Which Spatial Partition Trees Are Adaptive to Intrinsic Dimension?

Nakul Verma, Samory Kpotufe, and Sanjoy Dasgupta

{naverma, skpotufe, dasgupta}@ucsd.edu

University of California, San Diego

Why is this important?

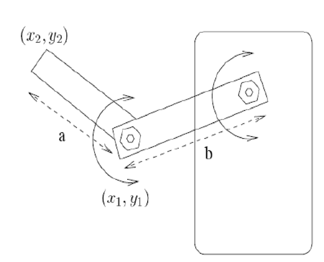
- Spatial trees are at the heart of many machine learning tasks (e.g. regression, near neighbor search, vector quantization).
- However, they tend to suffer from the *curse of dimensionality*: the rate at which the diameter of the data decreases as we go down the tree depends on the dimension of the space. In particular, we might require partitions of size $O(2^D)$ to attain small data diameters.
- Fortunately, many real world data have low intrinsic dimension (e.g. manifolds, sparse datasets), and we would like to benefit from such situations.

We show that

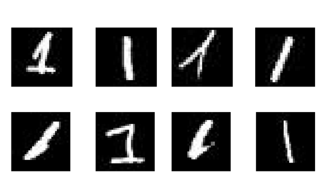
- Trees such as RPTree, PDTree and 2-MeansTree adapt to the intrinsic dimension of the data in terms of the rate at which they decrease diameter
- down the tree.
  This has strong implications on the performance of these trees on the various learning tasks they are used for.

Some real world data with low intrinsic dimension

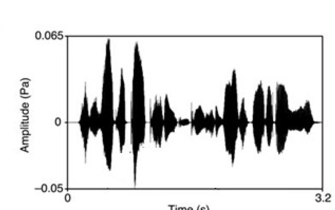
Rotating teapot. One degree of freedom (rotation angle).



Movement of a robotic arm. Two degrees of freedom. – one for each joint.



Handwritten characters. The tilt angle, thickness, etc. govern the final written form.



Speech. Few anatomical characteristics govern the spoken phonemes.

Hand gestures in Sign Language. Few gestures can follow other gestures.

Standard characterizations of intrinsic dimension

Common notions of intrinsic dimension (e.g. Box dimension, Doubling dimension, etc.) originally emerged from fractal geometry. They, however, have the following issues in the context of machine learning:

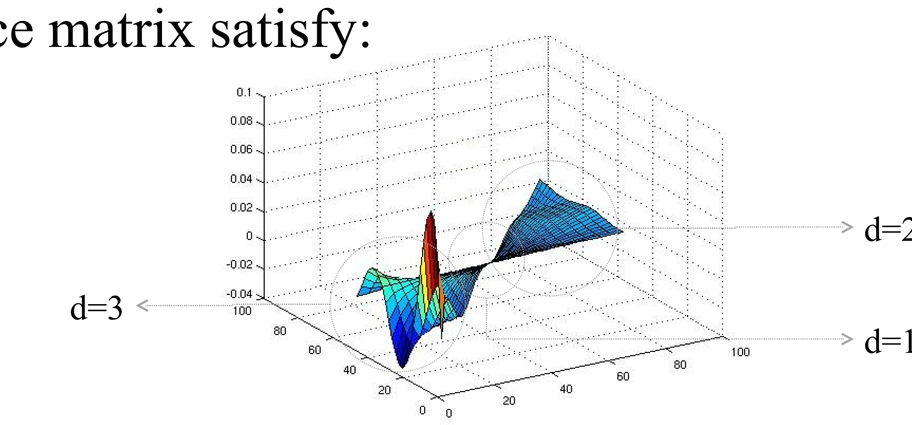
- These notions are purely geometrical and don't account for the underlying distribution.
- They are not robust to distributional noise: e.g. for a noisy manifold, these dimensions can be very high.
- They are difficult to verify empirically.

Need a more statistical notion of intrinsic dimension that characterizes the underlying distribution, is robust to noise, and is easy to verify for real world datasets.

