Random projection trees and low dimensional manifolds

Sanjoy Dasgupta and Yoav Freund
University of California, San Diego

I. The new nonparametrics

The new nonparametrics

The traditional bane of nonparametric statistics is the curse of dimensionality.

For data in R^D: convergence rates $n^{-\Omega(1/D)}$

But recently, some sources of rejuvenation:

1. Data near low-dimensional manifolds
2. Sparsity in data space or parameter space

Low dimensional manifolds

Motion capture:

N markers on a human body yields data in R^{3N}

Benefits of intrinsic low-dimensionality

Benefits you need to work for

Learning the structure of the manifold

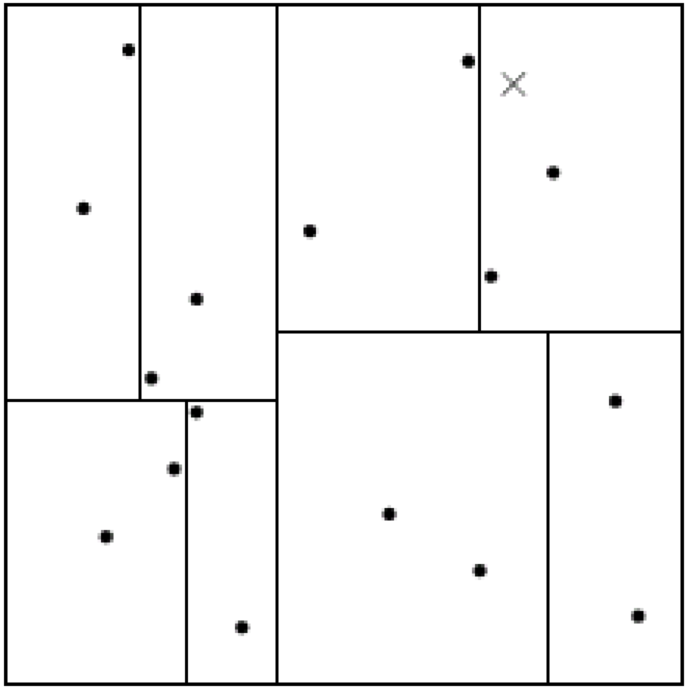
(a) Find explicit embedding $R^D \rightarrow R^d$, then work in low-dimensional space

(b) Use manifold structure for regularization

This talk:

Simple tweaks that make standard methods “manifold-adaptive”

The k-d tree

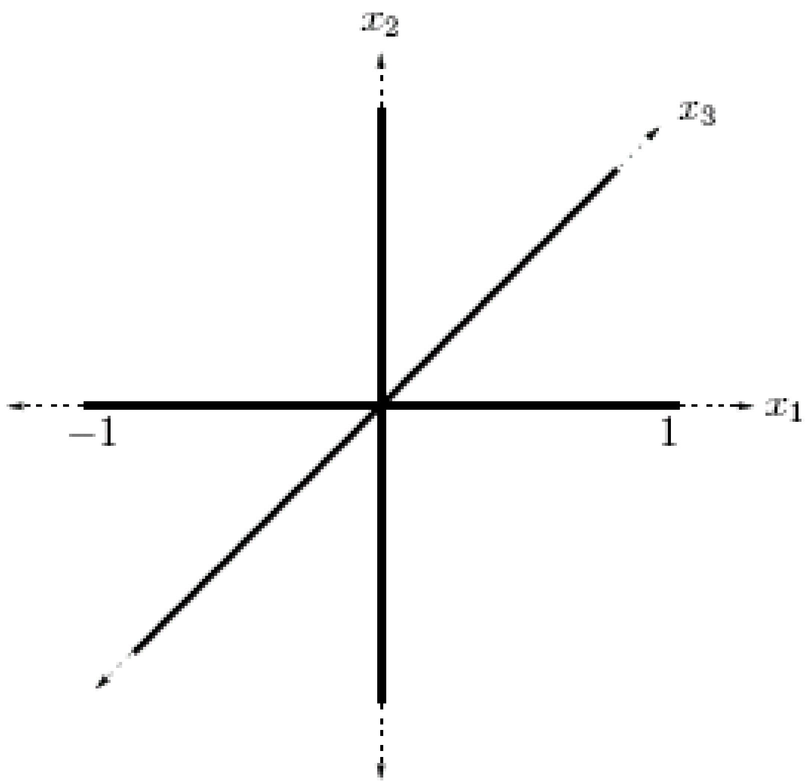


Problem: curse of dimensionality, as usual

Key notion in statistical theory of tree estimators: at what rate does cell diameter decrease as you move down the tree?

Rate of diameter decrease

Consider: $X = \cup_{i=1}^{D} \{te_i : \; -1 \leq t \leq 1\} \subset \mathbf{R}^D$

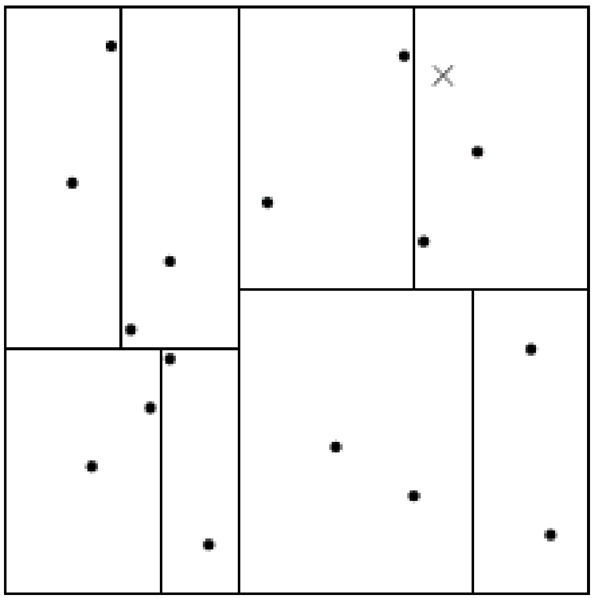


Need at least D levels to halve the diameter

Intrinsic dimension of this set is d = log D (or perhaps even 1, depending on your definition)

