“Dimension versus Distortion”

and

“Intrinsic Dimension of Metric Spaces”

Anupam Gupta
Carnegie Mellon University

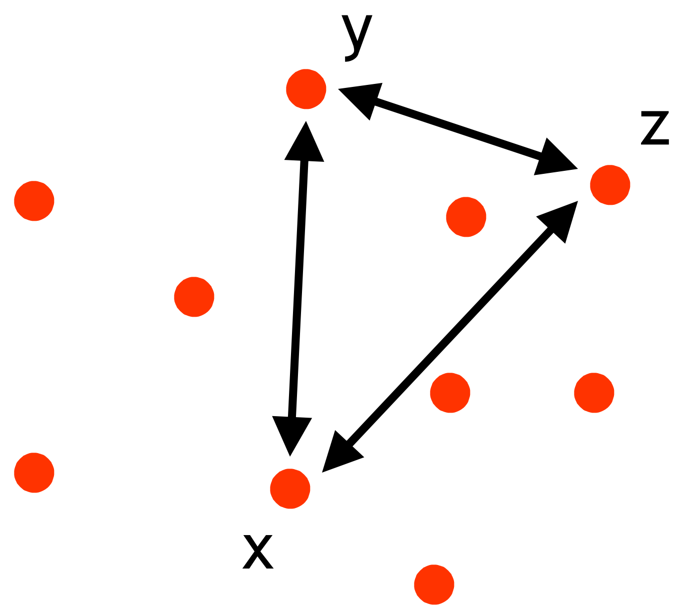
Metric space $M = (V, d)$

set V of points

distances $d(x,y)$

triangle inequality

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



in the previous two lectures…

We saw:

- every metric is (almost) a (random) tree
- every metric is (almost) an Euclidean metric

Today:

- low-dimensional embeddings
- low-dimensional metric spaces

why dimension reduction?

2-dimensional Euclidean space is *simple*

3-dim space is not so bad

4-dim ??

.

.

.

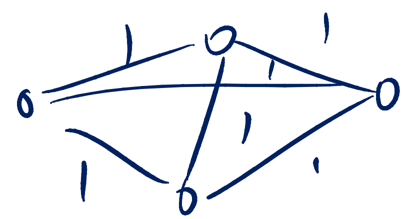
100000-dim space may be a bit much

harder for algorithms as well!

the “curse of dimensionality”

many (geometric) algorithms have running times that are exponential in the dimension.

so can we reduce the dimensions without changing the distances?



how about only changing the distances *by a small amount*?

hasn't this been solved?

There are many approaches to doing dimension reduction.

Many based on singular-value decompositions

e.g., PCA and many variants, ICA, factor analysis

also, multidimensional scaling…

However, they try to minimize some other notion of goodness

does not work for *distortion*

could stretch/shrink distances by large factors.

central question

How do we reduce the dimension
so that every distance is distorted by
only a small factor?

natural tension between dimension and distortion

exercise #2, problem 5

5. Dimension versus Distortion. The n-point uniform metric $U_n = (V, d)$ has interpoint distances $d(x, y) = 1$ for all $x \neq y \in V$.

a. Show that any embedding of U_n into $\mathbb{R}^k$ incurs a distortion of at least $\Omega(n^{1/k})$.

Hint: again, consider the vectors $x_1, x_2, \ldots, x_n$ giving the embedding. Suppose the map only expands distances, and the expansion of this map is D. What are the largest open balls around the x_i's which are disjoint? What is the smallest ball you can draw that is guaranteed to contain all the points, if the distortion is at most D?

If the volume of a radius-r ball in $\mathbb{R}^k$ is $c_k r^k$ for some constant c_k that depends only on k, what inequality does this give you?

b. Hence, if the distortion is at most $(1 + \varepsilon)$, then $k \geq \frac{\log n}{\epsilon}$.

the Johnson Lindenstrauss lemma

Theorem [1984]

Given any n-point subset V of Euclidean space R^m, one can map into $O(\log n/\epsilon^2)$-dimensional space while incurring a distortion at most $(1+\epsilon)$.

