On Low Distortion Embeddings of Statistical Distance Measures into Low Dimensional Spaces

Arnab Bhattacharyya
Manjish Pal
Purushottam Kar

Indian Institute of Technology, Kanpur

January 10, 2010

Introduction

Why "embed" distances anywhere at all

- Applications dealing with huge amounts of high dimensional data

Why "embed" distances anywhere at all

- Applications dealing with huge amounts of high dimensional data
- Prohibitive costs of performing point/range/k-NN queries in ambient space

Why “embed” distances anywhere at all

- Applications dealing with huge amounts of high dimensional data
- Prohibitive costs of performing point/range/k-NN queries in ambient space
  - Proximity queries become costlier with dimensionality - “Curse of Dimensionality”

Why "embed" distances anywhere at all

- Applications dealing with huge amounts of high dimensional data
- Prohibitive costs of performing point/range/k-NN queries in ambient space
    - Proximity queries become costlier with dimensionality - "Curse of Dimensionality"
    - Certain distance measures inherently difficult to compute (Earth Mover's distance)

Why “embed” distances anywhere at all

- Applications dealing with huge amounts of high dimensional data
- Prohibitive costs of performing point/range/k-NN queries in ambient space
  - Proximity queries become costlier with dimensionality - “Curse of Dimensionality”
  - Certain distance measures inherently difficult to compute (Earth Mover’s distance)
  - Absence of good index structures for non-metric distances

Existing Solutions

- Obtain easily estimable upper/lower-bounds on the distance measures (EMD/Edit distance)

Existing Solutions

- Obtain easily estimable upper/lower-bounds on the distance measures (EMD/Edit distance)
- Find embeddings which allow specific proximity queries

Existing Solutions

- Obtain easily estimable upper/lower-bounds on the distance measures (EMD/Edit distance)
- Find embeddings which allow specific proximity queries
- Embed into a metric space for which efficient algorithms for answering proximity queries exist

Statistical Distance Measures

- Prove to be challenging in context of embeddings

Statistical Distance Measures

- Prove to be challenging in context of embeddings
- Very useful in pattern recognition/database applications

Statistical Distance Measures

- Prove to be challenging in context of embeddings
- Very useful in pattern recognition/database applications
  - Bhattacharyya and Mahalanobis distances give better performance than ℓ_2 measure in image retrieval

Statistical Distance Measures

- Prove to be challenging in context of embeddings
- Very useful in pattern recognition/database applications
  - Bhattacharyya and Mahalanobis distances give better performance than ℓ_2 measure in image retrieval
  - Mahalanobis distance measure more useful than ℓ_2 when measuring distances between DNA sequences

Statistical Distance Measures

- Prove to be challenging in context of embeddings
- Very useful in pattern recognition/database applications
  - Bhattacharyya and Mahalanobis distances give better performance than ℓ_2 measure in image retrieval
  - Mahalanobis distance measure more useful than ℓ_2 when measuring distances between DNA sequences
  - Kullback-Leibler divergence well suited for use in time-critical texture retrieval from large databases

Statistical Distance Measures

- Prove to be challenging in context of embeddings
- Very useful in pattern recognition/database applications
  - Bhattacharyya and Mahalanobis distances give better performance than ℓ_2 measure in image retrieval
  - Mahalanobis distance measure more useful than ℓ_2 when measuring distances between DNA sequences
  - Kullback-Leibler divergence well suited for use in time-critical texture retrieval from large databases
  - Many other applications ...

Statistical Distance Measures

- Prove to be challenging in context of embeddings
- Very useful in pattern recognition/database applications
  - Bhattacharyya and Mahalanobis distances give better performance than ℓ_2 measure in image retrieval
  - Mahalanobis distance measure more useful than ℓ_2 when measuring distances between DNA sequences
  - Kullback-Leibler divergence well suited for use in time-critical texture retrieval from large databases
  - Many other applications ...
- However these are seldom metrics

Our contributions

- We examine 3 statistical distance measures with the goal of obtaining low-dimensional, low distortion embeddings

Our contributions

- We examine 3 statistical distance measures with the goal of obtaining low-dimensional, low distortion embeddings
- We present two techniques to prove non-embeddability results when concerned with embeddings of non-metrics into metric spaces

