ANN for ℓ_∞

Want: (r, cr) - ANN for $(\mathbb{R}^d, \|.\|_\infty)$.
LSH does not exist for $c \ll \sqrt{d}$.
Will use <u>data-dependent</u> partitions.
Data structure is <u>deterministic</u>.

<u>Th</u> (r, cr) - ANN for ℓ_∞ can be solved:

1) In space: $O(dn^{1+\varepsilon})$
2) Query time: $O(d \log n)$
3) $c = O\left(\frac{\log \log d}{\varepsilon}\right)$

Remarks:

- WLOG $r = 1$. Assume from now on.
- Doubly-logarithmic dependence is <u>tight</u>.

<u>Lm</u> $X \subset \mathbb{R}^d$ $|X| = n$. $\varepsilon > 0$.
Either: $\exists$ a ball (in ℓ_∞; aka cube) of radius $O\left(\frac{\log \log d}{\varepsilon}\right)$ that contains $\geq \frac{n}{2}$ points
Or: $\exists\, i \in [d]$ $\tau \in \mathbb{R}$ ("coordinate cut")
s.t. if $A = \{x \in X \mid x_i < \tau - 1\}$ $B = \{x \in X \mid \tau - 1 \leq x_i \leq \tau + 1\}$ $C = \{x \in X \mid x_i > \tau + 1\}$, then

(2)

- $|A|, |C| \geq \Omega\left(\frac{n}{d}\right)$
- $\left(\frac{|A|+|B|}{n}\right)^{1+\varepsilon} + \left(\frac{|B|+|C|}{n}\right)^{1+\varepsilon} \leq 1$

Pf Fix any coordinate (say $i=1$)

Choose τ to be the median

Repeat

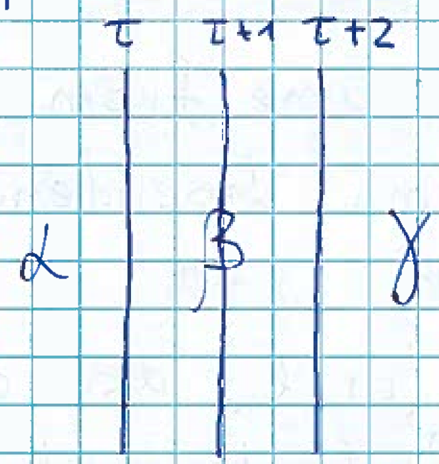
if $(1, \tau+1)$ is not a good cut

$\tau = \tau + 2$

Until $\frac{|C|}{n} \ll \frac{1}{d}$

Cl Either find a good cut or stop after $O\left(\frac{\log \log d}{\varepsilon}\right)$ steps.

Pf



to the right of τ – $(\beta+\gamma)$-fraction

to the right of $\tau+2$ – γ-fraction.

$(\alpha+\beta)^{1+\varepsilon} + (\beta+\gamma)^{1+\varepsilon} > 1$ otherwise a good cut.

$(1-\gamma)^{1+\varepsilon} + (\beta+\gamma)^{1+\varepsilon} > 1$

$\gamma < 1 - \left(1 - (\beta+\gamma)^{1+\varepsilon}\right)^{\frac{1}{1+\varepsilon}} \approx \frac{(\beta+\gamma)^{1+\varepsilon}}{1+\varepsilon}$ if

$(\beta+\gamma)$ - small.

$$(\text{frac. to the right of } \tau+2)^{"} \lesssim \frac{(\text{f. to the right of } \tau)^{"}}{1+\varepsilon}$$

In $\frac{\log \log d}{\varepsilon}$ steps $\frac{1}{2}$ becomes $\ll \frac{1}{d}$.

Can do the same for all coordinates and directions.

If no coordinate cut then ball centered in the median (along each coordinate) of radius $\frac{\log \log d}{\varepsilon}$ contains almost all the points.

Now to the data structure.

Preprocessing:

1) If dense ball, then a) store its center, any data point in it.

b) Recurse on the remainder.

2) If coordinate cut, recurse on $A \cup B$ and $B \cup C$.

Query:

1) Dense ball:

a) If within 1 from the ball, output any point from it

b) If not recurse on the remainder

2) Coordinate cut

a) If to the left, recurse on $A \cup B$

b) Right $\Rightarrow$ recurse on $B \cup C$.

Analysis:

1) Always return a point within

$$\underbrace{2R}_{\text{ball radius}} + 1 = O\left(\frac{\log \log d}{\varepsilon}\right)$$

2) Space: $n^{1+\varepsilon}$ points by induction

$$(\text{left})^{1+\varepsilon} + (\text{right})^{1+\varepsilon} \leq n^{1+\varepsilon}$$

since cut is good.

$O(d n^{1+\varepsilon})$ total

3) Query time

cost $d \Rightarrow \times (1 - \Omega(1))$

cost $1 \Rightarrow \times (1 - \Omega(\frac{1}{d}))$

$d \log n$ cost $\quad n \mapsto 1$.