There exists such a map that is a linear map.

What is this map?

Let $k = O(\log n/\epsilon^2)$

- Choose a uniformly random k-dimensional subspace and project the points down onto it.

  The distance between the projections is expected to be about sqrt{k/n} of their original distances.

  One can prove that the distances are within $(1+\epsilon)$ of this expected value with high probability.

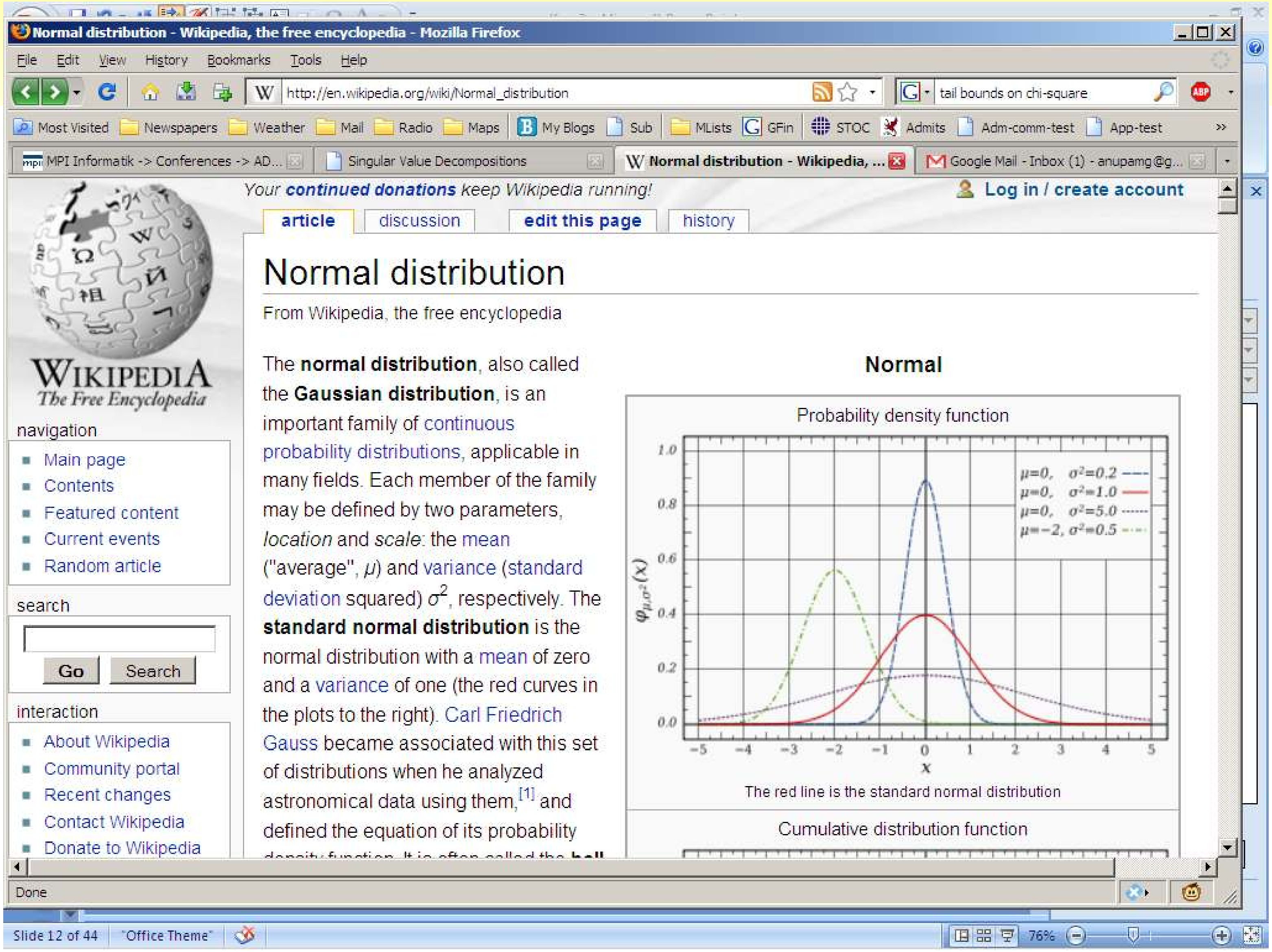
[JL, FM, DG]

