SIGTACS Seminar Series

Metric Embeddings and Applications in Computer Science

Presented by : Purushottam Kar

January 10, 2009

Outline

Basics

Definition (Metric)

A Metric is a structure (X, ρ) where ρ is a distance measure $\rho : X \times X \rightarrow \mathbb{R}$ which is non-negative, symmetric and satisfies the triangle inequality.

Definition (Embedding Distortion)

An embedding $f : X \rightarrow Y$ from a metric space (X, ρ) to another metric space (Y, σ) is said to have a distortion D if $D = \sup_{x,y \in X} \frac{\sigma(f(x), f(y))}{\rho(x,y)} \cdot \sup_{x,y \in X} \frac{\rho(x,y)}{\sigma(f(x), f(y))}$.

Such embeddings are also called *bi-Lipschitz* embeddings.

Embeddings

- Various criterion used to evaluate embeddings

Embeddings

- Various criterion used to evaluate embeddings
- Distortion, Stress, Residual Variance ...

Definition (Embedding Stress)

The stress for an embedding $f : X \rightarrow Y$ from a metric space (X, ρ) to another metric space (Y, σ) is defined to be $\sqrt{\frac{\sum_{x,y \in X} (\sigma(f(x), f(y)) - \rho(x,y))^2}{\sum_{x,y \in X} \rho(x,y)^2}}$.

Embeddings

- Various criterion used to evaluate embeddings
- Distortion, Stress, Residual Variance ...

Definition (Embedding Stress)

The stress for an embedding $f : X \rightarrow Y$ from a metric space (X, ρ) to another metric space (Y, σ) is defined to be $\sqrt{\frac{\sum_{x,y \in X} (\sigma(f(x),f(y)) - \rho(x,y))^2}{\sum_{x,y \in X} \rho(x,y)^2}}$.

- Lead to very interesting algorithmic questions

Application in Computer Science

- Started out as a branch of functional analysis

Application in Computer Science

- Started out as a branch of functional analysis
- Algorithmic applications

Application in Computer Science

- Started out as a branch of functional analysis
- Algorithmic applications
    - Metric Embeddings for datasets operating with a non-metric

Application in Computer Science

- Started out as a branch of functional analysis
- Algorithmic applications
  - Metric Embeddings for datasets operating with a non-metric
  - Dimensionality reduction to reduce storage space costs, processing time

Application in Computer Science

- Started out as a branch of functional analysis
- Algorithmic applications
  - Metric Embeddings for datasets operating with a non-metric
  - Dimensionality reduction to reduce storage space costs, processing time
  - Facilitate pruning procedures in database searches

Application in Computer Science

- Started out as a branch of functional analysis
- Algorithmic applications
  - Metric Embeddings for datasets operating with a non-metric
  - Dimensionality reduction to reduce storage space costs, processing time
  - Facilitate pruning procedures in database searches
  - Preserve residual variance (PCA), inter-point similarity (Random Projections), Stress (MDS)

Application in Computer Science

- Started out as a branch of functional analysis
- Algorithmic applications
  - Metric Embeddings for datasets operating with a non-metric
  - Dimensionality reduction to reduce storage space costs, processing time
  - Facilitate pruning procedures in database searches
  - Preserve residual variance (PCA), inter-point similarity (Random Projections), Stress (MDS)
- Streaming Algorithms

Embedding into l_∞

Theorem (Frétchet's Embedding)

Every n-point metric can be isometrically embedded into l_∞

- Fréchet's Embedding technique - non-expansive

Embedding into l_∞

Theorem (Frétchet's Embedding)

Every n-point metric can be isometrically embedded into l_∞

- Fréchet's Embedding technique - non-expansive
- Choose coordinates as projections onto some fixed sets

Embedding into l_∞

Theorem (Frétchet's Embedding)

Every n-point metric can be isometrically embedded into l_∞

- Fréchet's Embedding technique - non-expansive
- Choose coordinates as projections onto some fixed sets
- Triangle inequality ensures contractive embeddings

Embedding into l_∞

Theorem (Frétchet's Embedding)

Every n-point metric can be isometrically embedded into l_∞

- Fréchet's Embedding technique - non-expansive
- Choose coordinates as projections onto some fixed sets
- Triangle inequality ensures contractive embeddings
- Choice of "landmark" sets gives other algorithms

Embedding into l_∞

Theorem (Frétchet's Embedding)

Every n-point metric can be isometrically embedded into l_∞

- Fréchet's Embedding technique - non-expansive
- Choose coordinates as projections onto some fixed sets
- Triangle inequality ensures contractive embeddings
- Choice of "landmark" sets gives other algorithms
- Embedding dimension can be reduced to $O(qn^{\frac{1}{q}} \ln n)$ by tolerating a distortion of $2q - 1$.

