Exploiting low intrinsic dimensionality

Sanjoy Dasgupta

University of California, San Diego

The new nonparametrics

Nonparametric methods (kernel density estimation, tree-based regression, etc) can fit any function. But they suffer a severe curse of dimension.

The new nonparametrics

Nonparametric methods (kernel density estimation, tree-based regression, etc) can fit any function. But they suffer a severe curse of dimension.

Consider random pair (X, Y), where $X \in \mathbb{R}^d$ and $Y \in \mathbb{R}$.

- *Regression* problem: infer $f(x) = \mathbb{E}[Y|X = x]$.
- Let f_n be an estimator based on n data points. It is common to judge it by its squared loss

$$\mathbb{E}(f_n(X) - f(X))^2.$$

- Stone 1982: In general, loss $\geq n^{-2p/(2p+d)}$, where p captures the smoothness of f.

The new nonparametrics

Nonparametric methods (kernel density estimation, tree-based regression, etc) can fit any function. But they suffer a severe curse of dimension.

Consider random pair (X, Y), where $X \in \mathbb{R}^d$ and $Y \in \mathbb{R}$.

- *Regression* problem: infer $f(x) = \mathbb{E}[Y|X = x]$.
- Let f_n be an estimator based on n data points. It is common to judge it by its squared loss

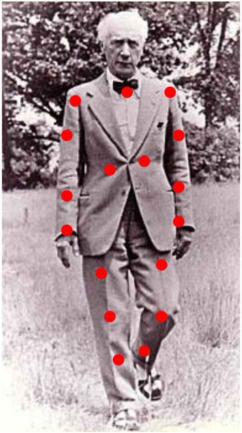
$$\mathbb{E}(f_n(X) - f(X))^2.$$

- Stone 1982: In general, loss $\geq n^{-2p/(2p+d)}$, where p captures the smoothness of f.

What if a high-dimensional data source actually has relatively few "degrees of freedom"?

Low dimensional manifolds

Sometimes data in a high-dimensional space $\mathbb{R}^d$ in fact lies close to a d_o-dimensional manifold, for $d_o \ll d$



1. Motion capture
   M markers on a human body yields data in $\mathbb{R}^{3M}$
2. Speech signals
   Representation can be made arbitrarily high dimensional by applying more filters to each window of the time series

This whole area: "Manifold learning"

Another example of low intrinsic dimension

Bag-of-words document model

- Fix a vocabulary of size, say, d
- A document is represented by a d-dimensional vector indicating, for each word, whether it appears (or how often)

Average number of nonzero entries in these vectors is $d_o \ll d$.

Another example of low intrinsic dimension

Bag-of-words document model

- Fix a vocabulary of size, say, d
- A document is represented by a d-dimensional vector indicating, for each word, whether it appears (or how often)

Average number of nonzero entries in these vectors is $d_o \ll d$.

There are several different and widely-occurring types of low intrinsic dimension. Can we:

- Find a broad notion of dimensionality that captures at least a few of these?
- Develop nonparametric estimators whose rates of convergence depend only on this refined notion rather than on the superficial ambient dimension?

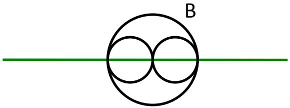
Doubling dimension

Set $S \subset \mathbb{R}^d$ has doubling dimension d_o if for any (Euclidean) ball B, the subset $S \cap B$ can be covered by 2^{d_o} balls of half the radius.

Doubling dimension

Set $S \subset \mathbb{R}^d$ has doubling dimension d_o if for any (Euclidean) ball B, the subset $S \cap B$ can be covered by 2^{d_o} balls of half the radius.

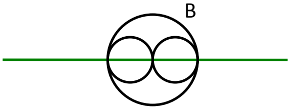
1. Example: $S =$ line has doubling dimension 1.



Doubling dimension

Set $S \subset \mathbb{R}^d$ has doubling dimension d_o if for any (Euclidean) ball B, the subset $S \cap B$ can be covered by 2^{d_o} balls of half the radius.

1. Example: $S =$ line has doubling dimension 1.

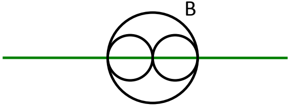


2. A k-dimensional flat has doubling dimension $c_o k$ for some absolute constant c_o.

Doubling dimension

Set $S \subset \mathbb{R}^d$ has doubling dimension d_o if for any (Euclidean) ball B, the subset $S \cap B$ can be covered by 2^{d_o} balls of half the radius.