Easier to deal with

Let $k = O(\log n/\epsilon^2)$

$N(\mu, \sigma^2) \sim \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

- Take a $k \times m$ matrix A
  each entry is an independent N(0,1) random value.
- the linear map is F(x) = Ax
- For any vector x and row a of A, $a.x \sim N(0,|x|^2)$
- For any vector x, $E[|Ax|^2] = k|x|^2$
- $\Pr[\ |Ax|^2 \text{ in } (1 \pm \epsilon)\, k|x|^2] \geq$



Normal distribution

From Wikipedia, the free encyclopedia

The **normal distribution**, also called the **Gaussian distribution**, is an important family of continuous probability distributions, applicable in many fields. Each member of the family may be defined by two parameters, *location* and *scale*: the mean ("average", μ) and variance (standard deviation squared) σ^2, respectively. The **standard normal distribution** is the normal distribution with a mean of zero and a variance of one (the red curves in the plots to the right). Carl Friedrich Gauss became associated with this set of distributions when he analyzed astronomical data using them,[1] and defined the equation of its probability

- Their product XY follows a distribution with density p given by

$$p(z) = \frac{1}{\pi\, \sigma_X\, \sigma_Y}\; K_0\!\left(\frac{|z|}{\sigma_X\, \sigma_Y}\right),$$ where K_0 is a modified Bessel function of the second kind.

- Their ratio follows a Cauchy distribution with $X/Y \sim \mathrm{Cauchy}(0, \sigma_X/\sigma_Y)$. Thus the Cauchy distribution is a special kind of ratio distribution.

4. If $X_1, \ldots, X_n$ are independent standard normal variables, then $X_1^2 + \cdots + X_n^2$ has a chi-square distribution with n degrees of freedom.
5. If $X_1, \ldots, X_n$ are independent standard normal variables, then the sample mean $\bar{X} = (X_1 + \cdots + X_n)/n$ and sample variance $S^2 = ((X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2)/(n-1)$ are independent. This property characterizes normal distributions (and helps to explain why the F-test is non-robust with respect to non-normality!)

Standardizing normal random variables [edit]

As a consequence of Property 1, it is possible to relate all normal random variables to the standard normal.

If $X \sim N(\mu,\sigma^2)$, then

$$Z = \frac{X - \mu}{\sigma}$$

is a standard normal random variable: $Z \sim N(0,1)$. An important consequence is that the cdf of a general normal

google

Note that $k\hat{d}_{l_2}/d_{l_2}$ follows a Chi-square distribution with k degrees of freedom, χ^2_k. Therefore, it is easy to prove the following Lemma about the tail bounds:

Lemma 1

$$\mathbf{Pr}(\hat{d}_{l_2} - d_{l_2} \geq \varepsilon d_{l_2}) \leq \exp\left(-\frac{k}{2}(\varepsilon - \log(1+\varepsilon))\right) = \exp\left(-k\frac{\varepsilon^2}{G_R}\right), \quad \varepsilon > 0,$$

$$\mathbf{Pr}(\hat{d}_{l_2} - d_{l_2} \leq -\varepsilon d_{l_2}) \leq \exp\left(-\frac{k}{2}(-\varepsilon - \log(1-\varepsilon))\right) = \exp\left(-k\frac{\varepsilon^2}{G_L}\right), \quad 0 < \varepsilon < 1,$$

where the constants

Easier to deal with

Let $k = O(\log n/\epsilon^2)$

- Take a $k \times m$ matrix A

  each entry is an independent N(0,1) random value.
- the linear map is $F(x) = Ax$
- For any vector x and row a of A, $a.x \sim N(0,|x|^2)$
- For any vector x, $E[|Ax|^2] = k|x|^2$
- $\Pr[\ |Ax|^2 \text{ in } (1 \pm \epsilon)\, k|x|^2] \geq\ 1 - \exp\{-\Omega(\epsilon^2 k)\}$

  $\geq 1 - \frac{1}{\text{poly}(n)}$
- So can take union bound over all n^2 difference vectors.