Embedding into l_2

Theorem (Bourgain's Embedding)

Every n-point metric can be $O(\log n)$-embedded into l_2

- Uses a random selection of the landmark sets

Embedding into l_2

Theorem (Bourgain's Embedding)

Every n-point metric can be $O(\log n)$-embedded into l_2

- Uses a random selection of the landmark sets
- Tight - The graph metric of a constant degree expander has $\Omega(\log n)$ distortion into any Euclidean space

Embedding into l_2

Theorem (Bourgain's Embedding)

Every n-point metric can be $O(\log n)$-embedded into l_2

- Uses a random selection of the landmark sets
- Tight - The graph metric of a constant degree expander has $\Omega(\log n)$ distortion into any Euclidean space
- Any embedding of the Hamming cube into l_2 incurs $\Omega\left(\sqrt{\log n}\right)$ distortion

Dimensionality Reduction in l_1

- Impossible - A D-embedding of n points may require $n^{\Omega(1/D^2)}$ dimensions

Dimensionality Reduction in l_1

- Impossible - A D-embedding of n points may require $n^{\Omega(1/D^2)}$ dimensions
- No “flattening” results known for other l_p metrics either ...

Dimensionality Reduction in l_1

- Impossible - A D-embedding of n points may require $n^{\Omega(1/D^2)}$ dimensions
- No "flattening" results known for other l_p metrics either ...
- Except for $p = 2$

The Johnson-Lindenstrauss Lemma

Theorem (The JL-Lemma)

Given $\epsilon > 0$ and integer n, let $k \geq k_0 = \mathcal{O}(\epsilon^{-2} \log n)$. For every set P of n points in $\mathbb{R}^d$ there exists $f : \mathbb{R}^d \longrightarrow \mathbb{R}^k$ such that for all $u, v \in P$

$$(1 - \epsilon)\|u - v\|^2 \leq \|f(u) - f(v)\|^2 \leq (1 + \epsilon)\|u - v\|^2.$$

- Implementation as a randomized algorithm

The Johnson-Lindenstrauss Lemma

Theorem (The JL-Lemma)

Given $\epsilon > 0$ *and integer* n*, let* $k \geq k_0 = \mathcal{O}(\epsilon^{-2} \log n)$*. For every set* P *of* n *points in* $\mathbb{R}^d$ *there exists* $f : \mathbb{R}^d \longrightarrow \mathbb{R}^k$ *such that for all* $u, v \in P$

$$(1-\epsilon)\|u-v\|^2 \leq \|f(u)-f(v)\|^2 \leq (1+\epsilon)\|u-v\|^2.$$

- Implementation as a randomized algorithm
- Equivalent interpretations - random projection vs. random rotation

The Johnson-Lindenstrauss Lemma

Theorem (The JL-Lemma)

Given $\epsilon > 0$ and integer n, let $k \geq k_0 = \mathcal{O}(\epsilon^{-2} \log n)$. For every set P of n points in $\mathbb{R}^d$ there exists $f : \mathbb{R}^d \longrightarrow \mathbb{R}^k$ such that for all $u, v \in P$

$$(1-\epsilon)\|u-v\|^2 \leq \|f(u)-f(v)\|^2 \leq (1+\epsilon)\|u-v\|^2.$$

- Implementation as a randomized algorithm
- Equivalent interpretations - random projection vs. random rotation
- Various Proofs known [IM98], [DG99], [AV99], [A01]

The Johnson-Lindenstrauss Lemma

Theorem (The JL-Lemma)