1. Example: $S =$ line has doubling dimension 1.

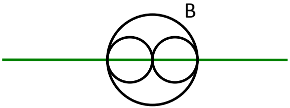


2. A k-dimensional flat has doubling dimension $c_o k$ for some absolute constant c_o.
3. If S has diameter Δ and doubling dimension d_o, then for any $\epsilon > 0$, it has an ϵ-cover of size $\leq (2\Delta/\epsilon)^{d_o}$.

Doubling dimension

Set $S \subset \mathbb{R}^d$ has doubling dimension d_o if for any (Euclidean) ball B, the subset $S \cap B$ can be covered by 2^{d_o} balls of half the radius.

1. Example: $S =$ line has doubling dimension 1.



2. A k-dimensional flat has doubling dimension $c_o k$ for some absolute constant c_o.
3. If S has diameter Δ and doubling dimension d_o, then for any $\epsilon > 0$, it has an ϵ-cover of size $\leq (2\Delta/\epsilon)^{d_o}$.
4. If S has doubling dimension d_o, then so does any subset of S.

The doubling dimension of sparse sets

Set $S \subset \mathbb{R}^d$ has doubling dimension d_o if for any (Euclidean) ball B, the subset $S \cap B$ can be covered by 2^{d_o} balls of half the radius.

1. A set of n points has doubling dimension at most $\log n$.
   Proof: It can be covered by n balls of any radius.

The doubling dimension of sparse sets

Set $S \subset \mathbb{R}^d$ has doubling dimension d_o if for any (Euclidean) ball B, the subset $S \cap B$ can be covered by 2^{d_o} balls of half the radius.

1. A set of n points has doubling dimension at most $\log n$.
   Proof: It can be covered by n balls of any radius.
2. If sets $S_1, \ldots, S_m$ each have doubling dimension $\leq d_o$, then $S_1 \cup \cdots \cup S_m$ has doubling dimension $\leq d_o + \log m$.
   Proof: $S_i \cap B$ can be covered by 2^{d_o} balls of half the radius. Therefore, at most $m2^{d_o}$ balls are needed for the union.

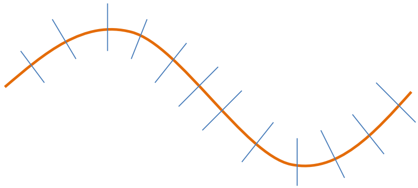
The doubling dimension of sparse sets

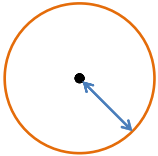
Set $S \subset \mathbb{R}^d$ has doubling dimension d_o if for any (Euclidean) ball B, the subset $S \cap B$ can be covered by 2^{d_o} balls of half the radius.

1. A set of n points has doubling dimension at most $\log n$.
   Proof: It can be covered by n balls of any radius.
2. If sets $S_1, \ldots, S_m$ each have doubling dimension $\leq d_o$, then $S_1 \cup \cdots \cup S_m$ has doubling dimension $\leq d_o + \log m$.
   Proof: $S_i \cap B$ can be covered by 2^{d_o} balls of half the radius. Therefore, at most $m2^{d_o}$ balls are needed for the union.
3. Suppose each point in $S \subset \mathbb{R}^d$ has $\leq k$ nonzero coordinates. Then S has doubling dimension $\leq c_o k + k \log d$.
   Proof: S is the union of $\binom{d}{k}$ flats of dimension k; we've seen that each flat has doubling dimension $\leq c_o k$.

The doubling dimension of manifolds

A Riemannian submanifold $M \subset \mathbb{R}^d$ has *condition number* $\leq 1/\tau$ if normals to M of length τ don't intersect:





If $M \subset \mathbb{R}^d$ is a k-dimensional manifold of condition number $1/\tau$, then its neighborhoods of radius τ have doubling dimension $O(k)$.

Exploiting low intrinsic dimension

Suppose we have data (X, Y), where the distribution of X is supported on a set $\mathcal{X} \subset \mathbb{R}^d$ of intrinsic dimension d_o.

Exploiting low intrinsic dimension

Suppose we have data (X, Y), where the distribution of X is supported on a set $\mathcal{X} \subset \mathbb{R}^d$ of intrinsic dimension d_o.

Some possibilities:

1. Find an embedding $\Phi : \mathbb{R}^d \to \mathbb{R}^k$ such that:
    - Φ is $1 - 1$ on $\mathcal{X}$,
    - k is much smaller than d, ideally $k = O(d_o)$, and
    - Φ preserves the neighborhood structure of $\mathcal{X}$ in some suitable sense.