[IM]

proof

embedding $\ell 2$ into $\ell 1$

Even easier to do

Database-Friendly Projections [A '01]

Again, let A be a k $\times$ m matrix

now fill it with uniformly random ±1 values.

(lower density: 0 w.p. 2/3, uniformly random ±1 otherwise.)

Again F(x) = Ax

Proof along similar general lines, but more involved.

Question: can we reduce the support of A any further?

faster maps

The time to compute the map was O(mk) = O(m log n).

"Fast JL Transform" [AC]

Time $O(m \log m + \epsilon^{-2} \log^3 n)$

Basic idea: uncertainty principle.

density of JL/DFP matrices can be reduced if x is "spread out"

support of x, Hx cannot both be small (H = Hadamard matrix)

so map F(x) = AHx

(actually AHDx, D is random diagonal sign matrix).

and faster still

The time to compute the map was $O(mk) = O(m \log n)$.

"Fast JL Transform" [AC]

Time $O(m \log m + k^3)$

[AL] $O(m \log k)$

[LAS] $O(m)$ --- for "well-spread-out" x.

dimension-distortion tradeoffs

What if we have a dimension k in mind?

exercise showed:

$\Omega(n^{1/k})$ distortion

best results:

$O(n^{2/\cancel{d}\,k} \log^{3/2} n)$ distortion

$\Omega(n^{1/\lfloor (k+1)/2 \rfloor})$

very interesting new results:

NP-hard to $n^{c/d}$-approximate distortion for all constant d.

tighter lower bound

Theorem [1984]

Given any n-point subset V of Euclidean space, one can map into $O(\log n/\epsilon^2)$-dimensional space while incurring a distortion at most $(1+\epsilon)$.

We've already seen a lower bound of $\Omega(\log n/\epsilon)$.

Theorem (Alon):

Lower bound of $\Omega(\log n/(\epsilon^2 \log \epsilon^{-1}))$

again for the "uniform" metric U_n

so what do we do now?

Note that we're again proving *uniform* results.
 ("All n-point Euclidean metrics…")

Can we give a better "per-instance" guarantee?

If the metric contains a copy of U_t
 we get a lower bound of $\Omega(\log t)$ dimensions
 if we want $O(1)$ distortion

But this is somewhat fragile
 (what if there's an *almost* uniform metric?)…

the doubling dimension

convenient way of capturing the property that
there are no large near-uniform metrics

gives us a notion of “intrinsic dimension” of a point set in Euclidean space

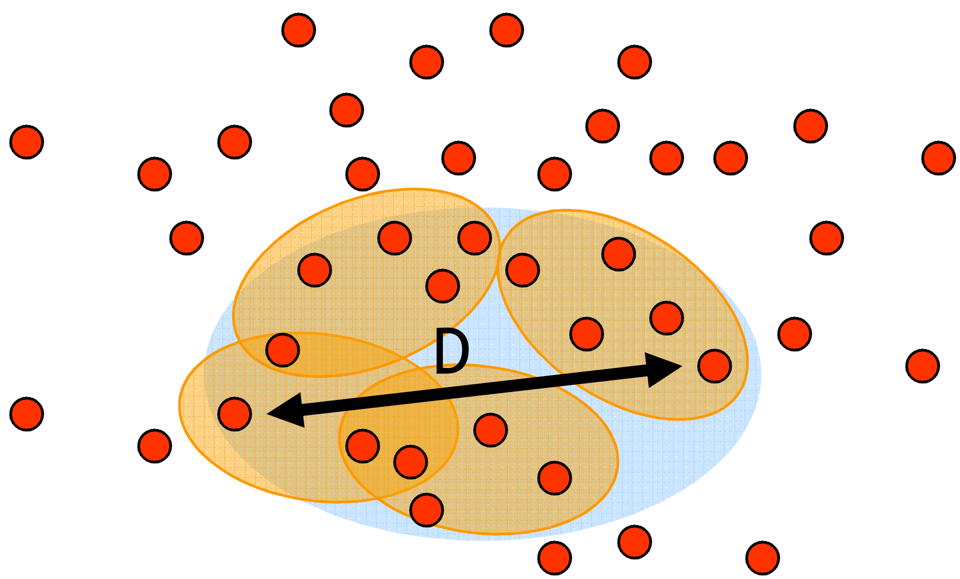
in fact, doubling dimension is defined for *any* metric space, not just for point sets in Euclidean space.