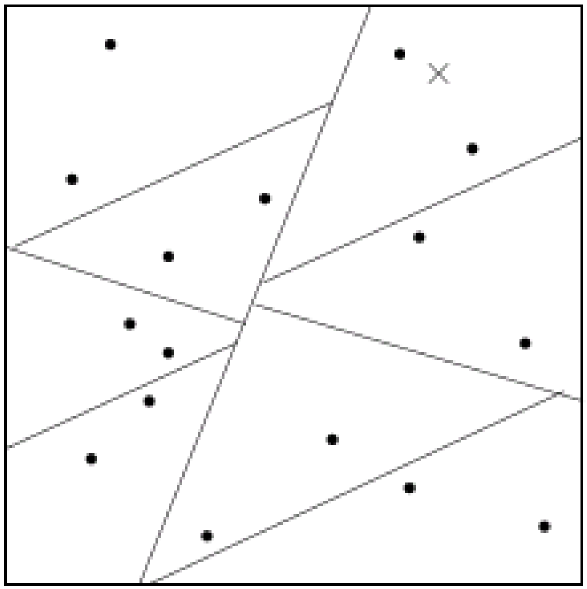
Random projection trees

K-d tree



Pick coordinate direction
Split at median

RP tree



Pick random direction
Split at median plus noise

If the data in R^D has intrinsic dimension d, then an RP tree halves the diameter in just d levels: no dependence on D.

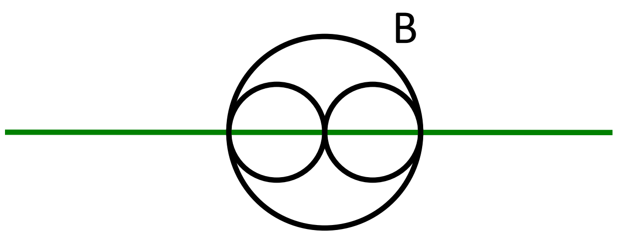
II. RP trees and Assouad dimension

Assouad dimension

Set $S \subset R^D$ has *Assouad dimension* $\leq$ d if: for any ball B, subset $S \cap B$ can be covered by 2^d balls of half the radius. Also called *doubling dimension*.

S = line

Assouad dimension = 1



S = k-dimensional affine subspace

Assouad dimension = O(k)

S = set of N points

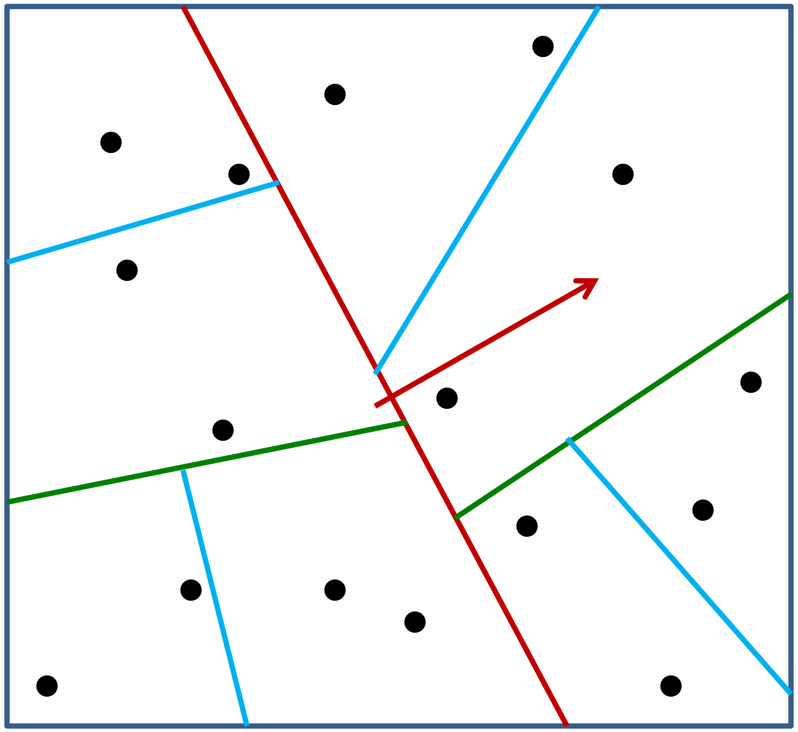
Assouad dimension $\leq$ log N

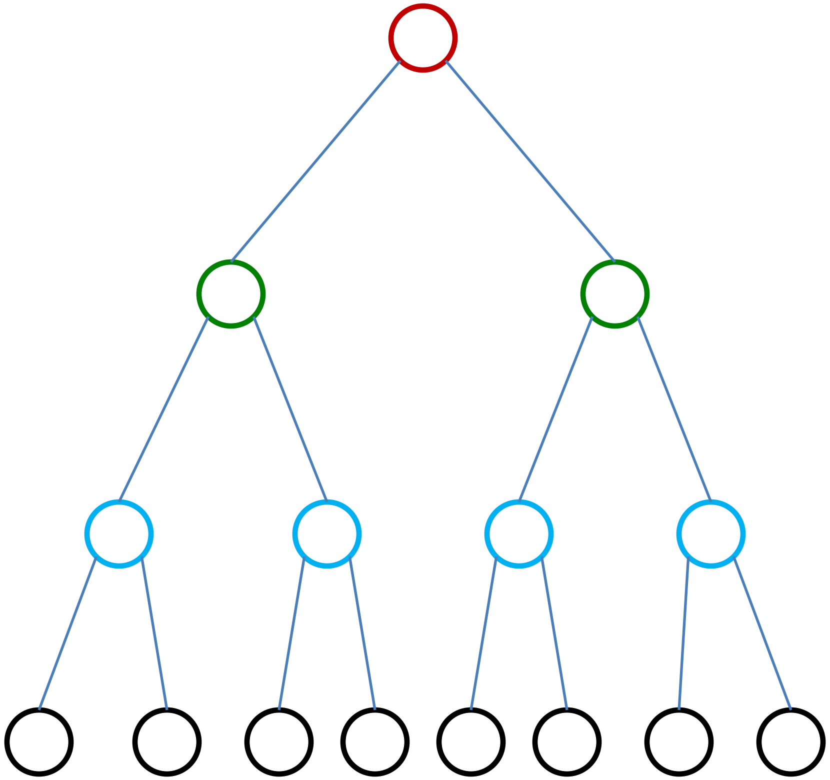
S = k-dim submanifold of R^D with finite *condition number*