Given $\epsilon > 0$ and integer n, let $k \geq k_0 = \mathcal{O}(\epsilon^{-2} \log n)$. For every set P of n points in $\mathbb{R}^d$ there exists $f : \mathbb{R}^d \longrightarrow \mathbb{R}^k$ such that for all $u, v \in P$

$$(1-\epsilon)\|u-v\|^2 \leq \|f(u)-f(v)\|^2 \leq (1+\epsilon)\|u-v\|^2.$$

- Implementation as a randomized algorithm
- Equivalent interpretations - random projection vs. random rotation
- Various Proofs known [IM98], [DG99], [AV99], [A01]
- Common Technique

$$\text{Point Drafting} \longrightarrow \text{Set Drafting} \overset{\text{Union Bound}}{\longrightarrow} \text{Set Embedding}$$

Enter Achlioptas

- Instead of choosing from an uncountably infinite domain, can we choose vectors from a finite set of vectors ?

Enter Achlioptas

- Instead of choosing from an uncountably infinite domain, can we choose vectors from a finite set of vectors ?
- Achlioptas: In fact 'choosing' from the d-dimensional Hamming Cube $\{1, -1\}^d$ works.

Enter Achlioptas

- Instead of choosing from an uncountably infinite domain, can we choose vectors from a finite set of vectors ?
- Achlioptas: In fact 'choosing' from the d-dimensional Hamming Cube $\{1, -1\}^d$ works.
- Consider a random vector $R = (X_1, X_2, \ldots, X_d)$, where each X_i is chosen from one of the two distributions:

$$D_1 = \frac{1}{\sqrt{d}} \begin{cases} -1 & \text{with probability } 1/2 \\ 1 & \text{with probability } 1/2 \end{cases}$$

$$D_2 = \frac{1}{\sqrt{d}} \begin{cases} -\sqrt{3} & \text{with probability } 1/6 \\ 0 & \text{with probability } 2/3 \\ \sqrt{3} & \text{with probability } 1/6 \end{cases}$$

Enter Achlioptas

- Pick k such random vectors $R_1, R_2, \ldots R_k$.

Enter Achlioptas

- Pick k such random vectors $R_1, R_2, \ldots R_k$.
- For a given unit vector $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_d)$, the low ($k$-)dimensional vector corresponding to α is

$$f(\alpha) = \sqrt{\frac{d}{k}}(\langle \alpha, R_1 \rangle, \langle \alpha, R_2 \rangle, \ldots, \langle \alpha, R_k \rangle)$$

Enter Achlioptas

- Pick k such random vectors $R_1, R_2, \ldots R_k$.
- For a given unit vector $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_d)$, the low ($k$-)dimensional vector corresponding to α is

$$f(\alpha) = \sqrt{\tfrac{d}{k}}(\langle \alpha, R_1 \rangle, \langle \alpha, R_2 \rangle, \ldots, \langle \alpha, R_k \rangle)$$

- Advantage: Simple and can be implemented as SQL queries.

Main Theorem

- Let $S = \langle \alpha, R_1 \rangle^2 + \langle \alpha, R_2 \rangle^2 + \cdots \langle \alpha, R_k \rangle^2$

Theorem (Main Theorem)

For every d-dimensional unit vector α*, integer* $k \geq 1$ *and* $\epsilon > 0$

$$\Pr\left[S \geq (1 \pm \epsilon)\frac{k}{d} \cdot 1\right] \leq e^{\frac{-k}{2}(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3})}$$

Main Theorem

- Let $S = \langle \alpha, R_1 \rangle^2 + \langle \alpha, R_2 \rangle^2 + \cdots \langle \alpha, R_k \rangle^2$

Theorem (Main Theorem)

For every d-dimensional unit vector α, integer $k \geq 1$ and $\epsilon > 0$

$$\Pr\left[S \geq (1 \pm \epsilon)\frac{k}{d} \cdot 1\right] \leq e^{\frac{-k}{2}(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3})}$$

- Hence, if $k \geq \frac{4+2\beta}{\epsilon^2/2-\epsilon^3/3} \log n$, this probability becomes smaller than $\frac{2}{n^{2+\beta}}$ which is inverse polynomial w.r.t n.