the doubling dimension

Dimension $\dim_D(M)$ is the smallest k such that

every set S with diameter D_S

can be covered by 2^k sets of diameter $\frac{1}{2}D_S$

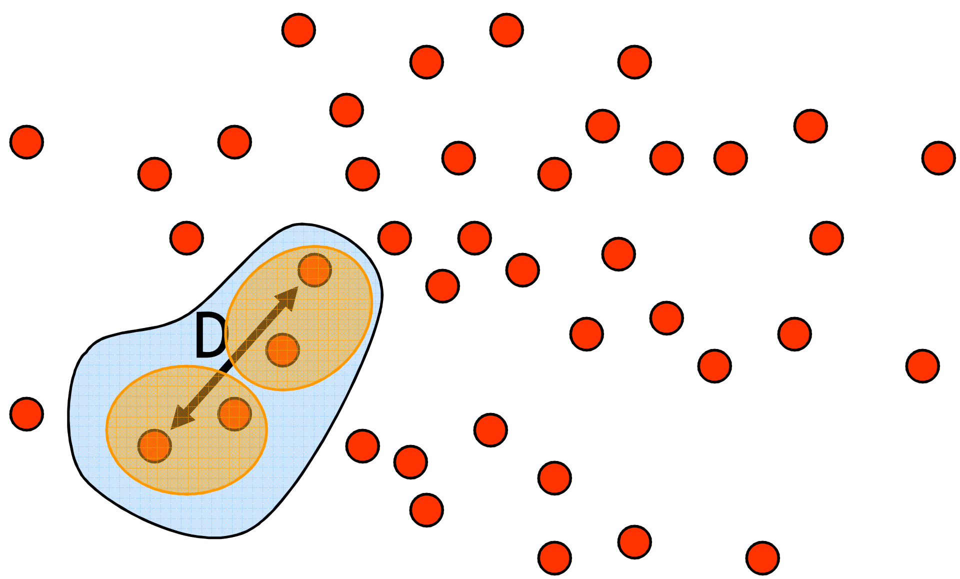


the doubling dimension

Dimension $\dim_D(M)$ is the smallest k such that

every set S with diameter D_S

can be covered by 2^k sets of diameter $\frac{1}{2}D_S$



Assouad

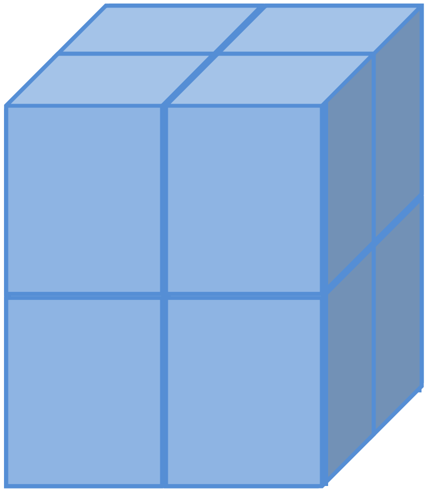
doubling generalizes geometric dimension

Take k-dim Euclidean space R^k

Claim: $\dim_D(R^k) \approx \Theta(k)$

Easy to see for boxes

Argument for spheres a bit more involved.