Assouad dimension = O(k) in small enough neighborhoods

Crucially: if S has Assouad dim $\leq$ d, so do subsets of S

RP trees





Spatial partitioning
Cell

Binary tree
Node

RP tree algorithm

procedure MAKETREE(S)

if |S| < MinSize:

  return (Leaf)

else:

  Rule ← CHOOSERULE(S)

  LeftTree ← MAKETREE({x ∈ S : Rule(x) = true})

  RightTree ← MAKETREE({x ∈ S : Rule(x) = false})

  return ([Rule,LeftTree,RightTree])

procedure CHOOSERULE(S)

choose a random unit direction $v \in R^D$

pick any point x ∈ S, and let y be the farthest point from it in S

choose δ uniformly at random in $[-1, 1] \cdot 6 \; \|x - y\|/D^{1/2}$

Rule(x) := $x \cdot v \leq (\text{median}(\{z \cdot v : z \in S\}) + \delta)$

return (Rule)

Performance guarantee

There is a constant c_0 with the following property.

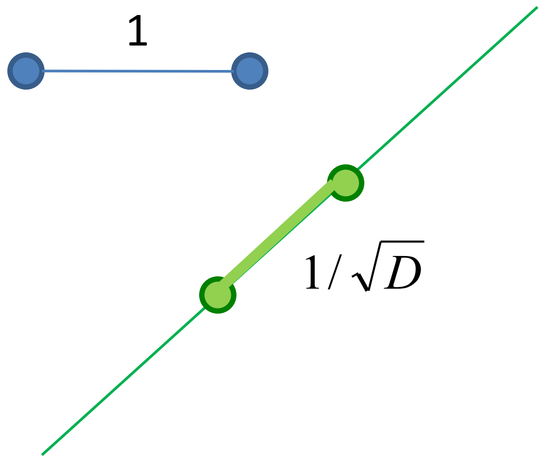
Build RP tree using data set $S \subset R^D$.

Pick any cell C in tree such that $S \cap C$ has Assouad dimension ≤ d.

Then, with prob $\geq 1/2$ (over construction of subtree rooted at C): for every descendant C′ that is more than c_0 d log d levels below C, we have *radius*(C′) ≤ *radius*(C)/2.

One-dimensional random projections

Projection from R^D onto (a random line) R^1: how does this affect the lengths of vectors? Very roughly: it shrinks them by $D^{1/2}$.



Lemma: Fix any vector $x \in R^D$. Pick a random unit vector $U \sim S^{D-1}$.

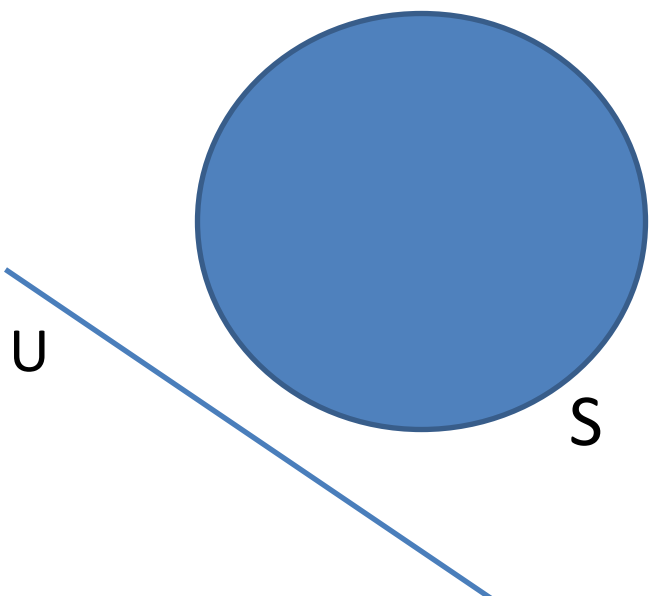
$$\text{(a)}\quad \Pr\left[|x \cdot U| \leq \alpha \cdot \frac{\|x\|}{\sqrt{D}}\right] \leq \sqrt{\frac{2}{\pi}}\alpha$$

$$\text{(b)}\quad \Pr\left[|x \cdot U| \geq \beta \cdot \frac{\|x\|}{\sqrt{D}}\right] \leq \frac{2}{\beta}e^{-\beta^2/2}$$

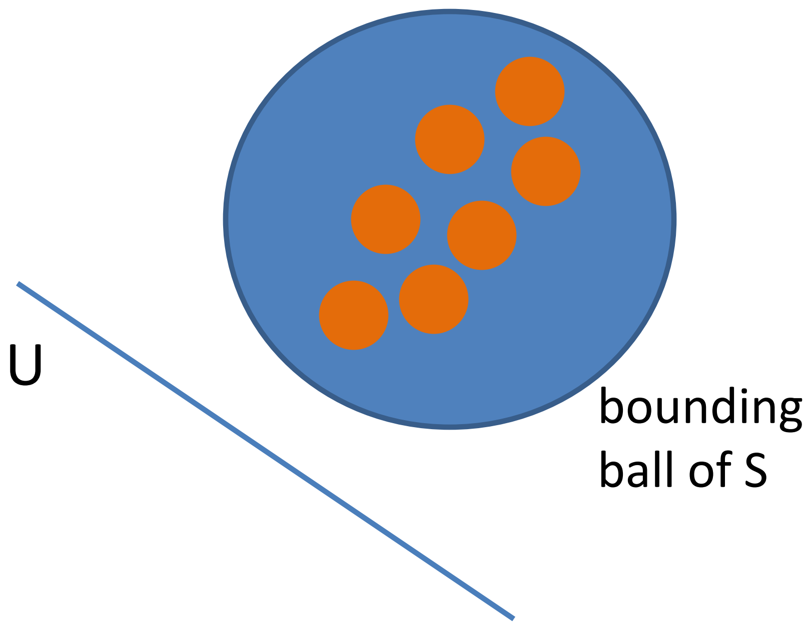
Effect of RP on diameter

Set $S \subset R^D$ is subjected to random projection U.
How does the diameter of $S \cdot U$ compare to that of S?