Expected Value of $\|f(\alpha)\|^2$

- On expectation the length of a unit vector α is preserved.

$$
\begin{aligned}
E\left[\|f(\alpha)\|^2\right] &= E\left[\sum_{i=1}^{k}\frac{d}{k}\left(\sum_{j=1}^{d}X_j\alpha_j\right)^2\right] \\
&= \frac{d}{k}\sum_{i=1}^{k}\left(\sum_{j=1}^{d}E[X_j^2]\alpha_j^2+\sum_{j<l}^{d}E[X_jX_l]\alpha_j\alpha_l\right) \\
&= \frac{d}{k}\sum_{i=1}^{k}\frac{1}{d}=1=\|\alpha\|^2
\end{aligned}
$$

Deviation from Expectation: Proof of Main Theorem

- By Markov inequality,

$$\Pr\left[S > (1+\epsilon)\frac{k}{d}\right] < E\left[e^{hS}\right] e^{-(1+\epsilon)\frac{hk}{d}}$$

$$\Pr\left[S < (1-\epsilon)\frac{k}{d}\right] < E\left[e^{-hS}\right] e^{(1-\epsilon)\frac{hk}{d}}$$

Deviation from Expectation: Proof of Main Theorem

- By Markov inequality,

$$\Pr\left[S > (1+\epsilon)\frac{k}{d}\right] < E\left[e^{hS}\right] e^{-(1+\epsilon)\frac{hk}{d}}$$
$$\Pr\left[S < (1-\epsilon)\frac{k}{d}\right] < E\left[e^{-hS}\right] e^{(1-\epsilon)\frac{hk}{d}}$$

- Since the vectors $R_i's$ are all chosen independently we can rewrite the above as

$$\Pr\left[S > (1+\epsilon)\frac{k}{d}\right] < \left(E\left[e^{hQ_1^2}\right]\right)^k e^{-(1+\epsilon)\frac{hk}{d}}$$
$$\Pr\left[S < (1-\epsilon)\frac{k}{d}\right] < \left(E\left[e^{-hQ_1^2}\right]\right)^k e^{(1-\epsilon)\frac{hk}{d}}$$

where $Q_1 = \langle \alpha, R_1 \rangle$

Proof of Main Theorem

- By Taylor's Expansion,

$$\Pr\left[S < (1-\epsilon)\frac{k}{d}\right] < \left(E\left[1 - hQ_1^2 + \frac{hQ_1^4}{2}\right]\right)^k e^{-(1+\epsilon)\frac{hk}{d}}$$
$$= \left(1 - \frac{h}{d} + \frac{h^2 E[Q_1^4]}{2}\right)^k e^{(1-\epsilon)\frac{hk}{d}}$$

Lemma

For $h \in [0, d/2)$ and all $d \geq 1$,

$$E\left[e^{hQ_1^2}\right] \leq \frac{1}{\sqrt{1-2h/d}} \quad (1)$$

$$E\left[Q_1^4\right] \leq \frac{3}{d^2} \quad (2)$$

Proof of Main Theorem using Inequalities (1) and (2)

- If we take $h = \frac{d\epsilon}{2(1+\epsilon)}$, for the upper bound we have the following:

$$\begin{aligned} \Pr\left[S > (1+\epsilon)\frac{k}{d}\right] &< \left(\frac{1}{\sqrt{1-2h/d}}\right)^k e^{-(1+\epsilon)\frac{hk}{d}} \\ &= \left((1+\epsilon)e^{-\epsilon}\right)^{k/2} < e^{\frac{-k}{2}\left(\frac{\epsilon^2}{2}-\frac{\epsilon^3}{3}\right)}. \end{aligned}$$

Proof of Main Theorem using Inequalities (1) and (2)

- If we take $h = \frac{d\epsilon}{2(1+\epsilon)}$, for the upper bound we have the following:

$$\begin{aligned}\Pr\left[S > (1+\epsilon)\frac{k}{d}\right] &< \left(\frac{1}{\sqrt{1-2h/d}}\right)^k e^{-(1+\epsilon)\frac{hk}{d}} \\ &= ((1+\epsilon)e^{-\epsilon})^{k/2} < e^{\frac{-k}{2}(\frac{\epsilon^2}{2}-\frac{\epsilon^3}{3})}.\end{aligned}$$

- For the same value of h, for the lower bound we get:

$$\begin{aligned}\Pr\left[S < (1-\epsilon)\frac{k}{d}\right] &< \left(1 - h/d + \frac{3h^2}{2d^2}\right)^k e^{(1-\epsilon)\frac{hk}{d}} \\ &< e^{\frac{-k}{2}(\frac{\epsilon^2}{2}-\frac{\epsilon^3}{3})}.\end{aligned}$$