Our contributions

- We examine 3 statistical distance measures with the goal of obtaining low-dimensional, low distortion embeddings
- We present two techniques to prove non-embeddability results when concerned with embeddings of non-metrics into metric spaces
- Applying them we get non-embeddability results (into metric spaces) for the Bhattacharyya and Kullback Leibler measures

Our contributions

- We examine 3 statistical distance measures with the goal of obtaining low-dimensional, low distortion embeddings
- We present two techniques to prove non-embeddability results when concerned with embeddings of non-metrics into metric spaces
- Applying them we get non-embeddability results (into metric spaces) for the Bhattacharyya and Kullback Leibler measures
- We also present dimensionality reduction schemes for the Bhattacharyya and the Mahalanobis distance measure

Preliminaries

Low distortion embeddings

- Ensure that notions of distance are almost preserved

Low distortion embeddings

- Ensure that notions of distance are almost preserved
- Preserve the geometry of the original space almost exactly

Low distortion embeddings

- Ensure that notions of distance are almost preserved
- Preserve the geometry of the original space almost exactly
- Give performance guarantees in terms of accuracy for all proximity queries

Low distortion embeddings

- Ensure that notions of distance are almost preserved
- Preserve the geometry of the original space almost exactly
- Give performance guarantees in terms of accuracy for all proximity queries
- Presence of several index structures for metric spaces motivates embeddings into metric spaces

Low distortion embeddings

- Ensure that notions of distance are almost preserved
- Preserve the geometry of the original space almost exactly
- Give performance guarantees in terms of accuracy for all proximity queries
- Presence of several index structures for metric spaces motivates embeddings into metric spaces
- In case the embedding is into ℓ_2, added benefit of dimensionality reduction

Some Preliminary definitions

Definition (Metric Space)

A pair $M = (X, \rho)$ where X is a set and $\rho : X \times X \longrightarrow \mathbb{R}^+ \cup \{0\}$ is called a metric space provided the distance measure ρ satisfies the properties of identity, symmetry and triangular inequality.

Definition (D-embedding and Distortion)

Given two metric spaces (X, ρ) and (Y, σ), a mapping $f : X \longrightarrow Y$ is called a D-embedding where $D \geq 1$, if there exists a number $r > 0$ such that for all $x, y \in X$,

$$r \cdot \rho(x, y) \leq \sigma\left(f(x), f(y)\right) \leq D \cdot r \cdot \rho(x, y)$$

The infimum of all numbers D such that f is a D-embedding is called the *distortion* of f.

The JL Lemma

- A classic result in the field of metric embeddings

The JL Lemma

- A classic result in the field of metric embeddings
- Makes it possible for large point sets in high-dimensional Euclidean spaces to be embedded into low-dimensional Euclidean spaces with arbitrarily small distortion

The JL Lemma

- A classic result in the field of metric embeddings
- Makes it possible for large point sets in high-dimensional Euclidean spaces to be embedded into low-dimensional Euclidean spaces with arbitrarily small distortion
- This result was made practically applicable to databases by Achlioptas

The JL Lemma

- A classic result in the field of metric embeddings
- Makes it possible for large point sets in high-dimensional Euclidean spaces to be embedded into low-dimensional Euclidean spaces with arbitrarily small distortion
- This result was made practically applicable to databases by Achlioptas

Theorem (Johnson-Lindenstrauss Lemma)

Let X be an n-point set in a d-dimensional Euclidean space (i.e. $(X, \ell_2) \subset \left(\mathbb{R}^d, \ell_2\right)$), and let $\epsilon \in (0, 1]$ be given. Then there exists a $(1+\epsilon)$-embedding of X into $(\mathbb{R}^k, \ell_2)$ where $k = O\left(\epsilon^{-2} \log n\right)$. Furthermore, this embedding can be found out in randomized polynomial time.