Exploiting low intrinsic dimension

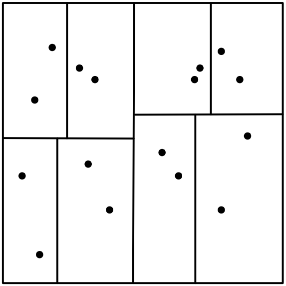
Suppose we have data (X, Y), where the distribution of X is supported on a set $\mathcal{X} \subset \mathbb{R}^d$ of intrinsic dimension d_o.

Some possibilities:

1. Find an embedding $\Phi : \mathbb{R}^d \to \mathbb{R}^k$ such that:
   - Φ is $1-1$ on $\mathcal{X}$,
   - k is much smaller than d, ideally $k = O(d_o)$, and
   - Φ preserves the neighborhood structure of $\mathcal{X}$ in some suitable sense.
2. Find a simpler representation of $\mathcal{X}$ that is easy to construct and provably adapts to the intrinsic dimension. Obvious candidate: tree-based spatial partition.

Spatial partitioning for nonparametric regression

e.g. the k-d tree:



To split a cell with points S:

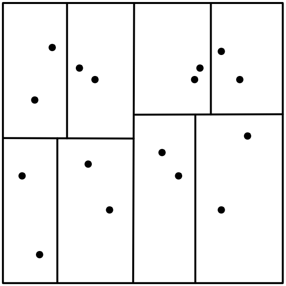
- Choose a coordinate direction
- Split at the median along that direction

Once the tree is built:

- Fit a simple model (e.g. constant) in each leaf.
- Answer a query by routing it to a leaf and applying the leaf's model.

Spatial partitioning for nonparametric regression

e.g. the k-d tree:



To split a cell with points S:

- Choose a coordinate direction
- Split at the median along that direction

Once the tree is built:

- Fit a simple model (e.g. constant) in each leaf.
- Answer a query by routing it to a leaf and applying the leaf's model.

These regressors are consistent if, as $n \to \infty$,

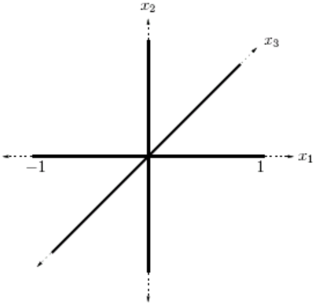
1. the diameter of the leaf cells goes to zero, and
2. the number of samples in each leaf goes to infinity.

Rate of convergence depends on relative speed of these two effects.

k-d trees are not adaptive to intrinsic dimension

As one moves down a k-d tree, how rapidly does the cell diameter shrink?

Consider the data set $S = \cup_{i=1}^{d} \{te_i : -1 \leq t \leq 1\}$.

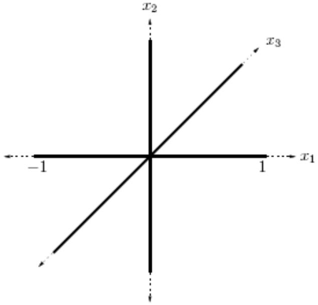


At least d levels are needed to halve the diameter.

k-d trees are not adaptive to intrinsic dimension

As one moves down a k-d tree, how rapidly does the cell diameter shrink?

Consider the data set $S = \cup_{i=1}^{d}\{te_i : -1 \leq t \leq 1\}$.

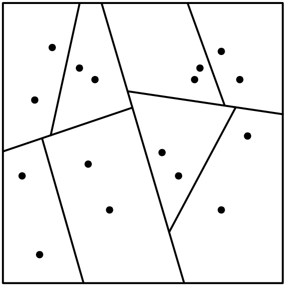


At least d levels are needed to halve the diameter.

Yet S has doubling dimension just $d_o = 1 + \log d$.

Random projection trees

A randomized variant of the k-d tree:

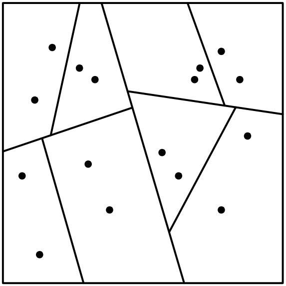


To split a cell with points $S \subset \mathbb{R}^d$:

- Choose a direction v at random from the unit sphere S^{d-1}
- Split at the median along that direction, perturbed slightly:
  - Pick any $x \in S$, and let $y \in S$ be the point farthest from it
  - Choose δ uniformly at random from $[-1, 1] \cdot 6\|x - y\|/\sqrt{d}$
  - Split at median$(\{z \cdot v : z \in S\}) + \delta$

Random projection trees

A randomized variant of the k-d tree:



To split a cell with points $S \subset \mathbb{R}^d$:

- Choose a direction v at random from the unit sphere S^{d-1}
- Split at the median along that direction, perturbed slightly:
  - Pick any $x \in S$, and let $y \in S$ be the point farthest from it
  - Choose δ uniformly at random from $[-1, 1] \cdot 6\|x - y\|/\sqrt{d}$
  - Split at median$(\{z \cdot v : z \in S\}) + \delta$

Theorem: There is a constant c_1 with the following property. Suppose an RP tree is built using data set $S \subset \mathbb{R}^d$. Pick any cell C in the RP tree; suppose that $S \cap C$ has doubling dimension $\leq d_o$. Then with probability at least $1/2$ (over the randomization in constructing the subtree rooted at C), for every descendant C' which is more than $c_1 d_o \log d_o$ levels below C, we have radius$(C') \leq$ radius$(C)/2$.

Properties of random projection

We choose random projections from $\mathbb{R}^d$ to $\mathbb{R}$ as follows:

- Pick U from the multivariate Gaussian $N(0, (1/d)I_d)$.
- Define projection $\Pi(x) = U \cdot x$.

Properties of random projection

We choose random projections from $\mathbb{R}^d$ to $\mathbb{R}$ as follows:

- Pick U from the multivariate Gaussian $N(0, (1/d)I_d)$.
- Define projection $\Pi(x) = U \cdot x$.

This shrinks an individual vector x by roughly $\sqrt{d}$.

Lemma: For any $\alpha, \beta > 0$:

(a) $\Pr\left[|\Pi(x)| \leq \alpha \cdot \frac{\|x\|}{\sqrt{d}}\right] \;\leq\; \sqrt{\frac{2}{\pi}}\,\alpha$; and

(b) $\Pr\left[|\Pi(x)| \geq \beta \cdot \frac{\|x\|}{\sqrt{d}}\right] \;\leq\; \frac{2}{\beta} e^{-\beta^2/2}$.

Properties of random projection

We choose random projections from $\mathbb{R}^d$ to $\mathbb{R}$ as follows:

- Pick U from the multivariate Gaussian $N(0, (1/d)I_d)$.
- Define projection $\Pi(x) = U \cdot x$.

This shrinks an individual vector x by roughly $\sqrt{d}$.

Lemma: For any $\alpha, \beta > 0$:

(a) $\Pr\left[|\Pi(x)| \leq \alpha \cdot \frac{\|x\|}{\sqrt{d}}\right] \leq \sqrt{\frac{2}{\pi}}\,\alpha$; and

(b) $\Pr\left[|\Pi(x)| \geq \beta \cdot \frac{\|x\|}{\sqrt{d}}\right] \leq \frac{2}{\beta} e^{-\beta^2/2}$.

Proof: $\Pi(x)$ also has a Gaussian distribution, $N(0, \|x\|^2/d)$.

Properties of random projection

We choose random projections from $\mathbb{R}^d$ to $\mathbb{R}$ as follows:

- Pick U from the multivariate Gaussian $N(0, (1/d)I_d)$.
- Define projection $\Pi(x) = U \cdot x$.

This shrinks an individual vector x by roughly $\sqrt{d}$.

Lemma: For any $\alpha, \beta > 0$:

(a) $\Pr\left[|\Pi(x)| \leq \alpha \cdot \frac{\|x\|}{\sqrt{d}}\right] \;\leq\; \sqrt{\frac{2}{\pi}}\,\alpha$; and

(b) $\Pr\left[|\Pi(x)| \geq \beta \cdot \frac{\|x\|}{\sqrt{d}}\right] \;\leq\; \frac{2}{\beta} e^{-\beta^2/2}$.

Proof: $\Pi(x)$ also has a Gaussian distribution, $N(0, \|x\|^2/d)$.

Corollary: Suppose $S \subset B(x_o, \Delta)$. With probability $> 1 - \delta$,

$$|\text{median}(\Pi(S)) - \Pi(x_o)| \;\leq\; \frac{\Delta}{\sqrt{d}} \cdot \sqrt{2 \ln \frac{2}{\delta}}.$$

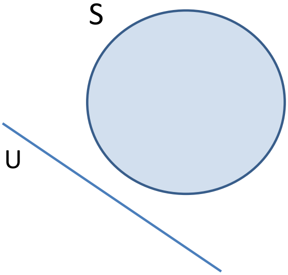
Random projection and diameter

For $S \subset \mathbb{R}^d$, how does the diameter of $\Pi(S)$ compare to that of S?

Random projection and diameter

For $S \subset \mathbb{R}^d$, how does the diameter of $\Pi(S)$ compare to that of S?