If S is full-dimensional:
$\text{diam}(S \cdot U) \leq \text{diam}(S)$.



If S has Assouad dimension d:
$\text{diam}(S \cdot U) \leq \text{diam}(S)\sqrt{d/D}$
(with high probability).

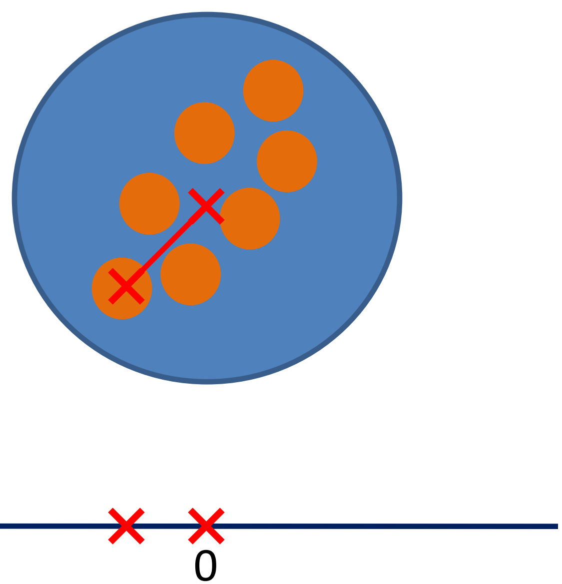


Diameter of projected set

$S \subset R^D$ has Assouad dim d. Pick random projection U. With high prob:

$$\mathrm{diam}(S \cdot U) \leq \mathrm{diam}(S) \cdot O\left(\sqrt{\frac{d \log D}{D}}\right)$$

Bounding ball of S is (wlog) B(0,1)



1. Can cover S by $(D/d)^{d/2}$ balls of radius $\sqrt{d/D}$
   Need 2^d balls of radius 1/2,
   4^d balls of radius 1/4,
   8^d balls of radius 1/8, ...,
   $(1/\epsilon)^d$ balls of radius ϵ

2. Pick any of these balls. Its projected center is fairly close to the origin.
   w.p. O(1): within $\sqrt{1/D}$
   w.p. $1\text{-}1/D^d$: within $\sqrt{(d \log D)/D}$

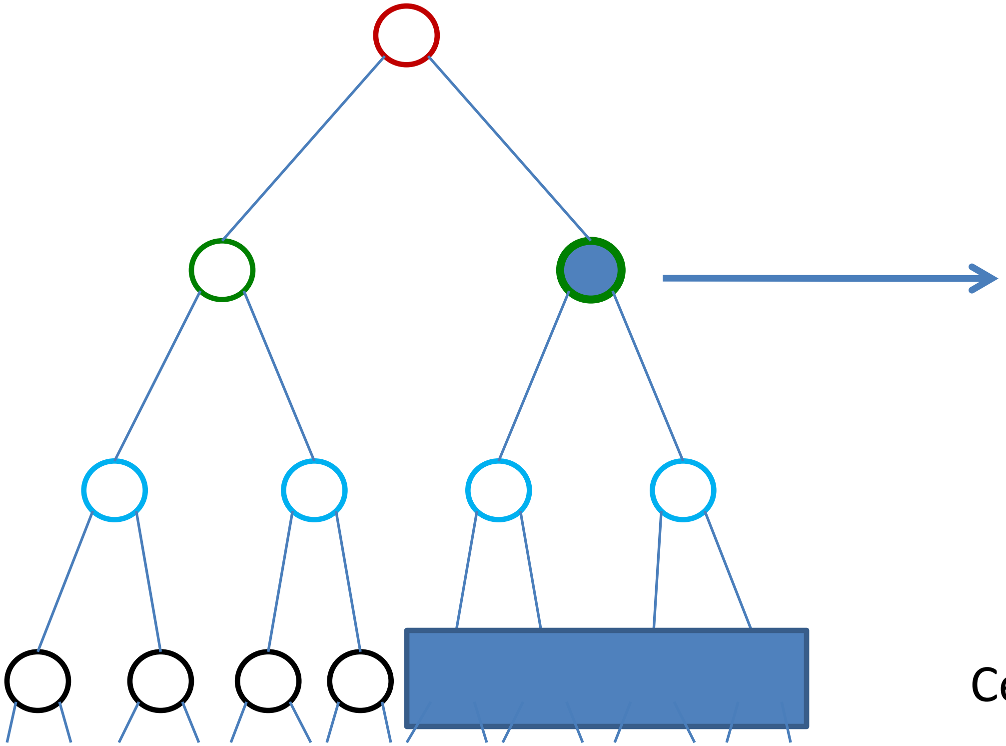
3. Do a union bound over all the balls.

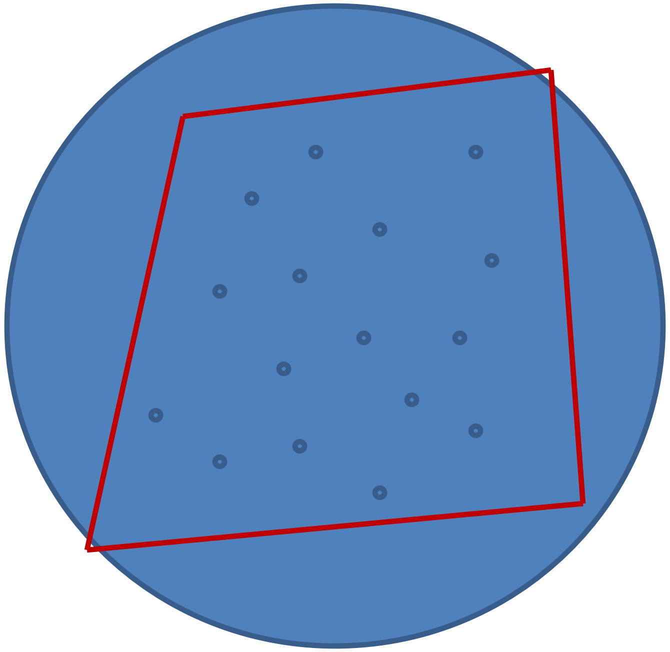
Proof outline

Pick any cell in the RP tree, and let $S \subset R^D$ be the data in it.