The JL Lemma

- The lemma ensures that even inner products are preserved to an arbitrarily low additive error

The JL Lemma

- The lemma ensures that even inner products are preserved to an arbitrarily low additive error
- Will be useful for dimensionality reduction with the Bhattacharyya distance measure

The JL Lemma

- The lemma ensures that even inner products are preserved to an arbitrarily low additive error
- Will be useful for dimensionality reduction with the Bhattacharyya distance measure

Corollary

Let u, v be unit vectors in $\mathbb{R}^d$. Then, for any $\epsilon > 0$, a random projection of these vectors to yield the vectors u' and v' respectively satisfies $\Pr\left[u \cdot v - \epsilon \leq u' \cdot v' \leq u \cdot v + \epsilon\right] \geq 1 - 4e^{\frac{-k}{2}\left(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3}\right)}$.

Some definitions ...

Definition (Representative vector)

Given a d-dimensional histogram $P = (p_1, \dots p_d)$, let $\sqrt{P}$ denote the unit vector $(\sqrt{p_1}, \dots, \sqrt{p_d})$. We shall call this the *representative vector* of P.

Some definitions ...

Definition (Representative vector)

Given a d-dimensional histogram $P = (p_1, \ldots p_d)$, let $\sqrt{P}$ denote the unit vector $(\sqrt{p_1}, \ldots, \sqrt{p_d})$. We shall call this the *representative vector* of P.

Definition (α-constrained histogram)

A histogram $P = (p_1, p_2, \ldots p_d)$ is said to be α-constrained if $p_i \geq \frac{\alpha}{d}$ for $i = 1, 2, \ldots, d$.

Some definitions ...

Definition (Representative vector)

Given a d-dimensional histogram $P = (p_1, \ldots p_d)$, let $\sqrt{P}$ denote the unit vector $(\sqrt{p_1}, \ldots, \sqrt{p_d})$. We shall call this the *representative vector* of P.

Definition (α-constrained histogram)

A histogram $P = (p_1, p_2, \ldots p_d)$ is said to be α-constrained if $p_i \geq \frac{\alpha}{d}$ for $i = 1, 2, \ldots, d$.

Observation

Given two α-constrained histograms P and Q, the inner product between the representative vectors is at least α, i.e., $\left\langle \sqrt{P}, \sqrt{Q} \right\rangle \geq \alpha$.

Some definitions ...

Definition (Representative vector)

Given a d-dimensional histogram $P = (p_1, \dots p_d)$, let $\sqrt{P}$ denote the unit vector $(\sqrt{p_1}, \dots, \sqrt{p_d})$. We shall call this the *representative vector* of P.

Definition (α-constrained histogram)

A histogram $P = (p_1, p_2, \dots p_d)$ is said to be α-constrained if $p_i \geq \frac{\alpha}{d}$ for $i = 1, 2, \dots, d$.

Observation

Given two α-constrained histograms P and Q, the inner product between the representative vectors is at least α, i.e., $\left\langle \sqrt{P}, \sqrt{Q} \right\rangle \geq \alpha$.

We will denote $\frac{\alpha}{d}$ by β

The Distance Measures

The Bhattacharyya Distance Measures

Definition (Bhattacharyya Coefficient)

For two histograms $P = (p_1, p_2, \ldots, p_d)$ and $Q = (q_1, q_2, \ldots q_d)$ with $\sum_{i=1}^{d} p_i = \sum_{i=1}^{n} q_i = 1$ and each $p_i, q_i \geq 0$, the *Bhattacharyya coefficient* is described as $BC(P, Q) = \sum_{i=1}^{n} \sqrt{p_i q_i}$.

We define two distance measures using this coefficient :

The Bhattacharyya Distance Measures

Definition (Bhattacharyya Coefficient)

For two histograms $P = (p_1, p_2, \dots, p_d)$ and $Q = (q_1, q_2, \dots q_d)$ with $\sum_{i=1}^{d} p_i = \sum_{i=1}^{n} q_i = 1$ and each $p_i, q_i \geq 0$, the *Bhattacharyya coefficient* is described as $BC(P, Q) = \sum_{i=1}^{n} \sqrt{p_i q_i}$.

We define two distance measures using this coefficient :

Definition (Hellinger Distance)

$H(P, Q) = 1 - BC(P, Q) = \frac{1}{2} \left(\left\| \sqrt{P} - \sqrt{Q} \right\| \right)^2$.