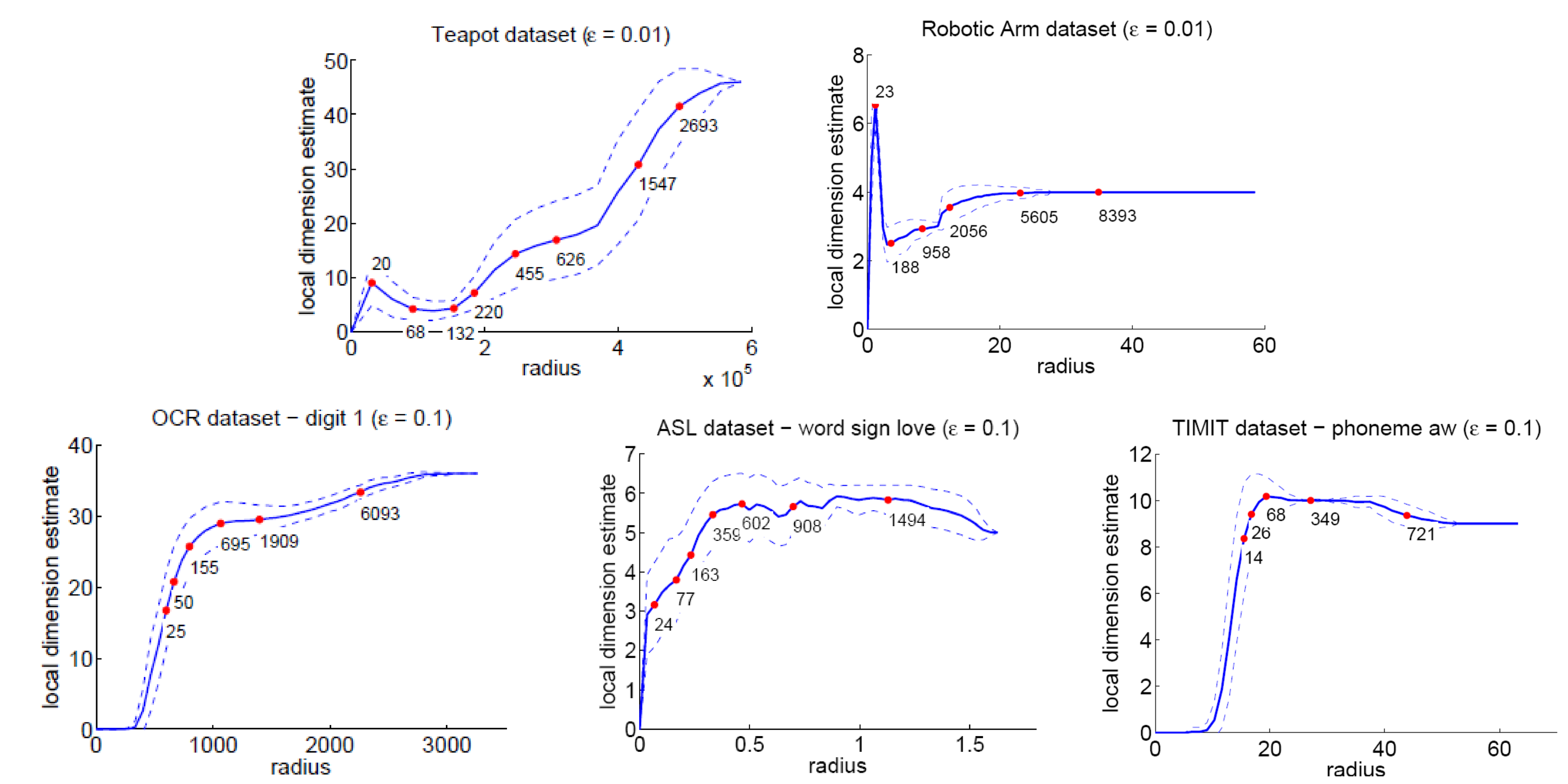
Local covariance dimension

A set $S \subset \Re^D$ is said to have *local covariance dimension (d, ϵ)* if the largest d eigenvalues of its covariance matrix satisfy:

$$\sigma_1^2 + \ldots + \sigma_d^2 \geq (1-\epsilon)\left(\sigma_1^2 + \ldots + \sigma_D^2\right)$$