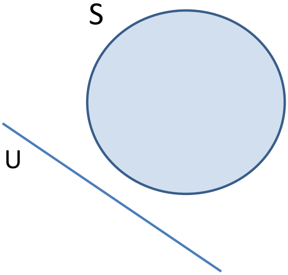
If S is full-dimensional, the diameter could be unchanged.



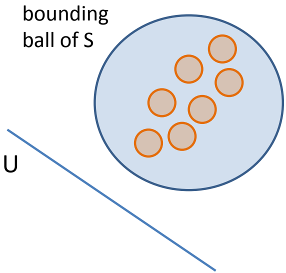
Random projection and diameter

For $S \subset \mathbb{R}^d$, how does the diameter of $\Pi(S)$ compare to that of S?

If S is full-dimensional, the diameter could be unchanged.



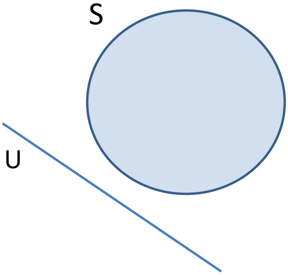
But if S has doubling dimension $d_o \ll d$, the diameter ought to shrink.



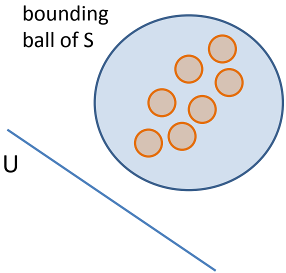
Random projection and diameter

For $S \subset \mathbb{R}^d$, how does the diameter of $\Pi(S)$ compare to that of S?

If S is full-dimensional, the diameter could be unchanged.



But if S has doubling dimension $d_o \ll d$, the diameter ought to shrink.



In the latter case, $\text{diam}(\Pi(S)) \;\approx\; \text{diam}(S) \cdot \sqrt{d_o/d}$.

Random projection and diameter

Theorem: If $S \subset \mathbb{R}^d$ has doubling dimension d_o, then with probability at least $1 - \delta$, the diameter of $\Pi(S)$ is at most

$$4 \cdot \frac{\text{diam}(S)}{\sqrt{d}} \cdot \sqrt{2\left(d_o + \ln\frac{2}{\delta}\right)}.$$

Random projection and diameter

Theorem: If $S \subset \mathbb{R}^d$ has doubling dimension d_o, then with probability at least $1-\delta$, the diameter of $\Pi(S)$ is at most

$$4 \cdot \frac{\mathrm{diam}(S)}{\sqrt{d}} \cdot \sqrt{2\left(d_o + \ln\frac{2}{\delta}\right)}.$$

Proof: We'll prove a weaker version with factor $\sqrt{(d_o \log d)/d}$.

Random projection and diameter

Theorem: If $S \subset \mathbb{R}^d$ has doubling dimension d_o, then with probability at least $1-\delta$, the diameter of $\Pi(S)$ is at most

$$4 \cdot \frac{\text{diam}(S)}{\sqrt{d}} \cdot \sqrt{2\left(d_o + \ln \frac{2}{\delta}\right)}.$$

Proof: We'll prove a weaker version with factor $\sqrt{(d_o \log d)/d}$.

1. WLOG S has diameter 1 and $S \subset B(0,1)$.

Random projection and diameter

Theorem: If $S \subset \mathbb{R}^d$ has doubling dimension d_o, then with probability at least $1-\delta$, the diameter of $\Pi(S)$ is at most

$$4 \cdot \frac{\text{diam}(S)}{\sqrt{d}} \cdot \sqrt{2\left(d_o + \ln \frac{2}{\delta}\right)}.$$

Proof: We'll prove a weaker version with factor $\sqrt{(d_o \log d)/d}$.

1. WLOG S has diameter 1 and $S \subset B(0,1)$.
2. Cover S by balls of radius $\sqrt{d_o/d}$. At most $(d/d_o)^{d_o/2}$ balls are needed.

Random projection and diameter

Theorem: If $S \subset \mathbb{R}^d$ has doubling dimension d_o, then with probability at least $1-\delta$, the diameter of $\Pi(S)$ is at most

$$4 \cdot \frac{\text{diam}(S)}{\sqrt{d}} \cdot \sqrt{2\left(d_o + \ln \frac{2}{\delta}\right)}.$$

Proof: We'll prove a weaker version with factor $\sqrt{(d_o \log d)/d}$.