The Bhattacharyya Distance Measures

Definition (Bhattacharyya Coefficient)

For two histograms $P = (p_1, p_2, \ldots, p_d)$ and $Q = (q_1, q_2, \ldots q_d)$ with $\sum_{i=1}^{d} p_i = \sum_{i=1}^{n} q_i = 1$ and each $p_i, q_i \geq 0$, the *Bhattacharyya coefficient* is described as $BC(P, Q) = \sum_{i=1}^{n} \sqrt{p_i q_i}$.

We define two distance measures using this coefficient :

Definition (Hellinger Distance)

$H(P, Q) = 1 - BC(P, Q) = \frac{1}{2}\left(\left\|\sqrt{P} - \sqrt{Q}\right\|\right)^2$.

Definition (Bhattacharyya Distance)

$BD(P, Q) = -\ln BC(P, Q)$.

Kullback Leibler Divergence

Definition (Kullback Leibler Divergence)

Given two histograms $P = \{p_1, p_2, \ldots, p_d\}$ and $Q = \{q_2, q_2 \ldots q_d\}$, the Kullback-Leibler divergence between the two distributions is defined as $KL(P, Q) = \sum_{i=1}^{d} p_i \ln \frac{p_i}{q_i}$.

Kullback Leibler Divergence

Definition (Kullback Leibler Divergence)

Given two histograms $P = \{p_1, p_2, \ldots, p_d\}$ and $Q = \{q_2, q_2 \ldots q_d\}$, the Kullback-Leibler divergence between the two distributions is defined as $KL(P, Q) = \sum_{i=1}^{d} p_i \ln \frac{p_i}{q_i}$.

- Non-symmetric and unbounded, i.e., for any given $c > 0$, one can construct histograms whose Kullback-Leibler divergence exceeds c.

Kullback Leibler Divergence

Definition (Kullback Leibler Divergence)

Given two histograms $P = \{p_1, p_2, \ldots, p_d\}$ and $Q = \{q_2, q_2 \ldots q_d\}$, the Kullback-Leibler divergence between the two distributions is defined as $KL(P, Q) = \sum_{i=1}^{d} p_i \ln \frac{p_i}{q_i}$.

- Non-symmetric and unbounded, i.e., for any given $c > 0$, one can construct histograms whose Kullback-Leibler divergence exceeds c.
- In order to avoid these singularities, we assume that the histograms are β-constrained

Kullback Leibler Divergence

Definition (Kullback Leibler Divergence)

Given two histograms $P = \{p_1, p_2, \ldots, p_d\}$ and $Q = \{q_2, q_2 \ldots q_d\}$, the Kullback-Leibler divergence between the two distributions is defined as $KL(P, Q) = \sum_{i=1}^{d} p_i \ln \frac{p_i}{q_i}$.

- Non-symmetric and unbounded, i.e., for any given $c > 0$, one can construct histograms whose Kullback-Leibler divergence exceeds c.
- In order to avoid these singularities, we assume that the histograms are β-constrained

Lemma

Given two β-constrained histograms P, Q, $0 \leq KL(P, Q) \leq \ln \frac{1}{\beta}$.

The Class of Quadratic Form Distance Measures

Definition (Quadratic Form Distance Measure)

A $d \times d$ positive definite matrix A defines a Quadratic Form Distance measure over $\mathbb{R}^d$ given by $Q_A(x, y) = \sqrt{(x - y)^T A(x - y)}$.

The Class of Quadratic Form Distance Measures

Definition (Quadratic Form Distance Measure)

A $d \times d$ positive definite matrix A defines a Quadratic Form Distance measure over $\mathbb{R}^d$ given by $Q_A(x, y) = \sqrt{(x - y)^T A(x - y)}$.

- Can be defined for any matrix but the resulting distance measure is a metric if and only if the matrix is positive definite

The Class of Quadratic Form Distance Measures

Definition (Quadratic Form Distance Measure)

A $d \times d$ positive definite matrix A defines a Quadratic Form Distance measure over $\mathbb{R}^d$ given by $Q_A(x, y) = \sqrt{(x-y)^T A(x-y)}$.

