Orthogonal Range Search in High Dimensions?

(app. Less exact nearest neighbors..)

k-d tree: $O(n)$ space, $O(n^{1-1/d})$ time

range tree: $O(n \log^d n)$ space, $O(\log^d n)$ time
nontriv. only for $d \ll \log \log \log n$

better analysis:
consider dominance range

$$Q_d(n) \leq Q_d(\frac{n}{2}) + Q_d(n)$$

$$\Rightarrow Q_d(n) \leq O\left(\left(\frac{\log d + d}{d}\right)\right)$$

$$\approx O\left(\left(\frac{\log n + d}{d}\right)\right)^d$$

nontriv for $d \leq 0.29 \log n$

larger $d$?
Consider offline case, given $n$ data pts (blue) & $m$ query pts (red)

Old Method
divide by median blue pt

$$T_d(n, m) \leq \max_{x \in [0,1]} T_d\left(\frac{n}{2}, (1-x)m\right) + T_d\left(\frac{n}{2}, (1-x)m\right) + T_d(n, m)$$
as above
New Method 1 (Impagliazzo et al. '14 / C.'15)

idea - rop-sided divide&conquer

\[(1-\alpha)n\] blue pts

\[\alpha n\] blue pts

\[\beta m\] red pts

\[(1-\beta)m\] red pts

choose dividing hyperplane s.t. \(\alpha = \beta\)

\[T_d(n,m) \leq \max_{\alpha \in [0,1]} T_d((1-\alpha)n, \alpha m) + T_d(\alpha n, (1-\alpha)m) + T_{d-1}((1-\alpha)n, (1-\alpha)m)\]

\[\Rightarrow T_d(n,n) \leq n^{2-\Theta(c\log c)} \quad \text{for } d = c \log n.\]

nontriv for \(c \ll \log n\) i.e. \(d \ll (\log^2 n)\).

Online queries?

Method 2 (C.'17)

idea - rop-sided d & c again!

\[(1-\alpha)n\] blue pts

\[\alpha n\] blue pts

pick \(\alpha \sim \frac{1}{(c\log c)^4}\)

- pick a rand. axis to divide
- store range trees for all subsets of \(c\log n\) pts
class

\[\Rightarrow O(n^{1+\epsilon})\] space, \(n^{1-\Theta(c\log c)}\) query time (rand.)
Method 3 (Abiad-Williams-Yao '16)

\[ \text{back to offline: } \left( \frac{N^2}{e} \right)^{\Theta(g)} \text{ time for } d = \log n \]

Idea: reduce to \( \{0,1\} \) case by divide & conquer
- solve \( \{0,1\} \) case by the polynomial method + matrix mult.

Divide into \( \frac{N}{5} \) groups of 5 pts.
Given groups A, B,

\[ F(A, B) = \bigwedge_{a \in A, b \in B} [a \text{ not dom by } b] \]

\[ = \bigwedge_{a \in A, b \in B} \bigvee_{i=1}^{d} a \cdot \bar{b}_i \]

poly of large deg
\[ \left( \bigvee_{x_i} = 1 - \prod_{i=1}^{d} (-x_i) \right) \]

Rand. trick: replace \( \bigvee_{x_i} \) by \( \bigvee_{i=1}^{d} \sum_{j=1}^{c} x_i \mod 2 \)

err prob \( \frac{1}{2^t} \)

set \( t = \Theta(\log s) \)
poly of \( O(\log s) \) deg.

Conditional Lower Bd.
If \( \exists \) algm for \( \{0,1\} \) dominance on \( N \) pts
with \( O(N^{1.99}) \) time for \( d = \Theta(\log N) \)
then \( \exists \) algm for SAT with \( kn \) clauses
with \( O(1.999^n) \) time for all const \( k \)
(dispersing Strong Exp Time Hypothesis (SETH))
Pf by Ex:

\[(x_1 \lor x_2 \lor x_3 \lor x_4) \land (x_1 \lor \overline{x}_2 \lor \overline{x}_3 \lor x_4) \land \ldots\]

is satisfiable

iff some pt (\(x_1 \lor x_2, \overline{x}_1 \lor \overline{x}_2, \ldots\))

is dominated by (\(x_3 \lor x_4, \overline{x}_3 \lor x_4, \ldots\)).

\(-k-SAT\) reduces to dominance for \(N = 2^{m/2} \text{ pts}\) in dim \(d \leq kN \leq O(k \log N)\)

\(N^{1.99} = 2^{1.995n}\) \(Q.E.D\)