2^3 boxes to cover larger box in R^3

doubling metrics

Dimension at most k if

every set S with diameter D_S can be covered by 2^k sets of diameter $\frac{1}{2}D_S$

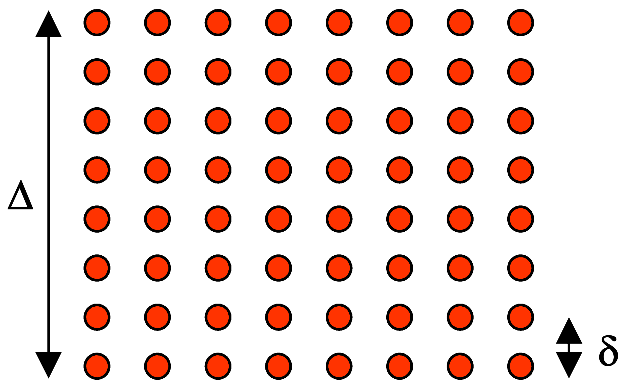
A family of metric spaces is called "doubling"

if there exists a constant k such that the doubling dimension of these metrics is bounded by k.

what is not a doubling metric?

The uniform metric U_t on t points have dimension $\Omega(\log t)$

small near-uniform metrics



this 2-dim set has $(\Delta/d)^2$ points

Suppose a metric (X,d) has doubling dimension k.

If any subset $S \subseteq X$ of points has

all inter-point distances lying between δ and Δ

there are at most $(\Delta/\delta)^{O(k)}$ points in S.

the (simple) proof

advantages of this fact

Thm: Doubling metrics admit O(dim(M))-padded decompositions

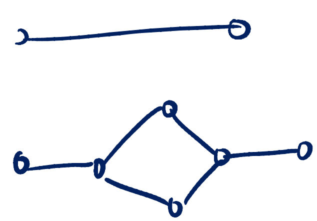
$\Rightarrow$ doubling metrics embed into ℓ2 with distortion sqrt{log n}.

Q: Do all doubling metrics embed into ℓ2 with distortion O(1)?

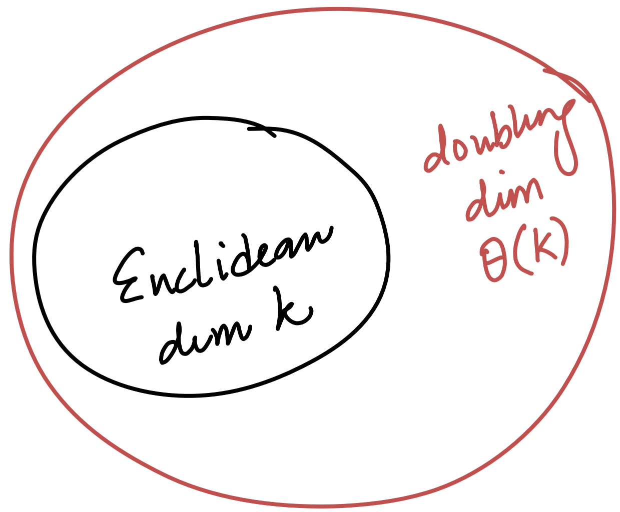
$\sqrt{\log n}$

G_0

G_1



$G_{\log_2 n}$



Many geometric algorithms can be extended to doubling spaces…

Near neighbor search
Compact routing
Distance labeling
Network triangulation
Sensor placements

Small-world networks
Traveling Salesman
Sparse Spanners
Approx. inference
Network Design

Clustering problems
Well-separated pair decomposition
Data structures
Learnability

example application

Assign labels $L(x)$ to each host x in a metric space
Looking just at $L(x)$ and $L(y)$, can infer distance $d(x,y)$

