ANN in High Dims: Offline Case (L_2)

given n query pts in advance

LSH: $\tilde{O}(n^{1+1/c^2})$ total time for approx factor $c = 1+\varepsilon$

$= \tilde{O}(n^{1+1/(1+\varepsilon)^2}) = \tilde{O}(n^{2-2\varepsilon+O(\varepsilon^2)})$

data-dependent LSH: $\tilde{O}(n^{1+1/(2c^2-1)}) = \tilde{O}(n^{2-4\varepsilon+O(\varepsilon^2)})$

Better?

Valiant '12 $\tilde{O}(n^{2-\Theta(\sqrt{\varepsilon})})$ [beats LSH for suff small ε]

Alman-C-Williams '16 $\tilde{O}(n^{2-\tilde{\Theta}(\varepsilon^{1/3})})$

by the polynomial method

By dim reduction (JL lemma), assume $d = O(\frac{1}{\varepsilon^2}\log n)$

Focus on ~~farthest neighbor~~ farthest pair

Divide input into $\frac{n}{s}$ groups of s pts.

Consider two groups A, B of s pts.

Want $F(A,B) = \max_{p\in A, q\in B} \|p-q\|$

Idea 1 - replace max with sum of k^{th} powers (k even)

compute $\hat{F}(A,B) = \sum_{p\in A, q\in B} \|p-q\|^k$

Then $F^k \leq \hat{F} \leq s^2 F^k$

$\Rightarrow F \leq \hat{F}^{1/k} \leq s^{2/k} F$

$\Rightarrow$ approx factor $s^{2/k} = e^{\frac{2}{k}\ln s}$

$\boxed{\text{set } k = \frac{2}{\varepsilon}\ln s}$ $\leq e^{\varepsilon} \sim 1+\varepsilon$

Why is $\hat{F}$ nice? it's polynomial in $2ds$ vars of deg k

e.g. $((x_1-y_1)^2 + (x_2-y_2)^2 + \dots)^{k/2} + \dots$

Fact Given polynomial $F: \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$
& set P, Q of N vectors in $\mathbb{R}^D$,
can evaluate $F(p,q)\ \forall p \in P, q \in Q$
by multiplying $N \times D'$ and $D' \times N$ matrices.
in $M(N, D', N)$ time
where $D' = \#$ terms (monomials) in F

Ex $F(x,y,z,w) = 2xyz + 3xw^2 + yzw$
to compute $F(x,y,z,w)$ for all $(x,y) \in P$, $(z,w) \in Q$

$$\begin{pmatrix} 2xy & 3x & y \\ & \vdots & \\ & & \end{pmatrix} \begin{pmatrix} z & & \\ w^2 & \cdots & \\ zw & & \end{pmatrix} = \begin{pmatrix} & & \\ & & \\ & & \end{pmatrix}$$

Fact $M(N,N,N) = O(N^{2.81})$ Strassen '69

$O(N^{2.373})$ Strother/Vassilevska-W. '13

$M(N, N^{0.17}, N) = \tilde{O}(N^2)$ Coppersmith '82

In our appl'n, $D = ds$, $N = \frac{n}{s}$

$$D' = s^2 \binom{2d+k}{k} \leftarrow \text{\# monomials of deg } k \text{ in } 2d \text{ vars}$$

$x_1^{k_1} x_2^{k_2} \cdots x_{2d}^{k_{2d}}$
$k_1 + \cdots + k_{2d} = k$

$$= s^2 \cdot \binom{O(\frac{1}{\varepsilon^2} \log n)}{O(\frac{1}{\varepsilon} \log s)}$$

$$\approx s^2 \left(\frac{1}{\varepsilon}\frac{\log n}{\log s}\right)^{O\left(\frac{1}{\varepsilon}\log s\right)} \qquad \left[\binom{a}{b} \approx \left(\frac{ea}{b}\right)^b.\right]$$

$$\approx n^{2\alpha}\left(\frac{1}{\varepsilon\alpha}\right)^{O\left(\frac{\alpha}{\varepsilon}\log n\right)} \qquad \text{set } s = n^\alpha$$

$$\approx n^{O\left(\frac{\alpha}{\varepsilon}\log\frac{1}{\varepsilon\alpha}\right)} \qquad \text{set } \alpha = \tilde{\Theta}(\varepsilon).$$

$$\ll \left(\frac{n}{s}\right)^{0.17}$$

$$\Rightarrow \text{total time } \tilde{O}\left(\left(\frac{n}{s}\right)^2\right) = \tilde{O}\left(n^{2-\tilde{O}(\varepsilon)}\right)$$

not better than LSH!

Idea 2 - Chebyshev polynomials

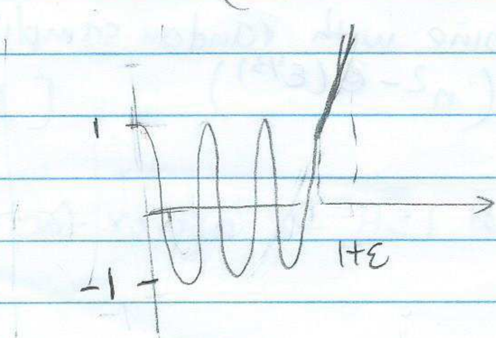
$T_2(x) = 2x^2 - 1$

$T_3(x) = 4x^3 - 3x$ $\qquad$ $\cos(k\theta) = T_k(\cos\theta)$

$T_4(x) = 8x^4 - 8x^2 + 1$

$\vdots$

$$T_k(x) = \begin{cases} \cos(k \arccos x) & \text{if } |x| \le 1 \\ \cosh(k \operatorname{arccosh} x) & \text{if } |x| > 1 \end{cases}$$



$\cosh x \sim 1 + \Theta(x^2)$

Fact $|x| \le 1 \Rightarrow |T_k(x)| \le 1$

$x \geq 1+\varepsilon \Rightarrow T_k(x) \geq \cosh(k \operatorname{arccosh}(1+\varepsilon))$

$\gtrsim \cosh(k\sqrt{\varepsilon})$

$\geq \frac{1}{2} e^{k\sqrt{\varepsilon}}$

Consider approx decis problem for fixed r.

Say $r=1$.

Compute $\tilde{F}(A,B) = \sum_{p\in A, q\in B} T_k(\|p-q\|)$ (k even)

Then farthest pair dist $\le 1 \Rightarrow \tilde{F} \le s^2 \cdot 1$

" " $\geq 1+\varepsilon \Rightarrow \tilde{F} \geq \frac{1}{2} e^{k\sqrt{\varepsilon}} - s^2$

$> s^2$

$\boxed{\text{Set } k = \frac{3}{\sqrt{\varepsilon}} \ln s}$

(instead of $\frac{2}{\varepsilon} \ln s$)

$\Rightarrow$ set $\alpha = \tilde{\Theta}(\sqrt{\varepsilon})$

$\Rightarrow \boxed{\tilde{O}\left(n^{2-\tilde{\Theta}(\sqrt{\varepsilon})}\right)}$

Idea 3 - combine with random sampling of dims

$\tilde{O}(n^{2-\tilde{\Theta}(\varepsilon^{1/3})})$ [Alman-C-Williams '16]

[open: beat LSH for approx factor 2?]