Chapter 38

On Complexity, Sampling, and ε-Nets and ε-Samples

By Sariel Har-Peled, April 26, 2022[①]

> "I've never touched the hard stuff, only smoked grass a few times with the boys to be polite, and that's all, though ten is the age when the big guys come around teaching you all sorts to things. But happiness doesn't mean much to me, I still think life is better. Happiness is a mean son of a bitch and needs to be put in his place. Him and me aren't on the same team, and I'm cutting him dead. I've never gone in for politics, because somebody always stand to gain by it, but happiness is an even crummier racket, and their ought to be laws to put it out of business."
>
> Momo, Emile Ajar

In this chapter we will try to quantify the notion of geometric complexity. It is intuitively clear that a ● (i.e., disk) is a simpler shape than an ⬬ (i.e., ellipse), which is in turn simpler than a ☻ (i.e., smiley). This becomes even more important when we consider several such shapes and how they interact with each other. As these examples might demonstrate, this notion of complexity is somewhat elusive.

To this end, we show that one can capture the structure of a distribution/point set by a small subset. The size here would depend on the complexity of the shapes/ranges we care about, but surprisingly it would be independent of the size of the point set.

38.1. VC dimension

Definition 38.1.1. A ***range space*** S is a pair $(\mathsf{X}, \mathcal{R})$, 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 ***points*** and the elements of $\mathcal{R}$ are ***ranges***.

Our interest is in the size/weight of the ranges in the range space. For technical reasons, it will be easier to consider a finite subset x as the underlining ground set.

Definition 38.1.2. Let $\mathsf{S} = (\mathsf{X}, \mathcal{R})$ be a range space, and let x be a finite (fixed) subset of X. For a range $\mathbf{r} \in \mathcal{R}$, its ***measure*** is the quantity

$$\overline{m}(\mathbf{r}) = \frac{|\mathbf{r} \cap \mathsf{x}|}{|\mathsf{x}|}.$$

While x is finite, it might be very large. As such, we are interested in getting a good estimate to $\overline{m}(\mathbf{r})$ by using a more compact set to represent the range space.

Definition 38.1.3. Let $\mathsf{S} = (\mathsf{X}, \mathcal{R})$ be a range space. For a subset N (which might be a multi-set) of x, its ***estimate*** of the measure of $\overline{m}(\mathbf{r})$, for $\mathbf{r} \in \mathcal{R}$, is the quantity

$$\overline{s}(\mathbf{r}) = \frac{|\mathbf{r} \cap N|}{|N|}.$$

The main purpose of this chapter is to come up with methods to generate a sample N, such that $\overline{m}(\mathbf{r}) \approx \overline{s}(\mathbf{r})$, for all the ranges $\mathbf{r} \in \mathcal{R}$.

It is easy to see that in the worst case, no sample can capture the measure of all ranges. Indeed, given a sample N, consider the range $\mathsf{x} \setminus N$ that is being completely missed by N. As such, we need to concentrate on range spaces that are "low dimensional", where not all subsets are allowable ranges. The notion of VC dimension (named after Vapnik and Chervonenkis [VC71]) is one way to limit the complexity of a range space.

Definition 38.1.4. Let $\mathsf{S} = (\mathsf{X}, \mathcal{R})$ be a range space. For $Y \subseteq \mathsf{X}$, let

$$\mathcal{R}_{|Y} = \left\{ \mathbf{r} \cap Y \;\middle|\; \mathbf{r} \in \mathcal{R} \right\} \tag{38.1}$$

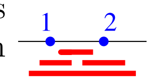
denote the ***projection*** of $\mathcal{R}$ on Y. The range space S projected to Y is $\mathsf{S}_{|Y} = \left(Y, \mathcal{R}_{|Y}\right)$.

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

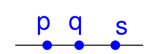
The ***Vapnik-Chervonenkis*** dimension (or VC ***dimension***) of S, denoted by $\dim_{\mathsf{VC}}(\mathsf{S})$, is the maximum cardinality of a shattered subset of X. If there are arbitrarily large shattered subsets, then $\dim_{\mathsf{VC}}(S) = \infty$.

38.1.1. Examples

Intervals. Consider the set X to be the real line, and consider $\mathcal{R}$ to be the set of all intervals on the real line. Consider the set $Y = \{1, 2\}$. Clearly, one can find four intervals that contain all possible subsets of Y. Formally, the projection $\mathcal{R}_{|Y} = \{\{\,\}, \{1\}, \{2\}, \{1, 2\}\}$. The intervals realizing each of these subsets are depicted on the right.