Proof of Inequality (2)

- For inequality (2)
  $$E[Q_1^4] = (\sum_{i=1}^{d} X_i\alpha_i)^4 = \sum_i E[X_i^4]\alpha_i^4 + \binom{4}{1,3}\sum_{i<j} E[X_i^3]E[X_j]\alpha_i^3\alpha_i + \binom{4}{2,2}\sum_{i<j} E[X_i^2]E[X_j^2]\alpha_i^2\alpha_j^2 + \binom{4}{2,1,1}\sum_{i<j<k} E[X_i^2]E[X_j]E[X_k]\alpha_i^2\alpha_j\alpha_k + \binom{4}{1,1,1,1}\sum_{i<j<k<l} E[X_i]E[X_j]E[X_k]E[X_l]\alpha_i\alpha_j\alpha_k\alpha_l$$
  $$= \frac{1}{d^2}(\alpha^4 + 6\sum_{i<j}\alpha_i^2\alpha_j^2) \leq \frac{3}{d^2}.$$

Proof of Inequality (1)

- The idea is to first make the random variable Q_1 independent of α and then compare the even moments of Q_1 with a properly scaled normal distribution.

Lemma (Worst Vector Lemma)

For all unit vectors α, $E[Q_1^{2k}(\alpha)] \leq E[Q_1^{2k}(w)]$, where $w = \frac{1}{\sqrt{d}}(1, 1, \ldots, 1)$ for $k = 1, 2, \ldots$.

Lemma (Normal Bound Lemma)

If $T \sim N(0, 1/d)$, then $E[Q_1^{2k}(w)] \leq E[T^{2k}]$, where $w = \frac{1}{\sqrt{d}}(1, 1, \ldots, 1)$ for $k = 1, 2, \ldots$.

Proof of Inequality (1)

$$
\begin{aligned}
E\left[e^{hT^2}\right] &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\lambda^2/2} e^{h\lambda^2/d} d\lambda \\
&= \frac{1}{\sqrt{1-2h/d}} \\
&= E\left[\sum_{k=0}^{\infty} \frac{h^k T^{2k}}{k!}\right] \qquad \text{(using MCT)} \\
&= \sum_{k=0}^{\infty} \frac{h^k E\left[T^{2k}\right]}{k!} \\
&\geq \sum_{k=0}^{\infty} \frac{h^k E\left[Q_1^{2k}(w)\right]}{k!} = E\left[e^{hQ_1(w)^2}\right] \geq E\left[e^{hQ_1(\alpha)^2}\right]
\end{aligned}
$$

Proving the Worst Vector Lemma

- Let r_1 and r_2 be i.i.d. r.v. distributed as $\{-1, +1\}$ with equal probability. Furthermore let a, b, T be any reals and $c = \sqrt{(a^2 + b^2)/2}$ and $k > 0$ be any integer, then

$$E\left[(T + ar_1 + br_2)^{2k}\right] \leq E\left[(T + cr_1 + cr_2)^{2k}\right]$$

Proving the Worst Vector Lemma

- Let r_1 and r_2 be i.i.d. r.v. distributed as $\{-1, +1\}$ with equal probability. Furthermore let a, b, T be any reals and $c = \sqrt{(a^2 + b^2)/2}$ and $k > 0$ be any integer, then

$$E\left[(T + ar_1 + br_2)^{2k}\right] \leq E\left[(T + cr_1 + cr_2)^{2k}\right]$$

- Let $R_1 = \frac{1}{\sqrt{d}}(r_1, r_2, \ldots, r_d)$. Thus we have

$$\begin{aligned} E\left[Q_1(\alpha)^{2k}\right] &= \frac{1}{d^k}\sum_R E\left[(R + \alpha_1 r_1 + \alpha_2 r_2)^{2k}\right] \Pr\left[\sum_{i=3}^{d} \alpha_i r_i = \frac{R}{\sqrt{d}}\right] \\ &\leq \frac{1}{d^k}\sum_R E\left[(R + cr_1 + cr_2)^{2k}\right] \Pr\left[\sum_{i=3}^{d} \alpha_i r_i = \frac{R}{\sqrt{d}}\right] \\ &= E\left[Q_1(\theta)^{2k}\right] \end{aligned}$$