1. WLOG S has diameter 1 and $S \subset B(0,1)$.
2. Cover S by balls of radius $\sqrt{d_o/d}$. At most $(d/d_o)^{d_o/2}$ balls are needed.
3. Pick any of these balls. With probability $1-(1/d)^{d_o}$, its center is projected to a point within distance $\sqrt{(d_o \log d)/d}$ of the origin; and thus the entire projected ball lies in an interval within distance $\sqrt{(d_o \log d)/d} + \sqrt{d_o/d}$ of the origin.

Random projection and diameter

Theorem: If $S \subset \mathbb{R}^d$ has doubling dimension d_o, then with probability at least $1-\delta$, the diameter of $\Pi(S)$ is at most

$$4 \cdot \frac{\text{diam}(S)}{\sqrt{d}} \cdot \sqrt{2\left(d_o + \ln\frac{2}{\delta}\right)}.$$

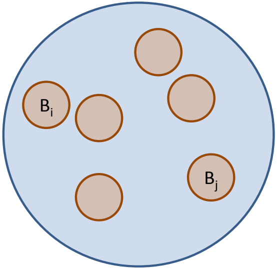
Proof: We'll prove a weaker version with factor $\sqrt{(d_o \log d)/d}$.

1. WLOG S has diameter 1 and $S \subset B(0,1)$.
2. Cover S by balls of radius $\sqrt{d_o/d}$. At most $(d/d_o)^{d_o/2}$ balls are needed.
3. Pick any of these balls. With probability $1-(1/d)^{d_o}$, its center is projected to a point within distance $\sqrt{(d_o \log d)/d}$ of the origin; and thus the entire projected ball lies in an interval within distance $\sqrt{(d_o \log d)/d} + \sqrt{d_o/d}$ of the origin.
4. Take a union bound over all the balls.

Proof outline for RP trees

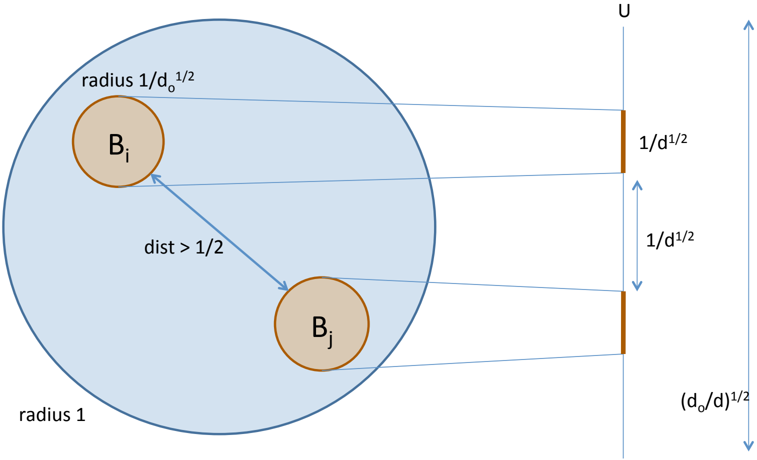
Suppose $S \subset \mathbb{R}^d$ has doubling dimension d_o and lies in a ball of radius 1. We need to show that if an RP tree is built on S, then with constant probability, every cell $O(d_o \log d_o)$ levels below is contained in ball of radius $1/2$.

Current cell (radius ≤ 1):



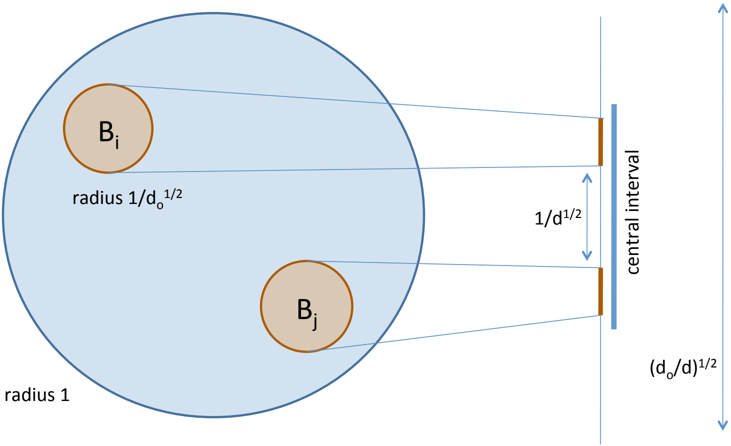
1. Cover S by $d_o^{d_o/2}$ balls B_i of radius $1/\sqrt{d_o}$.
2. Consider any pair of balls B_i, B_j that are distance $> 1/2 - 1/\sqrt{d_o}$ apart. We'll see that a single random split has constant probability of cleanly separating them.
3. There are at most $d_o^{d_o}$ such pairs, so after $O(d_o \log d_o)$ splits, with constant probability every faraway pair of balls will be separated. Thus all cells at that level will have radius $\leq 1/2$.