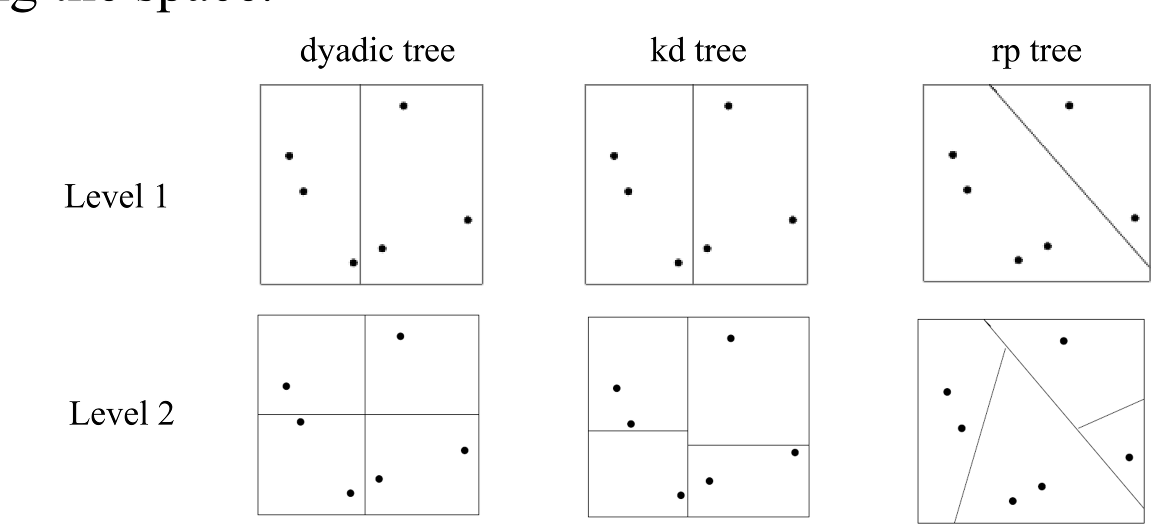
Empirical estimates of local covariance dimension



Loc. cov. dim. estimate at different scales for some real-world datasets.

Spatial partition trees

Builds a hierarchy of nested partitions of the data space by recursively bisecting the space.



The trees we consider:

dyadic tree: Pick a coordinate direction and split the data at the mid point along this direction.

kd tree: Pick a coordinate direction and split the data at the median along this direction.

RP tree: Pick a random direction and split the data at the median along this direction.

PCA/PD tree: Split the data at the median along the principal direction.

2Means tree: Compute the 2-means solution, and split the data as per the cluster assignment.

Theoretical guarantees

Diameter decrease rate (k): Smallest k such that data diameter is halved every k levels.

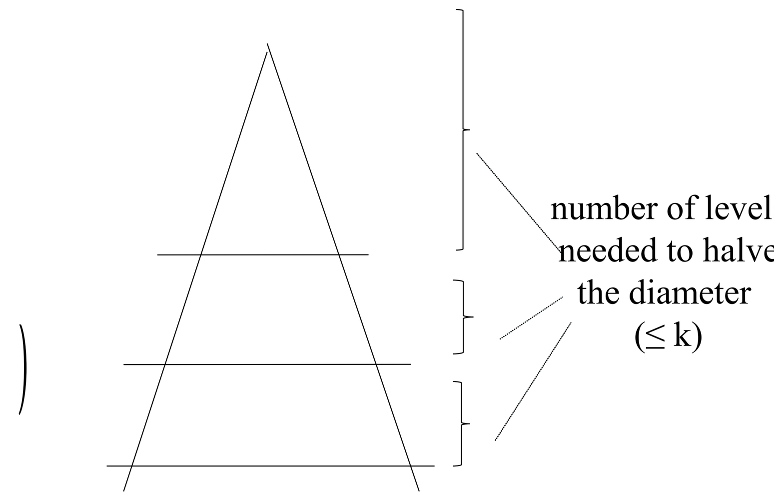
We show that:

