

- Problem: Given a set X of n points, a set \mathcal{R} of m disks (each of radius δ), and a positive integer k , determine whether there is a subset X' of X of size (at most) k such that every disk in \mathcal{R} has at least one point in X' . \leftarrow this problem is NP-hard
- associate a unit weight to each point in X ; for ℓ number of iterations, do the following -

(i) compute an ϵ -net N of X of size $ck \lg k$ (where c is a function of ϵ)

(ii) if every disk is hit by N

output N and exit

else

for some disk D that is not hit by N , double the weight of each point in D \leftarrow iterative reweighing

- Compute a random sample N of points in X while picking each point $x \in X$ into N based on the weight associated to x , and (let us call) the resultant set an ϵ -net of X . From the way an ϵ -net is constructed, for any disk $D \in \mathcal{R}$, if $\sum_{x \in X \cap D} w(x) \geq \epsilon \sum_{x \in X} w(x)$, then D contains at least one point of N with high probability. \leftarrow some more details, in terms of δ , are below

- Setting ℓ and ϵ :

At the start, the total weight $W_0 = n$. Significantly, at the end of i^{th} iteration, if a disk $D \in \mathcal{R}$ is not hit, since N is an ϵ -net, the total weight of points in D is $< \epsilon W_{i-1}$ with high probability. Hence, $W_i \leq W_{i-1} + \epsilon W_{i-1} \Rightarrow W_i \leq n(1 + \epsilon)^i \leq ne^{i\epsilon}$.

For any point j in an optimal solution, let $t_i(j)$ be the number of times j 's weight got doubled by the end of i^{th} iteration. Then, the weight of optimal hitting set at the end of i^{th} iteration OPT_i is $\geq \sum_{j=1}^k 2^{t_i(j)}$. Considering the latter is a convex function, it is minimized when $t_i(1), t_i(2), \dots, t_i(k)$ are as equal to each other as possible. Since at least one point's weight in optimal solution is doubled in each iteration, the minimum occurs when $t_i(j) = \lfloor \frac{i}{k} \rfloor$ for all j . Hence, the rightmost expression is $\geq k 2^{\lfloor \frac{i}{k} \rfloor}$.

Noting $OPT_i \leq W_i$, $k 2^{\frac{i}{k}} \leq ne^{i\epsilon} \Rightarrow (\lg k) + \frac{i}{k} \leq (\lg n) + (i\epsilon)(\lg_2 e) \Rightarrow i(\frac{1}{k} - 2\epsilon) \leq \lg(\frac{n}{k})$.

By choosing $\epsilon = \frac{1}{4k}$, $i \leq (2k) \lg \frac{n}{k}$.

That is, if the algorithm runs for more than $(2k) \lg(\frac{n}{k})$ iterations, there cannot be a hitting set of size k with high probability. Hence, set ℓ equal to $(2k) \lg(\frac{n}{k})$.

- If there is a hitting set of size k , then this algorithm outputs a hitting set of size $ck \lg k$ with high probability.

For the optimization version of this problem (in which X' of minimum cardinality needs to be found), run the above pseudocode by assigning values to k via binary searching in $[1, n]$. Hence, for a fixed ϵ , approx factor is $O(\lg(OPT))$.

- time complexity: number of iterations is $O(k \lg(n/k)) = O(k \lg n)$; generating an ϵ -net via (weighted) random sampling of X takes $O(n)$ time per iteration; checking whether $ck \lg k$ elements of an ϵ -net hits each of the m disks takes $cmk \lg k$ time per iteration

-
- A *range space* is a pair (X, \mathcal{R}) , where X is a ground set (finite or infinite) and \mathcal{R} is a (finite or infinite) family of subsets of X . The elements of X are (*usually*) *points* and the elements of \mathcal{R} are *ranges*.
 - Let $S = (X, \mathcal{R})$ be a range space. For $Y \subseteq X$, let $\mathcal{R}|_Y = \{r \cap Y \mid r \in \mathcal{R}\}$ denote the *projection* (a.k.a. *restriction*) of \mathcal{R} on Y . The range space S projected to Y is $S|_Y = (Y, \mathcal{R}|_Y)$.

If $\mathcal{R}|_Y$ contains all the subsets of Y (i.e., if Y is finite, we have $|\mathcal{R}|_Y| = 2^{|Y|}$), then Y is said to have been *shattered* by \mathcal{R} (or equivalently Y is shattered by S).

The VC^1 -dimension of S , denoted by $\dim_{VC}(S)$, is the maximum cardinality of a shattered subset of X . If there are arbitrarily large shattered subsets, then $\dim_{VC}(S) = \infty$.

examples:

- set X is the real line and \mathcal{R} is the set of all intervals on the real line – VC-dimension is 2
- set X is \mathbb{R}^2 and \mathcal{R} is the set of disks in the plane – VC-dimension is 3
- set X is \mathbb{R}^2 and \mathcal{R} is the set of all (closed) convex sets in the plane – VC-dimension is unbounded
- *Sauer's lemma*: If (X, \mathcal{R}) is a range space of VC dimension δ with $|X| = n$, then $|\mathcal{R}| \leq n^\delta$. — proof to be filled-in
- Given a range sapce (X, \mathcal{R}) , $x \subseteq X$, and any $0 < \epsilon \leq 1$, a subset $N \subseteq x$ is an ϵ -net for (X, \mathcal{R}) iff for every range $r \in \mathcal{R}$ with $|r \cap x| \geq \epsilon|x|$, r contains at least one point of N .

ϵ -Net theorem: Given any range space (X, \mathcal{R}) with VC dimension δ , a finite subset $x \subseteq X$, $0 < \epsilon \leq 1$, and $0 < \varphi < 1$, any random subset $N \subseteq x$ of cardinality $m \geq \max(\frac{4}{\epsilon} \lg \frac{4}{\varphi}, \frac{8\delta}{\epsilon} \lg \frac{16}{\epsilon})$ is an ϵ -net for (X, \mathcal{R}) with probability at least $1 - \varphi$. (That is, assuming δ and φ are constants, $m \geq c \cdot \frac{1}{\epsilon} \cdot \lg(\frac{1}{\epsilon})$ for a sufficiently large constant c .) — proof to be filled-in

References:

Geometric Approximation Algorithms by S. Har-Peled.

¹stands for Vapnik-Chervonenkis