- Can be defined for any matrix but the resulting distance measure is a metric if and only if the matrix is positive definite
- The Mahalanobis distance is a special case of QFD where the underlying distance measure is the covariance matrix of some distribution

Results on Dimensionality Reduction

Hellinger Distance

The fact that $H(P, Q)$ is the Euclidean distance between the points $\sqrt{P}$ and $\sqrt{Q}$ allows us to state the following theorem.

Theorem

The Hellinger distance admits a low distortion dimensionality reduction.

Proof. (Sketch).

Given a set of histograms, subject the corresponding set of representative vectors to a JL-type embedding and output a set of vectors for which the embedded set of vectors are the representatives. □

Bhattacharyya Distance

- The Bhattacharyya distance is unbounded even on the probability simplex - precisely when α is small

Bhattacharyya Distance

- The Bhattacharyya distance is unbounded even on the probability simplex - precisely when α is small
- Our result works well if distributions are α-constrained for large α

Bhattacharyya Distance

- The Bhattacharyya distance is unbounded even on the probability simplex - precisely when α is small
- Our result works well if distributions are α-constrained for large α

Theorem

The Bhattacharyya distance admits a low additive distortion dimensionality reduction.

Bhattacharyya Distance

- The Bhattacharyya distance is unbounded even on the probability simplex - precisely when α is small
- Our result works well if distributions are α-constrained for large α

Theorem

The Bhattacharyya distance admits a low additive distortion dimensionality reduction.

Proof. (Sketch).

Given a set of α-constrained histograms, subject them to a JL-type embedding with the error parameter set to $\epsilon' = \frac{\epsilon \cdot \alpha}{2}$. With high probability the following occurs : if P, Q are embedded respectively to P', Q', then $BD(P,Q) - \epsilon \leq BD(P', Q') \leq BD(P,Q) + \epsilon$. □

Quadratic Form Distance Measures

Theorem

The family of metric quadratic form distance measures admit a low distortion JL-type embedding into a Euclidean spaces.

Proof.

Every quadratic form distance measure forming a metric is characterized by a positive definite matrix A. Such matrices can be subjected to a Cholesky Decomposition of the form $A = L^T L$. Given a set of vectors subject them to the transformation $x \longmapsto Lx$ and subject there resulting vectors to a JL-type embedding. □

The proposed transformation essentially reduces the problem to an undistorted Euclidean space where the JL Lemma can be applied.

How to prove Non-embeddability results into metric spaces

The Asymmetry Technique

Definition (γ-Relaxed Symmetry)

A set X equipped with a distance function $d : X \times X \longrightarrow \mathbb{R}^+ \cup \{0\}$, is said to satisfy *$\gamma$-relaxed symmetry* if there exists $\gamma \geq 0$ such that for all point pairs $p, q \in X$, the following holds $|d(p, q) - d(q, p)| \leq \gamma$.

The Asymmetry Technique

Definition (γ-Relaxed Symmetry)

A set X equipped with a distance function $d : X \times X \longrightarrow \mathbb{R}^+ \cup \{0\}$, is said to satisfy *$\gamma$-relaxed symmetry* if there exists $\gamma \geq 0$ such that for all point pairs $p, q \in X$, the following holds $|d(p,q) - d(q,p)| \leq \gamma$.

Metrics satisfy the γ-relaxed triangle inequality for $\gamma = 0$

The Asymmetry Technique

Definition (γ-Relaxed Symmetry)

A set X equipped with a distance function $d : X \times X \longrightarrow \mathbb{R}^+ \cup \{0\}$, is said to satisfy *$\gamma$-relaxed symmetry* if there exists $\gamma \geq 0$ such that for all point pairs $p, q \in X$, the following holds $|d(p,q) - d(q,p)| \leq \gamma$.

Metrics satisfy the γ-relaxed triangle inequality for $\gamma = 0$

Lemma

Given a set X equipped with a distance function d that does not satisfy the γ-relaxed symmetry such that $d(x,y) \leq M$ for all $x, y \in X$, any embedding of X into a metric space incurs a distortion of at least $1 + \frac{\gamma}{M}$.