RP tree: $k \leq O(d)$

PD tree: $k \leq O\left(\sum \sigma_i^2 / \sigma_1^2\right)^2$

2M tree: $k \leq O\left(\min\left\{d, \left(\sum \sigma_i^2 / \sigma_1^2\right)^2\right\}\right)$

dyadic tree / kd tree: $k \leq O(D \log D)$

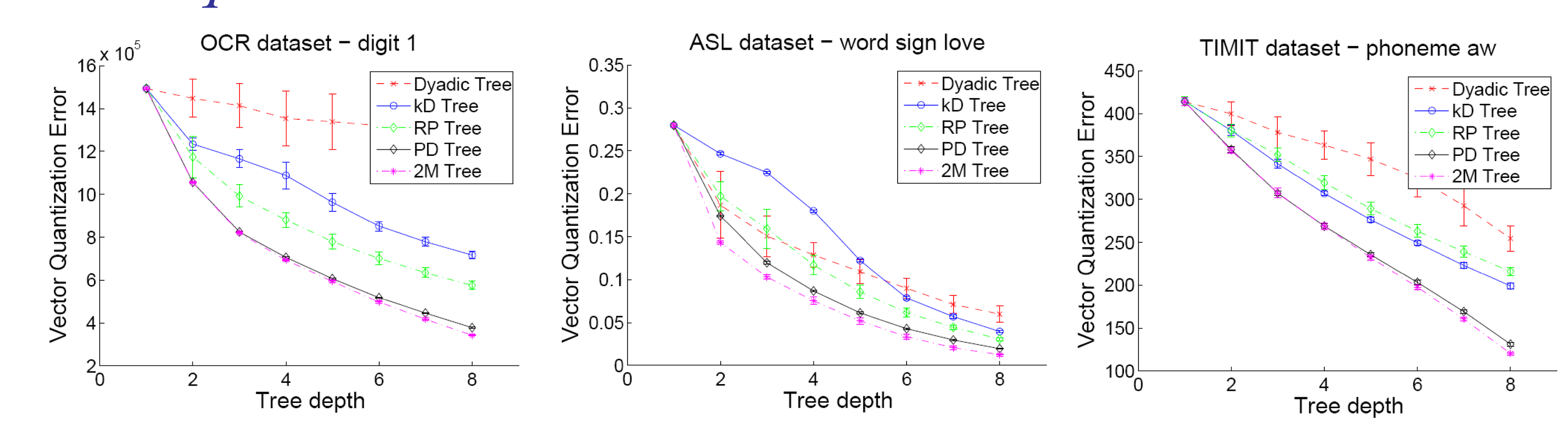


Axis parallel splitting rules (dyadic / kd tree) don't always adapt to intrinsic dimension; the upper bounds have matching lower bounds.

On the other hand, the irregular splitting rules (RP / PD / 2M trees) always adapt to intrinsic dimension. They therefore tend to perform better on real world tasks.