Suppose S has Assouad dim d and lies in a ball of radius 1.

Show: In every descendant cell d log d levels below, the data is contained in a ball of radius 1/2.

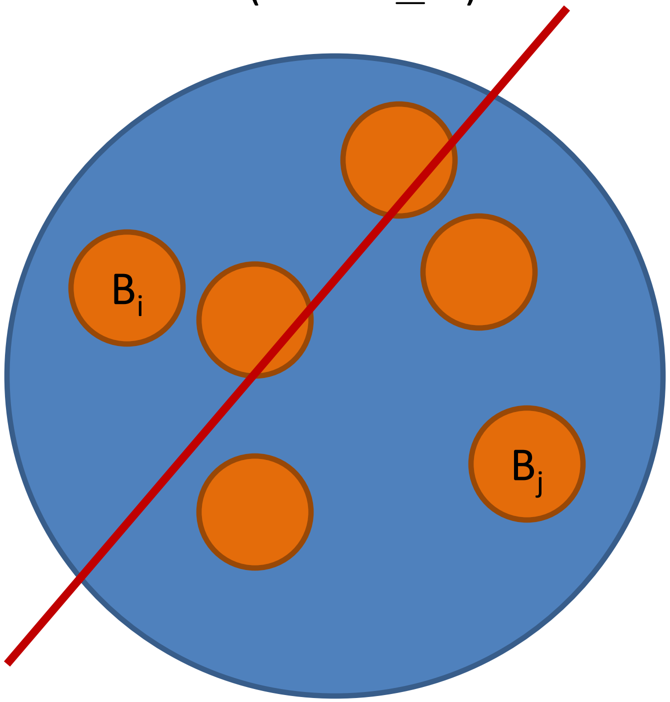




Cell has bounding ball of radius ≤ 1

Proof outline

Current cell (radius ≤ 1):



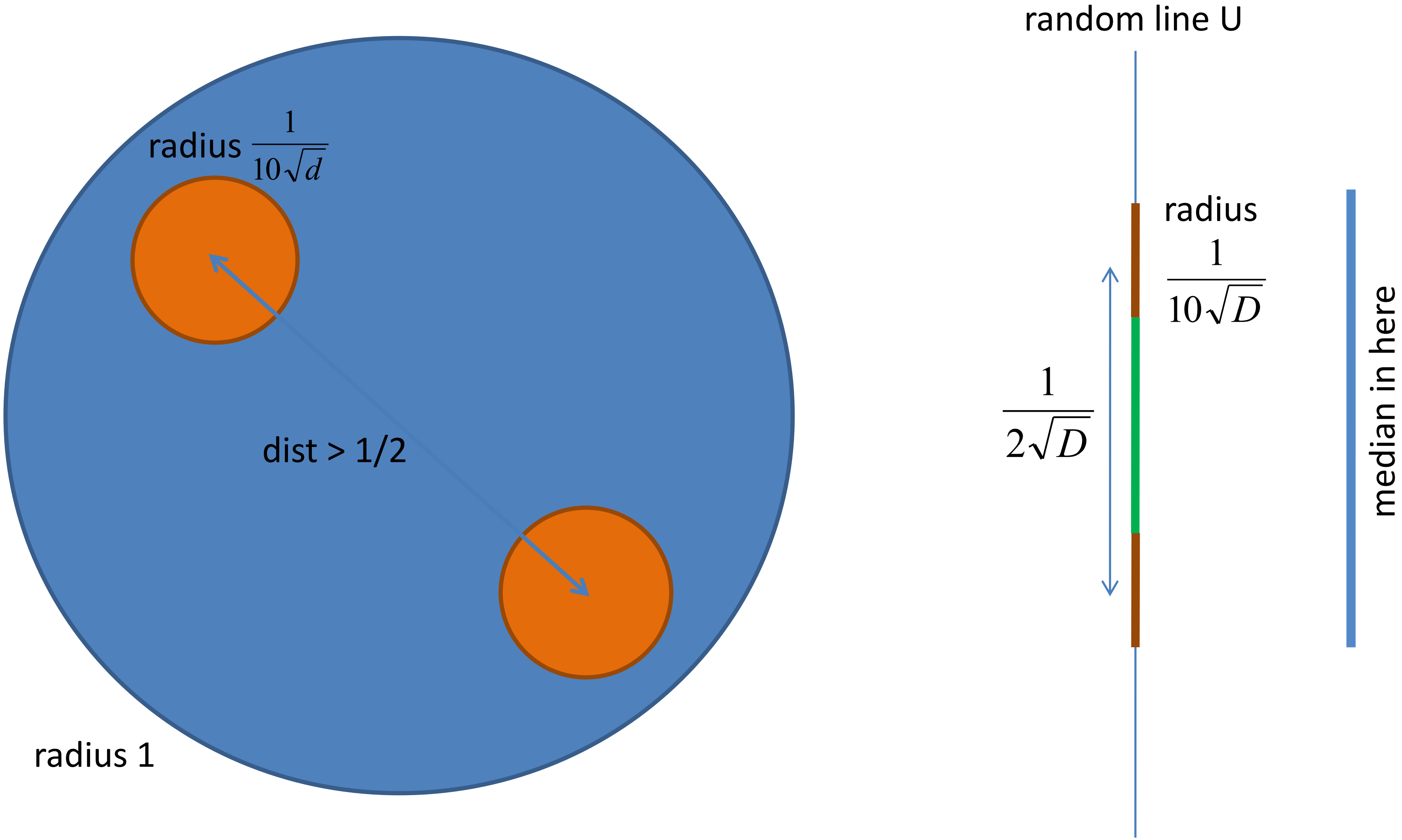
1. Cover S by $d^{d/2}$ balls B_i of radius $1/d^{1/2}$
2. Consider any pair of balls B_i, B_j at distance $\geq 1/2$ apart.

