain View
kai@google.com

Greg Corrado
Google Inc.
Mountain View
gcorrado@google.com

Jeffrey Dean
Google Inc.
Mountain View
jeff@google.com

Abstract

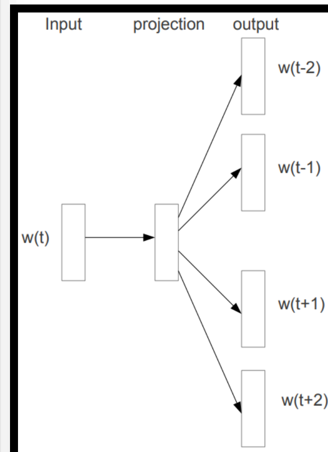
The recently introduced continuous Skip-gram model is an efficient method for learning high-quality distributed vector representations that capture a large number of precise syntactic and semantic word relationships. In this paper we present several extensions that improve both the quality of the vectors and the training speed. By subsampling of the frequent words we obtain significant speedup and also learn more regular word representations. We also describe a simple alternative to the hierarchical softmax called negative sampling.

An inherent limitation of word representations is their indifference to word order and their inability to represent idiomatic phrases. For example, the meanings of "Canada" and "Air" cannot be easily combined to obtain "Air Canada". Motivated by this example, we present a simple method for finding phrases in text, and show that learning good vector representations for millions of phrases is possible.

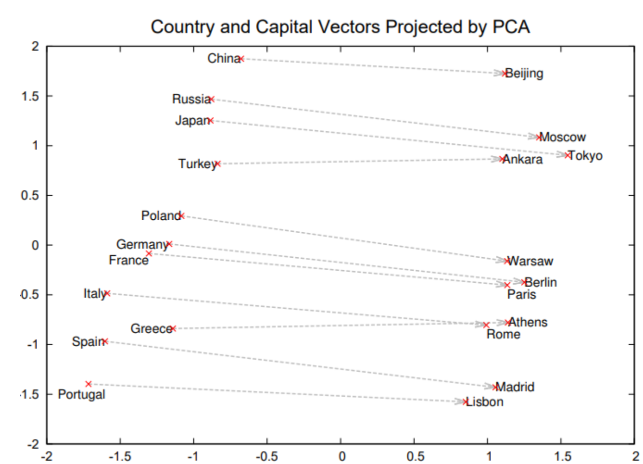
1 Introduction

Well-known example: Word2Vec

Download the embedding!



Skip-gram architecture: Predict neighborhood of a word



Challenging Computational Problems

- Is my data **embeddable**?
- Can you compute intrinsic **dimensionality**?
- Are two metric spaces **isometric**?
- How **similar** are two metric spaces?
- What is the **average** of two metric spaces?
- Can I embed into **non-Euclidean** spaces?

NP-Hardness Result

Robust Euclidean Embedding

Lawrence Cayton LCAYTON@CS.UCSD.EDU
Sanjoy Dasgupta DASGUPTA@CS.UCSD.EDU
Department of Computer Science and Engineering, University of California, San Diego
9500 Gilman Dr. La Jolla, CA 92093

Abstract

We derive a robust Euclidean embedding procedure based on semidefinite programming that may be used in place of the popular classical multidimensional scaling (cMDS) algorithm. We motivate this algorithm by arguing that cMDS is not particularly robust and has several other deficiencies. General-purpose semidefinite programming solvers are too memory intensive for medium to large sized applications, so we also describe a fast subgradient-based implementation of the robust algorithm. Additionally, since cMDS is often used for dimensionality reduction, we provide an in-depth look at reducing dimensionality with embedding procedures. In particular, we show that it is NP-hard to find optimal low-dimensional embeddings under a variety of cost functions.

choice for embedding seems to be c
sional scaling (cMDS). Its popula
ing relatively fast, parameter-free
and optimal for its cost function. In this work, we
look carefully at the algorithm and a
has some problematic features as wel
we argue that the cost function is n
conceptually awkward.

We propose a robust alternative to cM
clidean embedding (REE), that reta
desirable features of cMDS, but avo
pitfalls. We show that the global m
REE cost function can be found u
nite program (SDP). Though this is h
dard SDP-solvers can only manage the
gram for around 100 points. So th
used on more reasonably sized data
a subgradient-based implementation o

Dimensionality reduction is an important application

ℓ_1 EUCLIDEAN EMBEDDING
Input: A dissimilarity matrix $D = (d_{ij})$.
Output: An embedding into the line: $x_1, x_2, \ldots \in \mathbf{R}$
Goal: Minimize $\sum_{i,j} |d_{ij} - |x_i - x_j||$.

We show that this problem is NP-hard by reducing from a variant of not-all-equal 3SAT.

The hardness result can be extended to distortion functions of the form $\sum_{i,j} g\big(f(d_{ij}) - f(|x_i - x_j|)\big)$ We assume that f, g are

1. symmetric;
2. monotonically increasing in the absolute values of their arguments;
3. Lipschitz on $[0,1]$ with constant λ_U, that is, for $x, y \in [0,1]$, $|f(x) - f(y)| \le \lambda_U |x - y|$; and
4. similarly lower-bounded: for some $\lambda_L > 0$, for any $x, y \in [0,1]$, $|f(x) - f(y)| \ge \lambda_L |x - y| \max\{x, y\}$.

Notice that $f(x), g(x) \in \{x, x^2\}$ satisfy these conditions with $\lambda_U = 2, \lambda_L = 1$, meaning that $\|D - D^*\|_1$ and $\|D - D^*\|_2$ are both hard to minimize over one-dimensional embeddings.

Q:

What are some applications of this machinery?

Applications

- Reduce algorithmic runtime
- Compression
- Visualize data
- Interpolate
- Sample

...

http://thepioneerwoman.com/food-and-friends/chocolate-swiss-roll-cake/

Inverse Distance Problems

Justin Solomon
MIT, Spring 2019