Results

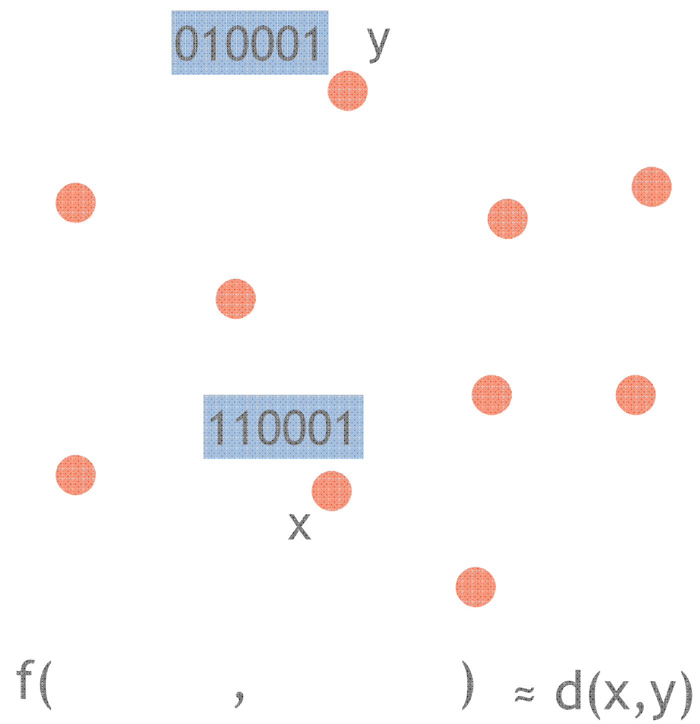
labels with $(O(1)/\varepsilon)^{\dim} \times \log n$ bits

estimates within $(1 + \varepsilon)$ factor

Contrast with

lower bound of n bit labels

in general for any factor < 2



another example application

For example:

[Arora 95] showed that TSP on R^k was $(1+\epsilon)$-approximable in time

$n^{\left(\frac{\log n}{\epsilon}\right)^{O(k)}}$

$n^{\left(1/\epsilon\right)^{O(k)}}$

[Talwar 04] extended this result to metrics with doubling dimension k

??

example in action:
sparse spanners for doubling metrics

[Chan G. Maggs Zhou]

r-nets

$N \subseteq S$ is an r-net $\quad r > 0$

if (1) $d(u,v) \geq r \quad \forall u,v \in N$

(2) $\forall x \in S \quad \exists y \in N$

$d(x,y) \leq r$

<u>Fact</u>: N is an r-net

$N' \subseteq N$ of diam R

$\Rightarrow |N'| \leq (R/r)^{O(k)}$

the construction

$(d(x,y) \geq 1 \quad \forall\, x,y \in S)$

$x \in Y_0 = S$

$\rightarrow Y_1 = 2\text{-net of } Y_0$

$\rightarrow Y_2 = 4\text{-net of } Y_1$

$S = Y_0 \supseteq Y_1 \supseteq Y_2 \quad . \quad . \quad . \quad \supseteq Y_i$