   A single random split has constant probability of cleanly separating them

3. There are at most d^d such pairs B_i, B_j

   So after d log d splits, every faraway pair of balls will be separated… which means all cells will have radius $\leq 1/2$

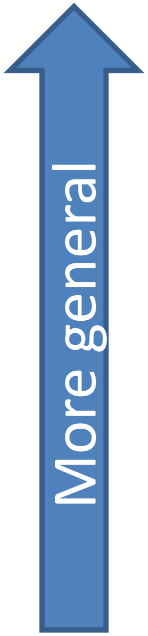
Big picture



Recall effect of random projection: lengths x $1/D^{1/2}$, diameter x $(d/D)^{1/2}$

III. RP trees and local covariance dimension

Intrinsic low dimensionality of sets



	Empirically verifiable?	Conducive to analysis?	Summary
Small covering numbers	Kind of	Yes, but too weak in some ways	Small global covers
Small Assouad dimension	Not really	Yes	AND: small local covers
Low-dimensional manifold	No	To some extent	AND: smoothness (local flatness)
Low-dimensional affine subspace	Yes	Yes	AND: global flatness

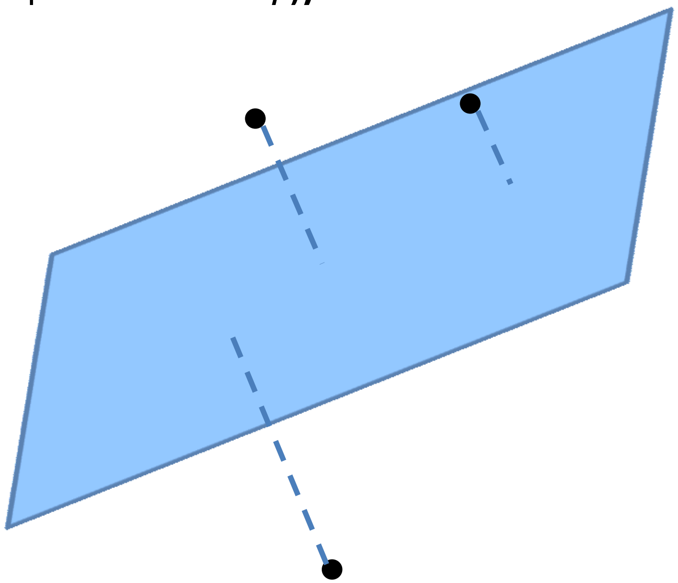
Obvious extension to distributions: at least $1-\delta$ of the probability mass lies within distance ϵ of a set of low intrinsic dimension

Local covariance dimension

A distribution over R^D has covariance dimension (d, ϵ) if its covariance matrix has eigenvalues $\lambda_1 \geq \cdots \geq \lambda_D$ that satisfy:

$$(\lambda_1 + \cdots + \lambda_d) \geq (1-\epsilon)(\lambda_1 + \cdots + \lambda_D).$$

That is, there is a d-dimensional affine subspace such that
(avg dist^2 from subspace)
$\leq \epsilon \cdot$ (avg dist^2 from mean)



We are interested in distributions that *locally* have this property, ie., for some partition of R^D, the restriction of the distribution to each region of the partition has covariance dimension (d,ϵ).

Performance guarantee

Instead of cell diameter, use vector quantization error:

VQ(cell) = avg squared dist from point in cell to mean(cell)

[Using slightly different RP tree construction.]
There are constants c_1, c_2 for which the following holds.

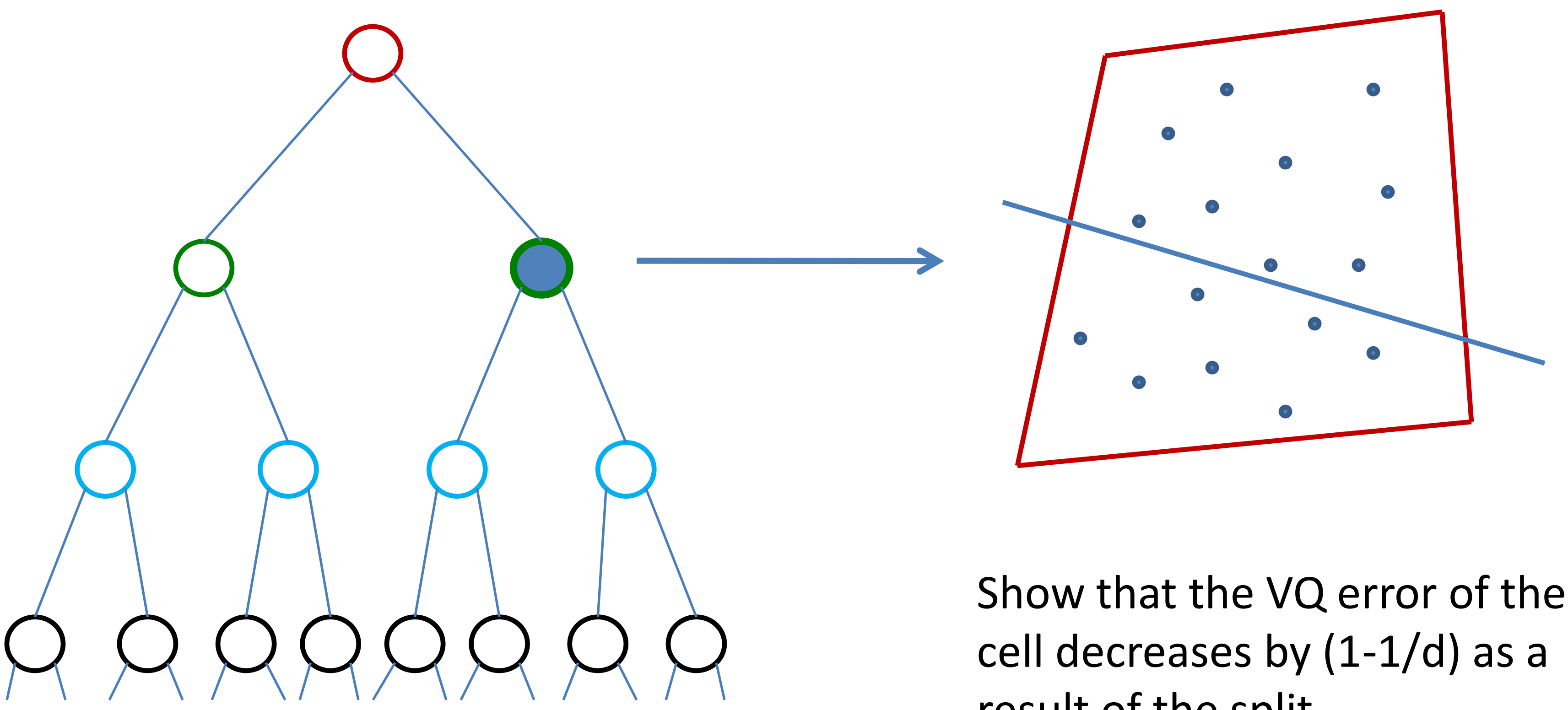
Build an RP tree from data $S \subset R^D$. Suppose a cell C has covariance dimension (d, c_1). Then for each of its children C':