The Relaxed Triangle Inequality Technique

Definition (λ-Relaxed Triangle Inequality)

A set X equipped with a distance function $d : X \times X \longrightarrow \mathbb{R}^+ \cup \{0\}$, is said to satisfy the λ-*relaxed triangle inequality* if there exists some constant $\lambda \leq 1$ such that for all triplets $p, q, r \in X$, the following holds $d(p, r) + d(r, q) \geq \lambda \cdot d(p, q)$.

The Relaxed Triangle Inequality Technique

Definition (λ-Relaxed Triangle Inequality)

A set X equipped with a distance function $d : X \times X \longrightarrow \mathbb{R}^+ \cup \{0\}$, is said to satisfy the λ-*relaxed triangle inequality* if there exists some constant $\lambda \leq 1$ such that for all triplets $p, q, r \in X$, the following holds $d(p, r) + d(r, q) \geq \lambda \cdot d(p, q)$.

Metrics satisfy the λ-relaxed triangle inequality for $\lambda = 1$

The Relaxed Triangle Inequality Technique

Definition (λ-Relaxed Triangle Inequality)

A set X equipped with a distance function $d : X \times X \longrightarrow \mathbb{R}^+ \cup \{0\}$, is said to satisfy the λ-*relaxed triangle inequality* if there exists some constant $\lambda \leq 1$ such that for all triplets $p, q, r \in X$, the following holds $d(p, r) + d(r, q) \geq \lambda \cdot d(p, q)$.

Metrics satisfy the λ-relaxed triangle inequality for $\lambda = 1$

Lemma

Any embedding of a set X equipped with a distance function d that does not satisfy the λ-relaxed triangle inequality into a metric space incurs a distortion of at least $\frac{1}{\lambda}$.

Non-"metric-embeddability" Results

Metric Embeddings for Bhattacharyya Distance

Theorem

There exist d-dimensional β-constrained distributions such that any embedding of these distributions under the Bhattacharyya distance measure into a metric space must incur a distortion of

$$D = \begin{cases} \Omega\left(\frac{\ln \frac{1}{d\beta}}{\ln d}\right) & \text{when } \beta > \frac{4}{d^2} \\ \Omega\left(\frac{\ln \frac{1}{\beta}}{\ln d}\right) & \text{when } \beta \leq \frac{4}{d^2} \end{cases}$$

Metric Embeddings for Bhattacharyya Distance

Theorem

There exist d-dimensional β-constrained distributions such that any embedding of these distributions under the Bhattacharyya distance measure into a metric space must incur a distortion of

$$D = \begin{cases} \Omega\left(\frac{\ln \frac{1}{d\beta}}{\ln d}\right) & \text{when } \beta > \frac{4}{d^2} \\ \Omega\left(\frac{\ln \frac{1}{\beta}}{\ln d}\right) & \text{when } \beta \leq \frac{4}{d^2} \end{cases}$$

Proof. (Sketch).

Choose three distributions that violate the relaxed triangle inequality with appropriate λ. □

Metric Embeddings for Bhattacharyya Distance

Theorem

For any two d-dimensional β-constrained distributions P and Q with $\beta < \frac{1}{2d}$, we have $H(P,Q) \leq BD(P,Q) \leq \frac{d}{1-2\beta d} \ln \frac{1}{(d-1)\beta} H(P,Q)$.

- Since the Hellinger distance forms a metric in the positive orthant, this constitutes a metric embedding

Metric Embeddings for Bhattacharyya Distance

Theorem

For any two d-dimensional β-constrained distributions P and Q with $\beta < \frac{1}{2d}$, we have $H(P, Q) \leq BD(P, Q) \leq \frac{d}{1-2\beta d} \ln \frac{1}{(d-1)\beta} H(P, Q)$.

- Since the Hellinger distance forms a metric in the positive orthant, this constitutes a metric embedding
- The result can be interpreted to show that the non-embeddability theorem stated earlier is tight upto an $O(d \ln d)$ factor

Metric Embeddings for Bhattacharyya Distance

Theorem

For any two d-dimensional β-constrained distributions P and Q with $\beta < \frac{1}{2d}$, we have $H(P,Q) \leq BD(P,Q) \leq \frac{d}{1-2\beta d} \ln \frac{1}{(d-1)\beta} H(P,Q)$.