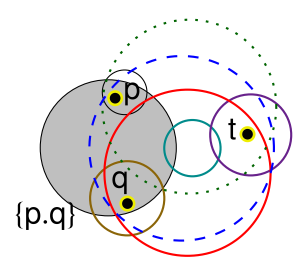


However, this is false for a set of three points $B = \{\mathsf{p}, \mathsf{u}, \mathsf{v}\}$, since there is no interval that can contain the two extreme points p and v without also containing u. Namely, the subset $\{\mathsf{p}, \mathsf{v}\}$ is not realizable for intervals, implying that the largest shattered set by the range space (real line, intervals) is of size two. We conclude that the VC dimension of this space is two.

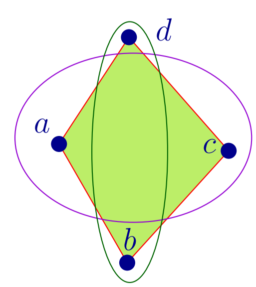


Disks. Let $\mathsf{X} = \mathbb{R}^2$, and let $\mathcal{R}$ be the set of disks in the plane. Clearly, for any three points in the plane (in general position), denoted by p, u, and v, one can find eight disks that realize all possible 2^3 different subsets. See the figure on the right.

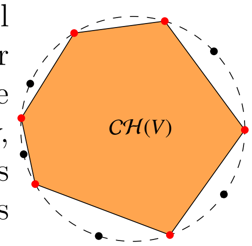
But can disks shatter a set with four points? Consider such a set P of four points. If the convex hull of P has only three points on its boundary, then the subset X having only those three vertices (i.e., it does not include the middle point) is impossible, by convexity. Namely, there is no disk that contains only the points of X without the middle point.



Alternatively, if all four points are vertices of the convex hull and they are a, b, c, d along the boundary of the convex hull, either the set $\{a, c\}$ or the set $\{b, d\}$ is not realizable. Indeed, if both options are realizable, then consider the two disks D_1 and D_2 that realize those assignments. Clearly, ∂D_1 and ∂D_2 must intersect in four points, but this is not possible, since two circles have at most two intersection points. See the figure on the left. Hence the VC dimension of this range space is 3.



Convex sets. Consider the range space $\mathsf{S} = (\mathbb{R}^2, \mathcal{R})$, where $\mathcal{R}$ is the set of all (closed) convex sets in the plane. We claim that $\dim_{\mathsf{VC}}(\mathsf{S}) = \infty$. Indeed, consider a set U of n points $p_1, \ldots, p_n$ all lying on the boundary of the unit circle in the plane. Let V be any subset of U, and consider the convex hull $\mathcal{CH}(V)$. Clearly, $\mathcal{CH}(V) \in \mathcal{R}$, and furthermore, $\mathcal{CH}(V) \cap U = V$. Namely, any subset of U is realizable by S. Thus, S can shatter sets of arbitrary size, and its VC dimension is unbounded.



Complement. Consider the range space $\mathsf{S} = (\mathsf{X}, \mathcal{R})$ with $\delta = \dim_{\mathsf{VC}}(\mathsf{S})$. Next, consider the complement space, $\overline{\mathsf{S}} = \left(\mathsf{X}, \overline{\mathcal{R}}\right)$, where

$$\overline{\mathcal{R}} = \left\{ \mathsf{X} \setminus \mathbf{r} \;\middle|\; \mathbf{r} \in \mathcal{R} \right\}.$$

Namely, the ranges of $\overline{\mathsf{S}}$ are the complement of the ranges in S. What is the VC dimension of $\overline{\mathsf{S}}$? Well, a set $B \subseteq \mathsf{X}$ is shattered by $\overline{\mathsf{S}}$ if and only if it is shattered by S. Indeed, if S shatters B, then for any $Z \subseteq B$, we have that $(B \setminus Z) \in \mathcal{R}_{|B}$, which implies that $Z = B \setminus (B \setminus Z) \in \overline{\mathcal{R}}_{|B}$. Namely, $\overline{\mathcal{R}}_{|B}$ contains all the subsets of B, and $\overline{\mathsf{S}}$ shatters B. Thus, $\dim_{\mathsf{VC}}\left(\overline{\mathsf{S}}\right) = \dim_{\mathsf{VC}}(\mathsf{S})$.