$$\mathbf{E}[VQ(C')] \leq VQ(C)\,(1 - c_2/d)$$

where the expectation is over the split at C.

Proof outline

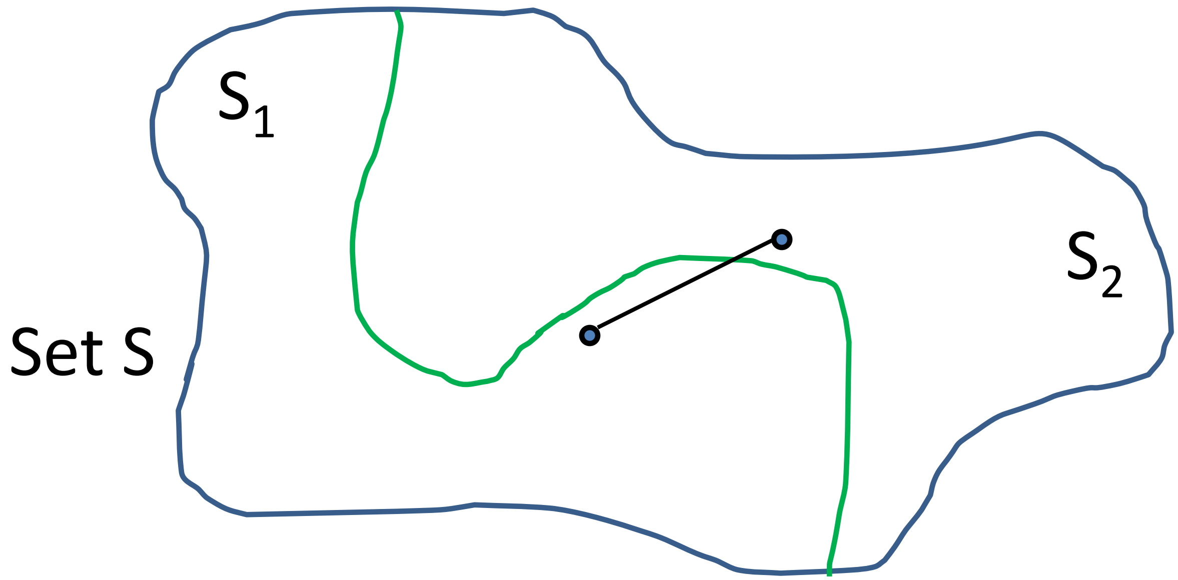
Pick any cell in the RP tree, and let $S \subset R^D$ be the data in it.



Show that the VQ error of the cell decreases by (1-1/d) as a result of the split.

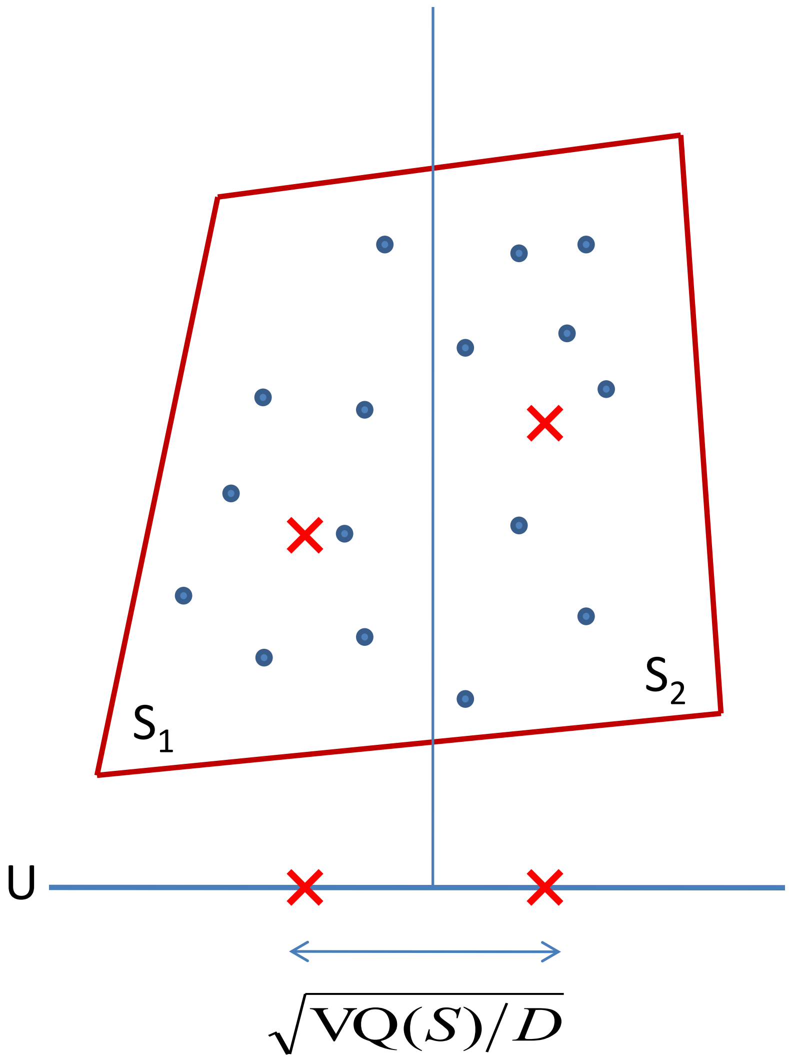
The change in VQ error

If a set S is split into two pieces S_1 and S_2 with equal numbers of points, by how much does its VQ error drop?