- Since the Hellinger distance forms a metric in the positive orthant, this constitutes a metric embedding
- The result can be interpreted to show that the non-embeddability theorem stated earlier is tight upto an $O(d \ln d)$ factor
- Additionally this embedding allows for dimensionality reduction as well

Metric Embeddings for Kullback-Leibler Divergence

An application of the Asymmetry technique gives us the following result

Metric Embeddings for Kullback-Leibler Divergence

An application of the Asymmetry technique gives us the following result

Theorem

For sufficiently large d and small β, there exists a set S of d-dimensional β-constrained histograms and a constant $c > 0$ such that any embedding of S into a metric space incurs a distortion of at least $1 + c$.

Metric Embeddings for Kullback-Leibler Divergence

An application of the Asymmetry technique gives us the following result

Theorem

For sufficiently large d and small β, there exists a set S of d-dimensional β-constrained histograms and a constant $c > 0$ such that any embedding of S into a metric space incurs a distortion of at least $1 + c$.

- It can be shown that this proof technique cannot give more than a constant lower bound in this case

Metric Embeddings for Kullback-Leibler Divergence

An application of the Asymmetry technique gives us the following result

Theorem

For sufficiently large d and small β, there exists a set S of d-dimensional β-constrained histograms and a constant $c > 0$ such that any embedding of S into a metric space incurs a distortion of at least $1 + c$.

- It can be shown that this proof technique cannot give more than a constant lower bound in this case
- However the situation is much worse ...

Metric Embeddings for Kullback-Leibler Divergence

An application of the Relaxed Triangle Inequality Technique gives us the following result

Metric Embeddings for Kullback-Leibler Divergence

An application of the Relaxed Triangle Inequality Technique gives us the following result

Theorem

For sufficiently large d, there exist d-dimensional β-constrained distributions such that embedding these under the Kullback-Leibler divergence into a metric space must incur a distortion of

$$\Omega\left(\frac{\ln\frac{1}{d\beta}}{\ln\left(d\ln\frac{1}{\beta}\right)}\right).$$

Metric Embeddings for Kullback-Leibler Divergence

An application of the Relaxed Triangle Inequality Technique gives us the following result

Theorem

For sufficiently large d, there exist d-dimensional β-constrained distributions such that embedding these under the Kullback-Leibler divergence into a metric space must incur a distortion of

$$\Omega\left(\frac{\ln \frac{1}{d\beta}}{\ln\left(d \ln \frac{1}{\beta}\right)}\right).$$

- The lower bound diverges for small β

Metric Embeddings for Kullback-Leibler Divergence

An application of the Relaxed Triangle Inequality Technique gives us the following result

Theorem

For sufficiently large d, there exist d-dimensional β-constrained distributions such that embedding these under the Kullback-Leibler divergence into a metric space must incur a distortion of

$$\Omega\left(\frac{\ln \frac{1}{d\beta}}{\ln\left(d \ln \frac{1}{\beta}\right)}\right).$$

- The lower bound diverges for small β
- Thus, by choosing point sets appropriately, we can force the embedding distortion to be arbitrarily large !

Metric Embeddings for Kullback-Leibler Divergence

- The above bounds show how the Kullback-Leibler divergence behaves near the uniform distribution and near the boundaries of the probability simplex

Metric Embeddings for Kullback-Leibler Divergence

- The above bounds show how the Kullback-Leibler divergence behaves near the uniform distribution and near the boundaries of the probability simplex
- Near the uniform distribution, asymmetry makes the Kullback-Leibler divergence hard to approximate by a metric

Metric Embeddings for Kullback-Leibler Divergence

- The above bounds show how the Kullback-Leibler divergence behaves near the uniform distribution and near the boundaries of the probability simplex
- Near the uniform distribution, asymmetry makes the Kullback-Leibler divergence hard to approximate by a metric
- As we move away from the uniform distribution the hardness is due to violation of the triangle inequality