Lemma 38.1.5. *For a range space $\mathsf{S} = (X, \mathcal{R})$ we have that $\dim_{\mathsf{VC}}(\mathsf{S}) = \dim_{\mathsf{VC}}\left(\overline{\mathsf{S}}\right)$, where $\overline{\mathsf{S}}$ is the complement range space.*

38.1.1.1. Halfspaces

Let $\mathsf{S} = (\mathsf{X}, \mathcal{R})$, where $\mathsf{X} = \mathbb{R}^d$ and $\mathcal{R}$ is the set of all (closed) halfspaces in $\mathbb{R}^d$. We need the following technical claim.

Claim 38.1.6. *Let $\mathsf{P} = \{\mathsf{p}_1, \ldots, \mathsf{p}_{d+2}\}$ be a set of $d+2$ points in $\mathbb{R}^d$. There are real numbers $\beta_1, \ldots, \beta_{d+2}$, not all of them zero, such that $\sum_i \beta_i \mathsf{p}_i = 0$ and $\sum_i \beta_i = 0$.*

Proof: Indeed, set $\mathsf{u}_i = (\mathsf{p}_i, 1)$, for $i = 1, \ldots, d+2$. Now, the points $\mathsf{u}_1, \ldots, \mathsf{u}_{d+2} \in \mathbb{R}^{d+1}$ are linearly dependent, and there are coefficients $\beta_1, \ldots, \beta_{d+2}$, not all of them zero, such that $\sum_{i=1}^{d+2} \beta_i \mathsf{u}_i = 0$. Considering only the first d coordinates of these points implies that $\sum_{i=1}^{d+2} \beta_i \mathsf{p}_i = 0$. Similarly, by considering only the $(d+1)$st coordinate of these points, we have that $\sum_{i=1}^{d+2} \beta_i = 0$. ■

To see what the VC dimension of halfspaces in $\mathbb{R}^d$ is, we need the following result of Radon. (For a reminder of the formal definition of convex hulls, see Definition 38.5.1.)

Theorem 38.1.7 (Radon's theorem). *Let $\mathsf{P} = \{\mathsf{p}_1, \ldots, \mathsf{p}_{d+2}\}$ be a set of $d+2$ points in $\mathbb{R}^d$. Then, there exist two disjoint subsets C and D of P, such that $\mathcal{CH}(C) \cap \mathcal{CH}(D) \neq \emptyset$ and $C \cup D = \mathsf{P}$.*

Proof: By Claim 38.1.6 there are real numbers $\beta_1, \ldots, \beta_{d+2}$, not all of them zero, such that $\sum_i \beta_i \mathsf{p}_i = 0$ and $\sum_i \beta_i = 0$.

Assume, for the sake of simplicity of exposition, that $\beta_1, \ldots, \beta_k \geq 0$ and $\beta_{k+1}, \ldots, \beta_{d+2} < 0$. Furthermore, let $\mu = \sum_{i=1}^{k} \beta_i = -\sum_{i=k+1}^{d+2} \beta_i$. We have that

$$\sum_{i=1}^{k} \beta_i \mathsf{p}_i = -\sum_{i=k+1}^{d+2} \beta_i \mathsf{p}_i.$$

In particular, $v = \sum_{i=1}^{k} (\beta_i/\mu) \mathsf{p}_i$ is a point in $\mathcal{CH}(\{\mathsf{p}_1, \ldots, \mathsf{p}_k\})$. Furthermore, for the same point v we have $v = \sum_{i=k+1}^{d+2} -(\beta_i/\mu) \mathsf{p}_i \in \mathcal{CH}(\{\mathsf{p}_{k+1}, \ldots, \mathsf{p}_{d+2}\})$. We conclude that v is in the intersection of the two convex hulls, as required. ■

The following is a trivial observation, and yet we provide a proof to demonstrate it is true.

Lemma 38.1.8. *Let* $\mathsf{P} \subseteq \mathbb{R}^d$ *be a finite set, let* v *be any point in* $\mathcal{CH}(\mathsf{P})$*, and let* h^+ *be a halfspace of* $\mathbb{R}^d$ *containing* v*. Then there exists a point of* P *contained inside* h^+*.*