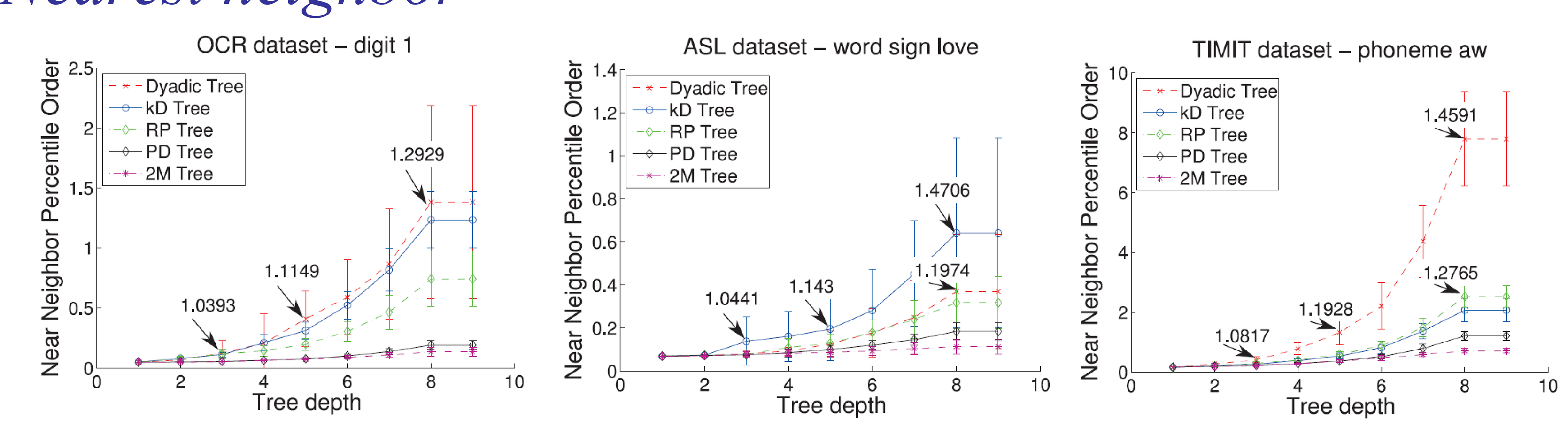
Experiments

Vector quantization



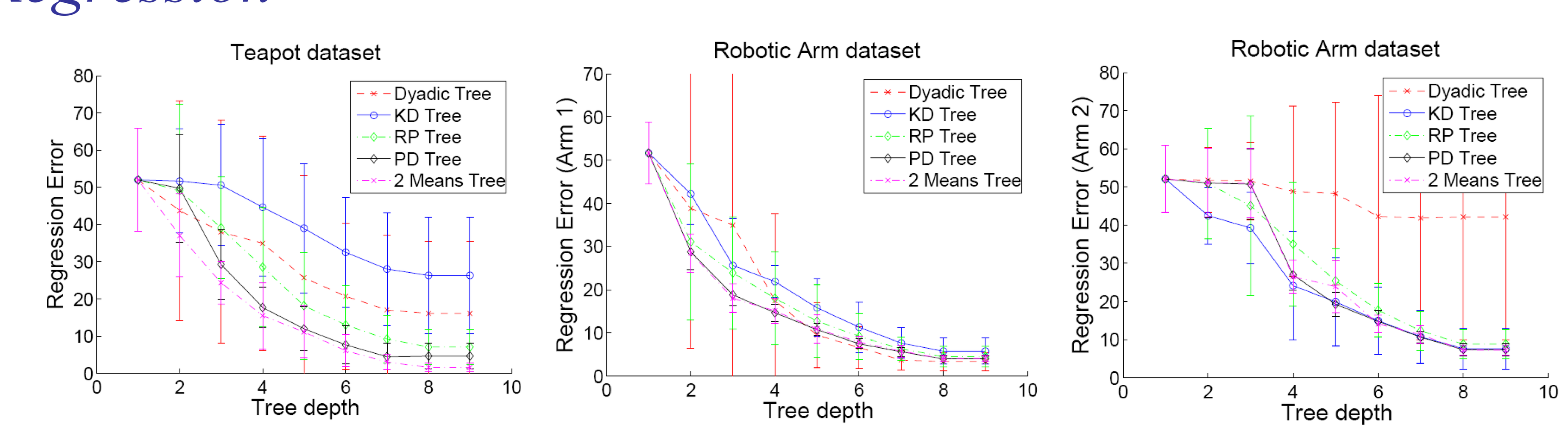
Quantization error of test data at different levels for various partition trees (built using separate training data). 2-means and PD trees perform the best.

Nearest neighbor



Quality of the found neighbor at various levels of the partition trees.

Regression



l_2 regression error in predicting the rotation angle at different tree levels.

All experiments are done with 10-fold cross validation.