Metric Embeddings for Kullback-Leibler Divergence

- The above bounds show how the Kullback-Leibler divergence behaves near the uniform distribution and near the boundaries of the probability simplex
- Near the uniform distribution, asymmetry makes the Kullback-Leibler divergence hard to approximate by a metric
- As we move away from the uniform distribution the hardness is due to violation of the triangle inequality
- For large β (say $\beta = \Omega\left(\frac{1}{d}\right)$), the Asymmetry Technique gives a better bound

Metric Embeddings for Kullback-Leibler Divergence

- The above bounds show how the Kullback-Leibler divergence behaves near the uniform distribution and near the boundaries of the probability simplex
- Near the uniform distribution, asymmetry makes the Kullback-Leibler divergence hard to approximate by a metric
- As we move away from the uniform distribution the hardness is due to violation of the triangle inequality
- For large β (say $\beta = \Omega\left(\frac{1}{d}\right)$), the Asymmetry Technique gives a better bound
- For smaller β (say $\beta = o\left(\frac{1}{d^4}\right)$) we get a better lower bound using the Relaxed Triangle Inequality Technique - this lower bound diverges

A Useful Embedding for Kullback-Leibler Divergence

Theorem

For any two d-dimensional β-constrained distributions P and Q,

$$\frac{\ell_2^2(P,Q)}{2} \le KL(P,Q) \le \left(\frac{1}{2\beta} + \frac{1}{3\beta^5}\right)\ell_2^2(P,Q).$$

A Useful Embedding for Kullback-Leibler Divergence

Theorem

For any two d-dimensional β-constrained distributions P and Q,

$$\frac{\ell_2^2(P,Q)}{2} \leq KL(P,Q) \leq \left(\frac{1}{2\beta} + \frac{1}{3\beta^5}\right)\ell_2^2(P,Q).$$

Uses a result from information theory called Pinsker's inequality

A Useful Embedding for Kullback-Leibler Divergence

Theorem

For any two d-dimensional β-constrained distributions P and Q,

$$\frac{\ell_2^2(P,Q)}{2} \leq KL(P,Q) \leq \left(\frac{1}{2\beta} + \frac{1}{3\beta^5}\right)\ell_2^2(P,Q).$$

Uses a result from information theory called Pinsker's inequality

- The ℓ_2^2 measure is not a metric - however very close to one

A Useful Embedding for Kullback-Leibler Divergence

Theorem

For any two d-dimensional β-constrained distributions P and Q,

$$\frac{\ell_2^2(P,Q)}{2} \leq KL(P,Q) \leq \left(\frac{1}{2\beta} + \frac{1}{3\beta^5}\right) \ell_2^2(P,Q).$$

Uses a result from information theory called Pinsker's inequality

- The ℓ_2^2 measure is not a metric - however very close to one
- It also admits dimensionality reduction via the JL Lemma

A Useful Embedding for Kullback-Leibler Divergence

Theorem

For any two d-dimensional β-constrained distributions P and Q,

$$\frac{\ell_2^2(P,Q)}{2} \leq KL(P,Q) \leq \left(\frac{1}{2\beta} + \frac{1}{3\beta^5}\right)\ell_2^2(P,Q).$$

Uses a result from information theory called Pinsker's inequality

- The ℓ_2^2 measure is not a metric - however very close to one
- It also admits dimensionality reduction via the JL Lemma
- Hence despite the poor bound on distortion, can be useful

Future Directions

Open Questions

- A low multiplicative distortion dimensionality reduction scheme for the Bhattacharyya distance measure

Open Questions

- A low multiplicative distortion dimensionality reduction scheme for the Bhattacharyya distance measure
- A low distortion dimensionality reduction scheme for the Kullback-Leibler distance measure

Open Questions

- A low multiplicative distortion dimensionality reduction scheme for the Bhattacharyya distance measure
- A low distortion dimensionality reduction scheme for the Kullback-Leibler distance measure
- Tightening of the bounds for the Bhattacharyya distance measure shown in this paper

Open Questions

- A low multiplicative distortion dimensionality reduction scheme for the Bhattacharyya distance measure
- A low distortion dimensionality reduction scheme for the Kullback-Leibler distance measure
- Tightening of the bounds for the Bhattacharyya distance measure shown in this paper
- In short - a theory of Non-Metric Embeddings