Proof: The halfspace h^+ can be written as $h^+ = \left\{ \mathsf{x} \in \mathbb{R}^d \;\middle|\; \langle \mathsf{x}, v \rangle \leq c \right\}$. Now $\mathsf{v} \in \mathcal{CH}(\mathsf{P}) \cap h^+$, and as such there are numbers $\alpha_1, \ldots, \alpha_m \geq 0$ and points $\mathsf{p}_1, \ldots, \mathsf{p}_m \in \mathsf{P}$, such that $\sum_i \alpha_i = 1$ and $\sum_i \alpha_i \mathsf{p}_i = \mathsf{v}$. By the linearity of the dot product, we have that

$$\mathsf{v} \in h^+ \implies \langle \mathsf{v}, v \rangle \leq c \implies \left\langle \sum_{i=1}^{m} \alpha_i \mathsf{p}_i, v \right\rangle \leq c \implies \beta = \sum_{i=1}^{m} \alpha_i \langle \mathsf{p}_i, v \rangle \leq c.$$

Setting $\beta_i = \langle \mathsf{p}_i, v \rangle$, for $i = 1, \ldots, m$, the above implies that β is a weighted average of $\beta_1, \ldots, \beta_m$. In particular, there must be a β_i that is no larger than the average. That is $\beta_i \leq c$. This implies that $\langle \mathsf{p}_i, v \rangle \leq c$. Namely, $\mathsf{p}_i \in h^+$ as claimed. ■

Let S be the range space having $\mathbb{R}^d$ as the ground set and all the close halfspaces as ranges. Radon's theorem implies that if a set Q of $d + 2$ points is being shattered by S, then we can partition this set Q into two disjoint sets Y and Z such that $\mathcal{CH}(Y) \cap \mathcal{CH}(Z) \neq \emptyset$. In particular, let v be a point in $\mathcal{CH}(Y) \cap \mathcal{CH}(Z)$. If a halfspace h^+ contains all the points of Y, then $\mathcal{CH}(Y) \subseteq h^+$, since a halfspace is a convex set. Thus, any halfspace h^+ containing all the points of Y will contain the point $\mathsf{v} \in \mathcal{CH}(Y)$. But $\mathsf{v} \in \mathcal{CH}(Z) \cap h^+$, and this implies that a point of Z must lie in h^+, by Lemma 38.1.8. Namely, the subset $Y \subseteq \mathsf{Q}$ cannot be realized by a halfspace, which implies that Q cannot be shattered. Thus $\dim_{\mathsf{VC}}(S) < d + 2$. It is also easy to verify that the regular simplex with $d + 1$ vertices is shattered by S. Thus, $\dim_{\mathsf{VC}}(S) = d + 1$.

38.2. Shattering dimension and the dual shattering dimension

The main property of a range space with bounded VC dimension is that the number of ranges for a set of n elements grows polynomially in n (with the power being the dimension) instead of exponentially. Formally, let the ***growth function*** be

$$\mathcal{G}_\delta(n) = \sum_{i=0}^{\delta} \binom{n}{i} \leq \sum_{i=0}^{\delta} \frac{n^i}{i!} \leq n^\delta, \tag{38.2}$$

for $\delta > 1$ (the cases where $\delta = 0$ or $\delta = 1$ are not interesting and we will just ignore them). Note that for all $n, \delta \geq 1$, we have $\mathcal{G}_\delta(n) = \mathcal{G}_\delta(n-1) + \mathcal{G}_{\delta-1}(n-1)$[2].

Lemma 38.2.1 (Sauer's lemma). *If* $(\mathsf{X}, \mathcal{R})$ *is a range space of* VC *dimension* δ *with* $|\mathsf{X}| = n$*, then* $|\mathcal{R}| \leq \mathcal{G}_\delta(n)$*.*

Proof: The claim trivially holds for $\delta = 0$ or $n = 0$.

Let x be any element of X, and consider the sets

$$\mathcal{R}_x = \left\{ \mathbf{r} \setminus \{x\} \;\middle|\; \mathbf{r} \cup \{x\} \in \mathcal{R} \text{ and } \mathbf{r} \setminus \{x\} \in \mathcal{R} \right\} \quad \text{and} \quad \mathcal{R} \setminus x = \left\{ \mathbf{r} \setminus \{x\} \;\middle|\; \mathbf{r} \in \mathcal{R} \right\}.$$

[2]Here is a cute (and standard) counting argument: $\mathcal{G}_\delta(n)$ is just the number of different subsets of size at most δ out of n elements. Now, we either decide to not include the first element in these subsets (i.e., $\mathcal{G}_\delta(n-1)$) or, alternatively, we include the first element in these subsets, but then there are only $\delta - 1$ elements left to pick (i.e., $\mathcal{G}_{\delta-1}(n-1)$).

