Approximate Nearest Neighbors in l_∞

Piotr Indyk

Stanford University

l_∞ metric

Properties:

- any n-point metric space can be $O(1)$-embedded into l_∞^d with $d = n^\alpha$ [Matousek'96]
- lower bounds for d $\rightarrow$ dimensionality reduction not possible
- embeddings of Hausdorff metric [Farach-Colton, Indyk'99]

Best result [Gabow, Bentley, Tarjan'84]:

- $O(\log^d n)$ query time
- $O(n \log^d n)$ space

Results [I'98]

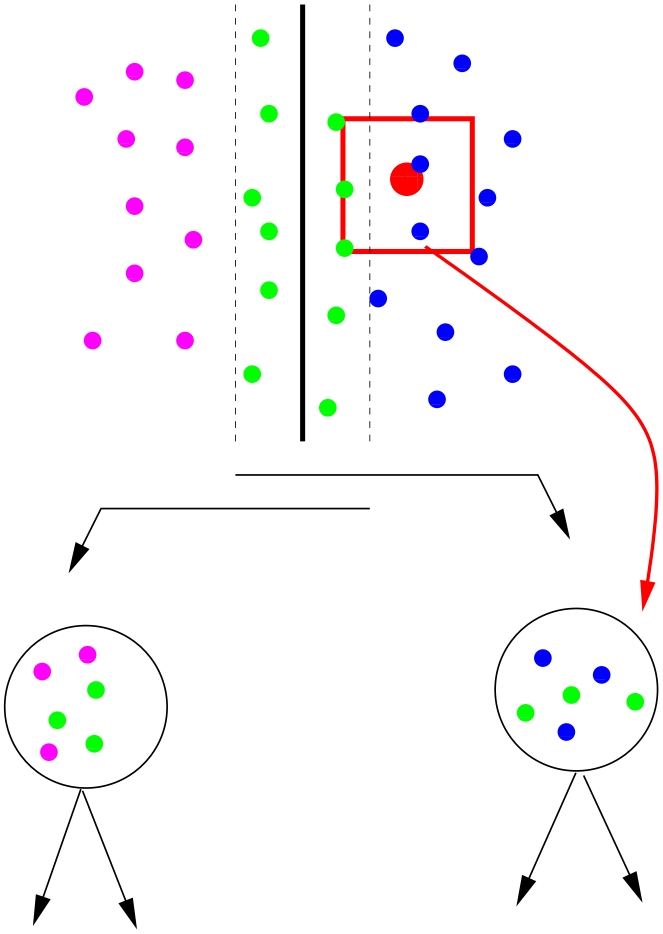
Upper bound (for ϵ-PLEB):

- query time: $O(d \log^{O(1)} n)$
- storage: $dn + n^{1+\rho}$
- error: $O(\log_{1+\rho} \log d)$

Lower bound:

- superset query reduces to ϵ-NN in l_∞^d with $\epsilon < 2$
- latter seems hard

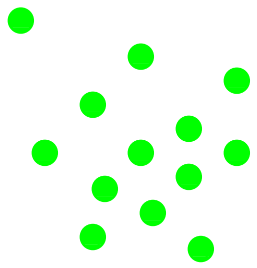
Idea I: separators



Query time ≈ tree depth.

Idea II

No balanced separator!