By *exactly* $\|\text{mean}(S_1) - \text{mean}(S_2)\|^2$.

Proof outline -- 3



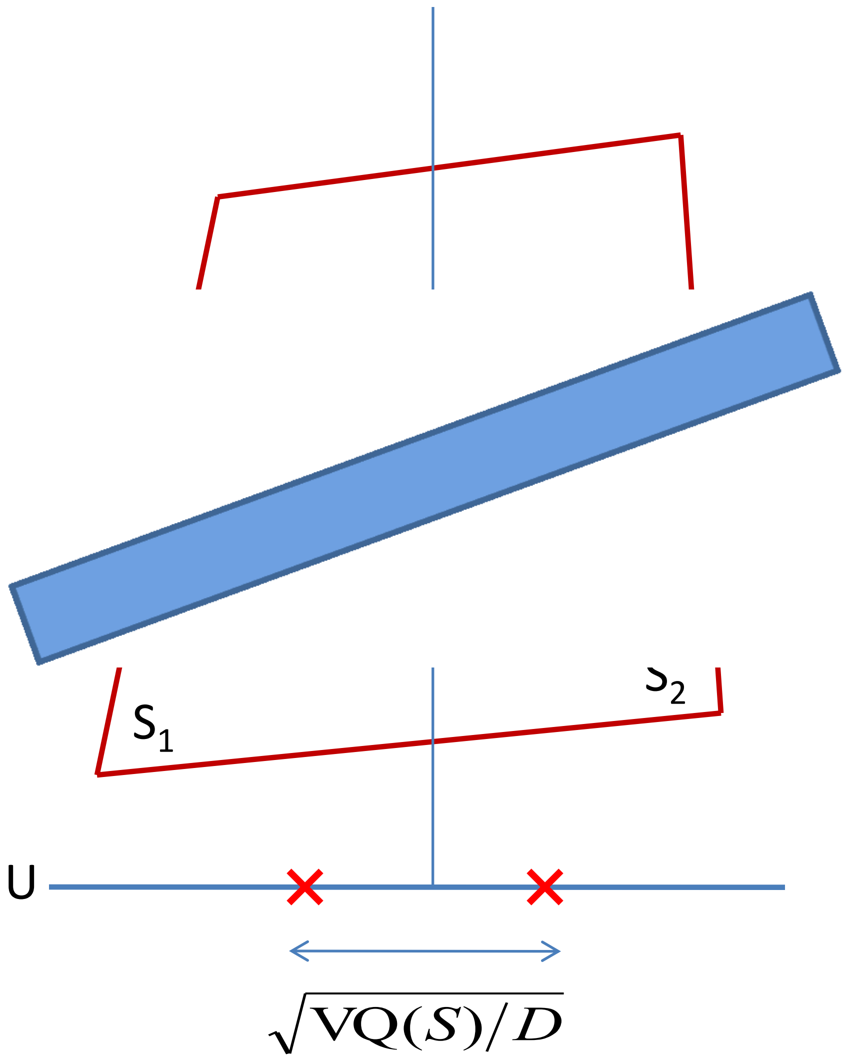
VQ(S)
= average squared distance to mean(S)
= (1/2) average squared interpoint distance
= "variance of S"

Projection onto U shrinks distances by $D^{1/2}$, so shrinks variance by D

Variance of projected S is roughly VQ(S)/D

Distance between projected means is at least $\sqrt{\mathrm{VQ}(S)/D}$

Proof outline -- 4



S is close to a d-dimensional affine subspace; so mean(S_1) and mean(S_2) lie very close to this subspace

The subspace has Assouad dimension O(d), so all vectors in it shrink to $\leq (d/D)^{1/2}$ their original length when projected onto U

Therefore the distance between mean(S_1) and mean(S_2) is at least $\sqrt{\mathrm{VQ}(S)/d}$

IV. Connections and open problems

The uses of k-d trees

1. Classification and regression

   Given data points $(x_1, y_1), \ldots, (x_n, y_n)$, build a tree on the x_i. For any subsequent query x, assign it a y-label that is an average or majority vote of y_i values in cell(x).

2. Near neighbor search

   Build tree on data base $x_1, \ldots, x_n$. Given query x, find an x_i close to it: return nearest neighbor in cell(x).

3. Nearest neighbor search

   Like (2), but may need to look beyond cell(x).

4. Speeding up geometric computations

   For instance, N-body problems in which all interactions between nearby pairs of particles must be computed.

Vector quantization

Setting: lossy data compression.

Data generated from some distribution P over R^D. Pick:

finite codebook $C \subset R^D$

encoding function $\alpha: R^D \to C$

such that $\mathbf{E}\,\|X - \alpha(X)\|^2$ is small.

Tree-based VQ in applications with large |C|.

Typical rate: VQ error $\leq e^{-r/D}$ (r = depth of tree).

RP trees have VQ error $e^{-r/d}$.

Compressed sensing

New model for working with D-dimensional data:

Never look at the original data X!

Work exclusively with a few random projections ϕ(X)

Candes-Tao, *Donoho*: sparse X can be reconstructed from ϕ(X).

Cottage industry of algorithms working exclusively with ϕ(X).

RP trees are compatible with this viewpoint.

Use the same random projection across a level of the tree

Precompute random projections

What next

1. Other tree data structures?

   e.g. nearest neighbor search [such as “cover trees”]

2. Other nonparametric estimators

   e.g. kernel density estimation

3. Other structure (such as clustering) that can be exploited to improve convergence rates of statistical estimators

Thanks

Yoav Freund, my co-author

Nakul Verma and Mayank Kabra, students who ran many experiments

National Science Foundation, for support under grants IIS-0347646 and IIS-0713540