The big picture



Recall that random projection shrinks diameter by $\sqrt{d_o/d}$ and individual vectors by $1/\sqrt{d}$.

The big picture



Most projected points (and the median) fall in a central interval of size $1/\sqrt{d}$.

Regression in spaces of low intrinsic dimension

Goal: regression with rates depending on d_o rather than d.

Regression in spaces of low intrinsic dimension

Goal: regression with rates depending on d_o rather than d.

Given data $(X_1, Y_1), \ldots, (X_n, Y_n) \in \mathbb{R}^d \times \mathbb{R}$, here's a typical tree-based regressor:

1. The data is used to construct a partition $\mathbb{C}$ of the underlying space.
2. A simple model is fit to each cell of $\mathbb{C}$.

For instance, piecewise-constant regressor:

$$f_{n,\mathbb{C}}(x) = \frac{\sum_{i=1}^{n} Y_i \cdot 1(X_i \in \mathbb{C}(x))}{\sum_{i=1}^{n} 1(X_i \in \mathbb{C}(x))},$$

where $\mathbb{C}(x)$ is the cell of $\mathbb{C}$ to which x belongs.

Regression in spaces of low intrinsic dimension

Goal: regression with rates depending on d_o rather than d.

Given data $(X_1, Y_1), \ldots, (X_n, Y_n) \in \mathbb{R}^d \times \mathbb{R}$, here's a typical tree-based regressor:

1. The data is used to construct a partition $\mathbb{C}$ of the underlying space.
2. A simple model is fit to each cell of $\mathbb{C}$.

For instance, piecewise-constant regressor:

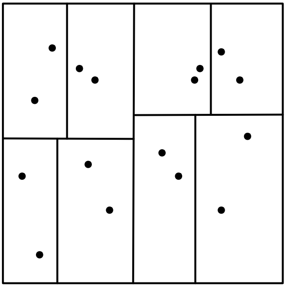
$$f_{n,\mathbb{C}}(x) = \frac{\sum_{i=1}^{n} Y_i \cdot 1(X_i \in \mathbb{C}(x))}{\sum_{i=1}^{n} 1(X_i \in \mathbb{C}(x))},$$

where $\mathbb{C}(x)$ is the cell of $\mathbb{C}$ to which x belongs.

We'll use the leaf-cells of an RP tree as the partition $\mathbb{C}$.

RP-tree based regression: analysis

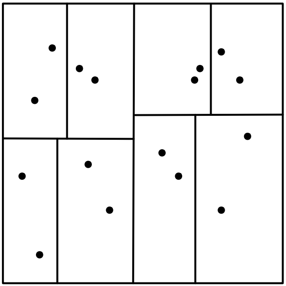
Standard analysis of a tree-based regressor, assuming the regression function is Lipschitz:



1. Bound the bias.
   This is proportional to the physical diameter of the cells of partition $\mathbb{C}$.
2. Bound the variance.
   Relate the empirical Y-mean within each cell to the true Y-mean, and relate the empirical probability mass of each cell to its true mass.

RP-tree based regression: analysis

Standard analysis of a tree-based regressor, assuming the regression function is Lipschitz:



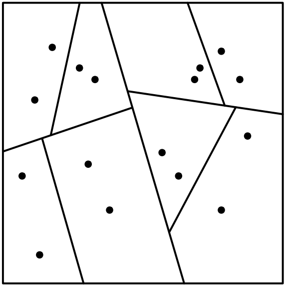
1. Bound the bias.
   This is proportional to the physical diameter of the cells of partition $\mathbb{C}$.
2. Bound the variance.
   Relate the empirical Y-mean within each cell to the true Y-mean, and relate the empirical probability mass of each cell to its true mass.

Both arguments fail in the context of RP trees.

Bounding cell diameters

The cells of an RP tree are convex but otherwise irregular. It is hard to bound their *physical* diameter

$$\Delta(C) = \max_{x,y \in C} \|x - y\|.$$



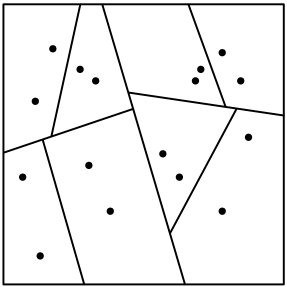
Bounding cell diameters

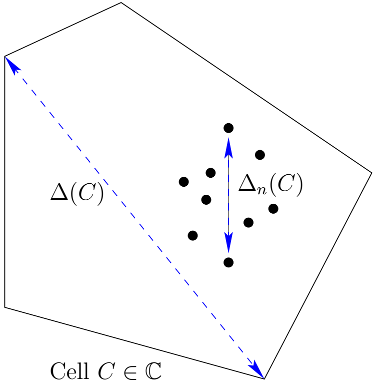
The cells of an RP tree are convex but otherwise irregular. It is hard to bound their *physical* diameter