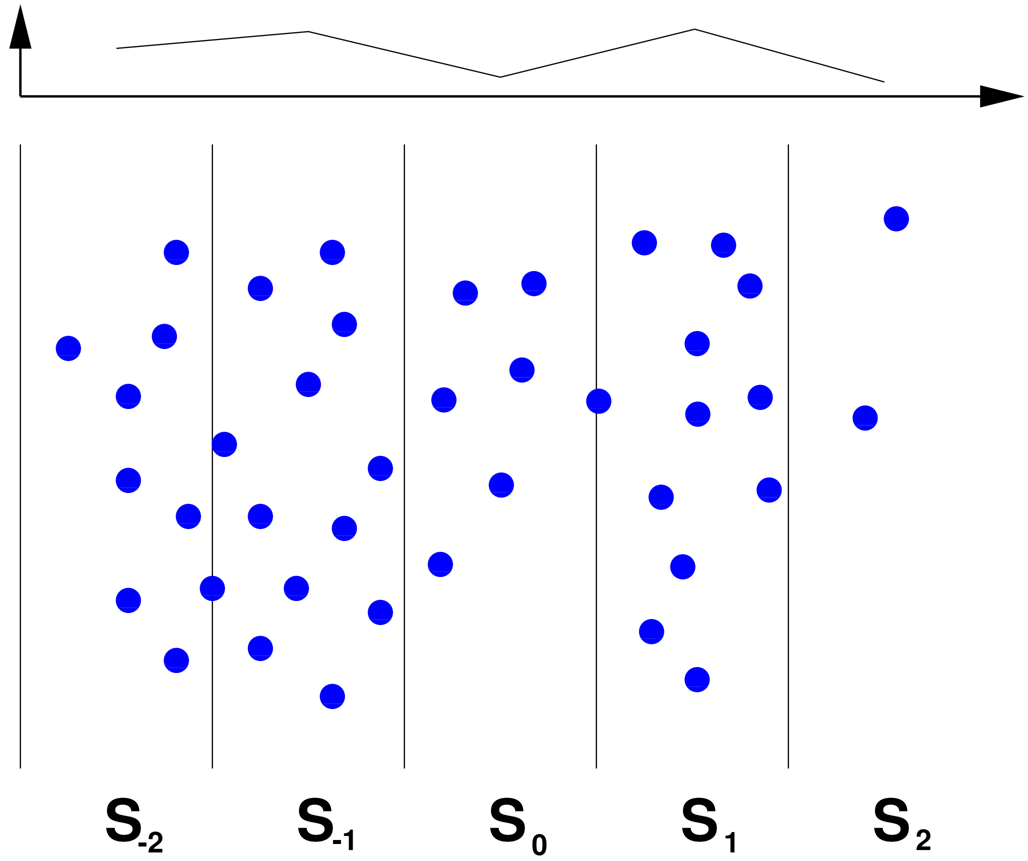


However, diameter is small

Approach:

- use separator, or
- find large cluster of small diameter

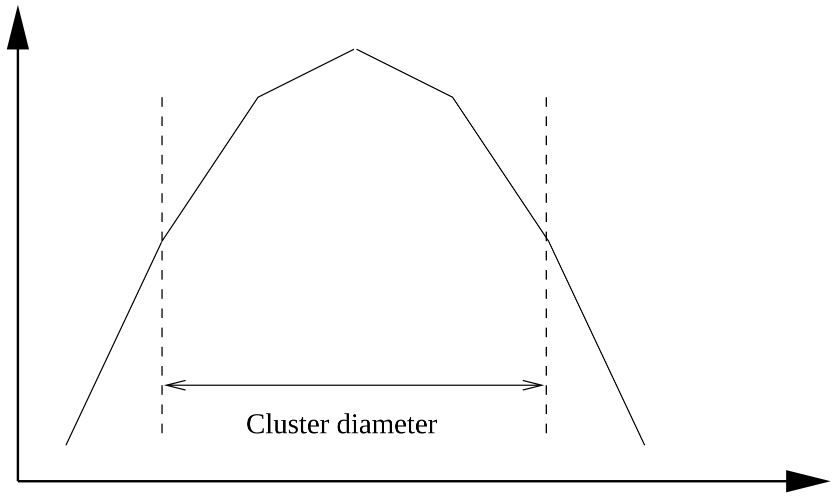
Profile



Case 1: Flat profile.

Profile II

Case 2: Steep profile



- make the tails $= \frac{1}{2d}$
- cut in all dimensions $\rightarrow 50\%$ left
- remove the cluster

Separator:details

Claim: Assume that

- αn points to the left
- βn points in the middle
- γn points to the right

Then we use $\tilde{O}(n^{1+L})$ storage, where

$$L = L(\beta, \gamma) = \log_{1/(\beta+\gamma)} \frac{\beta + \gamma}{\gamma}$$

In fact, if $L(\beta, \gamma) \leq \rho$, then storage is $\tilde{O}(n^{1+\rho})$.

Claim: Let $t_i = \sum_{j \geq i} s_j$. Assume $L(s_i, t_{i+1}) > \rho$. Then $t_{i+1} < t_i^{1+\rho}$.

Query time

Path to the leaf:

- separator node: $O(d \log n)$ times, $O(1)$ per node
- box node: $O(\log n)$ times, $O(d)$ per node

Total time: $O(d \log n)$.

Lower bound

Superset query problem (SQ):

Given: sets $S_1 \dots S_n \subset U$

Goal: a data structure which for any query set $Q \subset U$ returns the sets $S_i \subset Q$.

We reduce SQ to ϵ-PLEB for $\epsilon < 2$.

Lower bound - proof

Claim: There is a map $S_i \to f[S_i]$, $Q \to g[Q]$ such that

$$|g[Q] - f[S]|_\infty = \begin{cases} \frac{1}{3} & \text{if } S \subset Q \\ 1 & \text{otherwise} \end{cases}$$

Proof: Define

$$g(x) = \begin{cases} 0 & \text{if } x = 0 \\ \frac{2}{3} & \text{if } x = 1 \end{cases}$$

$$f(x) = \begin{cases} \frac{1}{3} & \text{if } x = 0 \\ 1 & \text{if } x = 1 \end{cases}$$

- if there is $a \in S - Q$, then
  $|g[Q] - f[S]|_\infty \geq |g(0) - f(1)| \geq 1$
- otherwise (if $S \subset Q$):
  - if $a \notin S$, then $|g(x) - f(0)| = \frac{1}{3}$
  - if $a \in S$ then $a \in Q$ and $|g(1) - f(1)| = \frac{1}{3}$