$Y_{i+1} = 2^{i+1}\text{ net of } Y_i$

$\forall x \in S$

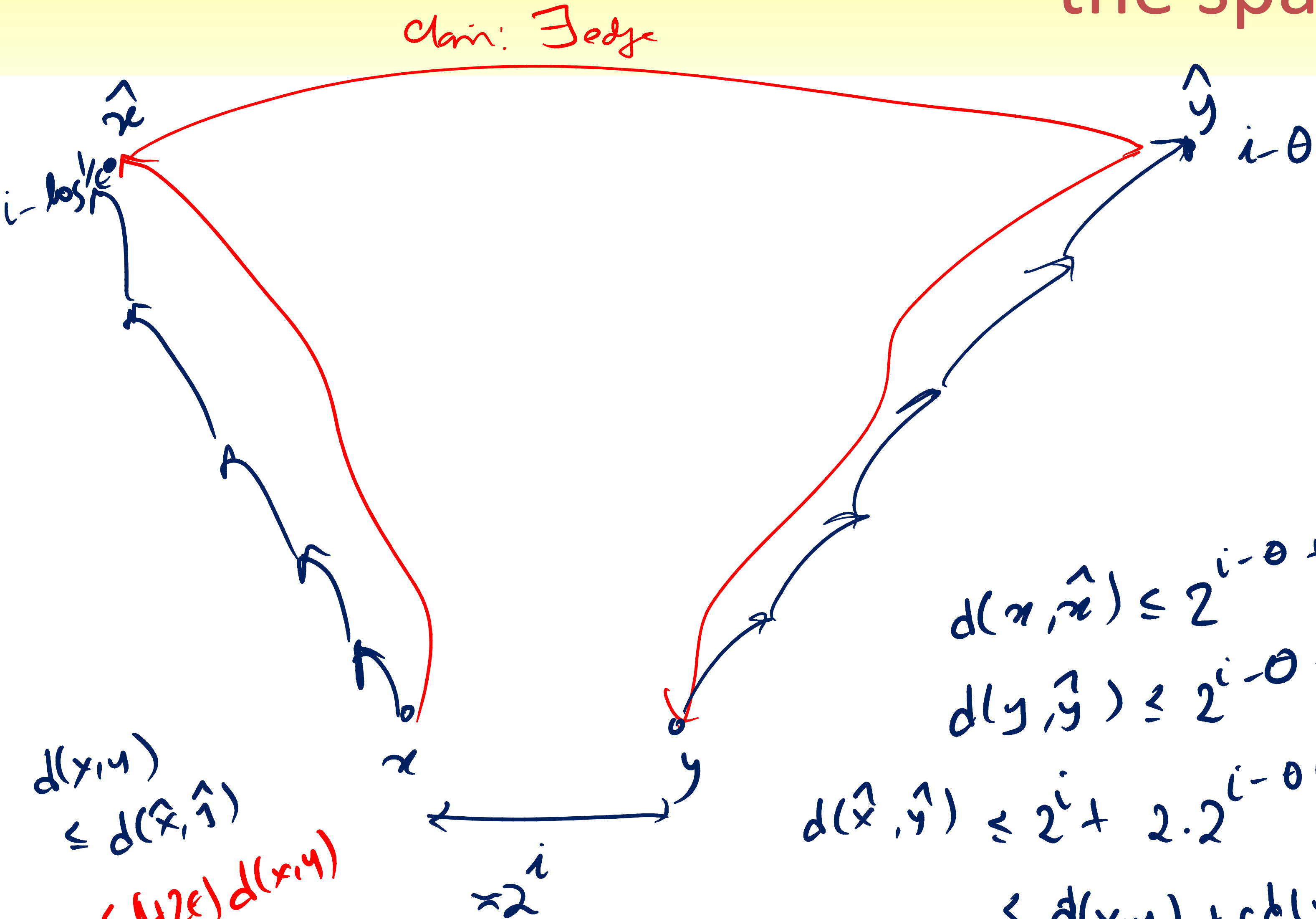
$\exists\, \hat{x} \in Y_i$

$d(x, \hat{x}) \leq 2^{i+1}$

$\vdots$

$|Y_{\log \mathrm{diam}}| = 1$

the sparsity



$d(x,y)$
$\leq d(\hat{x},\hat{y})$
$\leq (1+2\epsilon)\, d(x,y)$

$d(x,\hat{x}) \leq 2^{i-\theta+1}$

$d(y,\hat{y}) \leq 2^{i-\theta+1}$

$d(\hat{x},\hat{y}) \leq 2^i + 2 \cdot 2^{i-\theta+1}$

ϵ

$\leq d(x,y) + c\, d(x,y)\, 2^{-\theta}$

$\leq (1+\epsilon)\, d(x,y)$

the stretch

Back to dimensionality reduction

useful for distortion-dimension

Theorem: [Chan G. Talwar]

Any metric with doubling dimension k embeds into Euclidean space with T dimensions with distortion

$$O\left(\log n \cdot \sqrt{\frac{k}{T}}\right)$$

(where T $\in$ [k log log n, log n])

Can get very similar tradeoffs using results of [ABN'08]

For the case of

CGT, ABN

dimension	$\tilde{O}(\cancel{\dim}\ k)$	$O(\log n)$
Doubling	distortion $O(\log n)$	$\sqrt{\dim \cdot \log n}$
Doubling Euclidean	??	$O(1)$

JL

other notions of metric dimension

strong doubling dimension

correlation fractal dimension

negative-curvature

thank you!