$$\Delta(C) = \max_{x,y \in C} \|x - y\|.$$

But the RP tree results do give us a bound on their *data diameter*

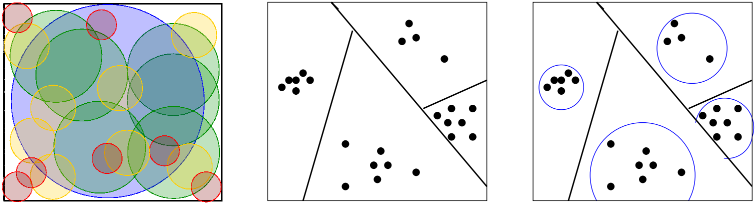
$$\Delta_n(C) = \max_{X_i, X_j \in C} \|X_i - X_j\|.$$





The two notions of diameter

Although the algorithm is forced to work with irregular partition $\mathbb{C}$, we define an alternate partition $\mathbb{C}'$ that is used in the analysis.



(a) Cover $\mathcal{B}$ (b) Partition $\mathbb{C}$ (c) Partition $\mathbb{C}'$

Each cell of $C \in \mathbb{C}$ corresponds to two cells of $C_1, C_2 \in \mathbb{C}'$:

- C_1 has physical diameter approximately equal to its data diameter.
- C_2 contains no data points.

An RP-tree based regressor

$\mathbb{C}_0 = \mathbb{R}^d$
Define $\alpha(n) = (\log^2 n) \log\log(n/\delta) + \log(1/\delta)$
For $i = 1, 2, \ldots$:
  For each cell $C \in \mathbb{C}_{i-1}$:
    Set the subtree rooted at C to `coreRPtree`$(C \cap S)$
  Let $\mathbb{C}_i$ be the partition of $\mathbb{R}^d$ defined by the leaves of the current tree
  If $\Delta_n^2(\mathbb{C}_i) \leq \Delta_n^2(\mathbb{C}_0) \cdot (\alpha(n)/n) \cdot 2^{\mathrm{depth}(\mathbb{C}_i)}$:
    Let $\mathbb{C}^*$ be either $\mathbb{C}_{i-1}$ or $\mathbb{C}_i$, whichever has smaller $\left(\frac{\alpha(n)}{n} \cdot |\mathbb{C}| + \Delta_n^2(\mathbb{C})\right)$
    Return $f_{n,\mathbb{C}^*}$

The `coreRPtree` subroutine takes as input a cell C and returns a subtree whose root corresponds to C and whose leaves have average data diameter half that of C.

Final risk bound

There are absolute constants C, c_o for the which the following holds. Suppose

- the regression function f is λ-Lipschitz and
- the instance space has doubling dimension d_o and diameter Δ.

Then with probability at least $1 - \delta$, the estimator f_n has loss

$$\|f_n - f\|^2 \;\leq\; C\lambda^2\Delta^2 \left(\frac{\log^2 n + \log 1/\delta}{n}\right)^{2/2+k},$$

where $k = c_o d_o \log d_o$.

Final risk bound

There are absolute constants C, c_o for the which the following holds. Suppose

- the regression function f is λ-Lipschitz and
- the instance space has doubling dimension d_o and diameter Δ.

Then with probability at least $1 - \delta$, the estimator f_n has loss

$$\|f_n - f\|^2 \;\leq\; C\lambda^2\Delta^2 \left(\frac{\log^2 n + \log 1/\delta}{n}\right)^{2/2+k},$$

where $k = c_o d_o \log d_o$.

We have gone from $k = d$ down to $k = O(d_o \log d_o)$.
Can we go down further, to $k = d_o$?

Open problems

1. Working in general metric spaces.
   Doubling dimension can be defined for any metric space, but an RP tree is confined to Euclidean space. What are good spatial partition trees for metric spaces and what kinds of adaptivity do they exhibit?

Open problems

1. Working in general metric spaces.
   Doubling dimension can be defined for any metric space, but an RP tree is confined to Euclidean space. What are good spatial partition trees for metric spaces and what kinds of adaptivity do they exhibit?
2. More general notions of intrinsic dimension.
   Can we get closer to the underlying "degrees of freedom" of the input space?

Thanks

To my co-authors Yoav Freund and Samory Kpotufe, and to the National Science Foundation.