where $c = \sqrt{(\alpha_1^2 + \alpha_2^2)/2}$

Proving the Worst Vector Lemma

- Let r_1 and r_2 be i.i.d. r.v. distributed as $\{-1, +1\}$ with equal probability. Furthermore let a, b, T be any reals and $c = \sqrt{(a^2 + b^2)/2}$ and $k > 0$ be any integer, then

$$E\left[(T + ar_1 + br_2)^{2k}\right] \leq E\left[(T + cr_1 + cr_2)^{2k}\right]$$

- Let $R_1 = \frac{1}{\sqrt{d}}(r_1, r_2, \ldots, r_d)$. Thus we have

$$\begin{aligned} E\left[Q_1(\alpha)^{2k}\right] &= \frac{1}{d^k} \sum_R E\left[(R + \alpha_1 r_1 + \alpha_2 r_2)^{2k}\right] \Pr\left[\sum_{i=3}^{d} \alpha_i r_i = \frac{R}{\sqrt{d}}\right] \\ &\leq \frac{1}{d^k} \sum_R E\left[(R + cr_1 + cr_2)^{2k}\right] \Pr\left[\sum_{i=3}^{d} \alpha_i r_i = \frac{R}{\sqrt{d}}\right] \\ &= E\left[Q_1(\theta)^{2k}\right] \end{aligned}$$

where $c = \sqrt{(\alpha_1^2 + \alpha_2^2)/2}$

- θ is a more "uniform" unit vector than α.

Proving the Normal Bound Lemma

- Let $\{T_i\}_{i=1}^{d}$ be i.i.d. normal r.v.. By stability of normal distribution

$$T = \frac{1}{d}\sum_{i=1}^{d} T_i \sim N(0, 1/d)$$

Proving the Normal Bound Lemma

- Let $\{T_i\}_{i=1}^{d}$ be i.i.d. normal r.v.. By stability of normal distribution

$$T = \frac{1}{d}\sum_{i=1}^{d} T_i \sim N(0, 1/d)$$

- We also have $Q_1(w) = \frac{1}{d}\sum_{i=1}^{d} r_1$

$$E[Q_1^{2k}(w)] = \frac{1}{d^{2k}}\sum_{i_1=1}^{d}\cdots\sum_{i_{2k}=1}^{d} E[r_{i_1}\ldots r_{i_{2k}}]$$

$$E[T^{2k}] = \frac{1}{d^{2k}}\sum_{i_1=1}^{d}\cdots\sum_{i_{2k}=1}^{d} E[T_{i_1}\ldots T_{i_{2k}}]$$

Proving the Normal Bound Lemma

- Let $\{T_i\}_{i=1}^{d}$ be i.i.d. normal r.v.. By stability of normal distribution

$$T = \frac{1}{d}\sum_{i=1}^{d} T_i \sim N(0, 1/d)$$

- We also have $Q_1(w) = \frac{1}{d}\sum_{i=1}^{d} r_1$

$$E[Q_1^{2k}(w)] = \frac{1}{d^{2k}}\sum_{i_1=1}^{d}\cdots\sum_{i_{2k}=1}^{d} E[r_{i_1}\ldots r_{i_{2k}}]$$

$$E[T^{2k}] = \frac{1}{d^{2k}}\sum_{i_1=1}^{d}\cdots\sum_{i_{2k}=1}^{d} E[T_{i_1}\ldots T_{i_{2k}}]$$

- For each index assignment we have

$$E[r_{i_1}\ldots r_{i_{2k}}] \leq E[T_{i_1}\ldots T_{i_{2k}}]$$

Open questions

- Plenty !

Open questions

- Plenty !
- No-flattening results for other l_p metrics, non metrics

Open questions

- Plenty !
- No-flattening results for other l_p metrics, non metrics
- Embeddability of non-metrics into metric spaces - useful in databases, learning

Open questions

- Plenty !
- No-flattening results for other l_p metrics, non metrics
- Embeddability of non-metrics into metric spaces - useful in databases, learning
- Information Theoretic Metrics - KL, Bhattacharyya, Mahalanobis - widely used

THANK YOU