Observe that $|\mathcal{R}| = |\mathcal{R}_x| + |\mathcal{R} \setminus x|$. Indeed, we charge the elements of $\mathcal{R}$ to their corresponding element in $\mathcal{R} \setminus x$. The only bad case is when there is a range $\mathbf{r}$ such that both $\mathbf{r} \cup \{x\} \in \mathcal{R}$ and $\mathbf{r} \setminus \{x\} \in \mathcal{R}$, because then these two distinct ranges get mapped to the same range in $\mathcal{R} \setminus x$. But such ranges contribute exactly one element to $\mathcal{R}_x$. Similarly, every element of $\mathcal{R}_x$ corresponds to two such "twin" ranges in $\mathcal{R}$.

Observe that $(X \setminus \{x\}, \mathcal{R}_x)$ has VC dimension $\delta - 1$, as the largest set that can be shattered is of size $\delta - 1$. Indeed, any set $B \subset \mathsf{X} \setminus \{x\}$ shattered by $\mathcal{R}_x$ implies that $B \cup \{x\}$ is shattered in $\mathcal{R}$.

Thus, we have

$$|\mathcal{R}| = |\mathcal{R}_x| + |\mathcal{R} \setminus x| \le \mathcal{G}_{\delta-1}(n-1) + \mathcal{G}_{\delta}(n-1) = \mathcal{G}_{\delta}(n),$$

by induction. ■

Interestingly, Lemma 38.2.1 is tight.

Next, we show pretty tight bounds on $\mathcal{G}_{\delta}(n)$. The proof is technical and not very interesting, and it is delegated to Section 38.4.

Lemma 38.2.2. *For $n \ge 2\delta$ and $\delta \ge 1$, we have* $\left(\frac{n}{\delta}\right)^{\delta} \le \mathcal{G}_{\delta}(n) \le 2\left(\frac{ne}{\delta}\right)^{\delta}$, *where* $\mathcal{G}_{\delta}(n) = \sum_{i=0}^{\delta} \binom{n}{i}$.

Definition 38.2.3 (Shatter function). Given a range space $\mathsf{S} = (\mathsf{X}, \mathcal{R})$, its ***shatter function*** $\pi_{\mathsf{S}}(m)$ is the maximum number of sets that might be created by S when restricted to subsets of size m. Formally,

$$\pi_{\mathsf{S}}(m) = \max_{\substack{B \subset \mathsf{X} \\ |B| = m}} \left|\mathcal{R}_{|B}\right|;$$

see Eq. (38.1).

Our arch-nemesis in the following is the function $x/\ln x$. The following lemma states some properties of this function, and its proof is left as an exercise.

Lemma 38.2.4. *For the function $f(x) = x/\ln x$ the following hold.*

(A) $f(x)$ is monotonically increasing for $x \ge e$.
(B) $f(x) \ge e$, for $x > 1$.
(C) For $u \ge \sqrt{e}$, if $f(x) \le u$, then $x \le 2u \ln u$.
(D) For $u \ge \sqrt{e}$, if $x > 2u \ln u$, then $f(x) > u$.
(E) For $u \ge e$, if $f(x) \ge u$, then $x \ge u \ln u$.

38.2.1. Mixing range spaces

Lemma 38.2.5. *Let $\mathsf{S} = (X, \mathcal{R})$ and $\mathsf{T} = (X, \mathcal{R}')$ be two range spaces of VC dimension δ and δ', respectively, where $\delta, \delta' > 1$. Let $\widehat{\mathcal{R}} = \{\mathbf{r} \cup \mathbf{r}' \mid \mathbf{r} \in \mathcal{R}, \mathbf{r}' \in \mathcal{R}'\}$. Then, for the range space $\widehat{\mathsf{S}} = (X, \widehat{\mathcal{R}})$, we have that $\dim_{\mathsf{VC}}(\widehat{\mathsf{S}}) = O(\delta + \delta')$.*

Proof: As a warm-up exercise, we prove a somewhat weaker bound here of $O((\delta + \delta') \log(\delta + \delta'))$. The stronger bound follows from Theorem 38.2.6 below. Let B be a set of n points in X that are shattered by $\widehat{\mathsf{S}}$. There are at most $\mathcal{G}_{\delta}(n)$ and $\mathcal{G}_{\delta'}(n)$ different ranges of B in the range sets $\mathcal{R}_{|B}$ and $\mathcal{R}'_{|B}$, respectively, by Lemma 38.2.1. Every subset C of B realized by $\widehat{r} \in \widehat{\mathcal{R}}$ is a union of two subsets $B \cap \mathbf{r}$ and $B \cap \mathbf{r}'$, where $\mathbf{r} \in R$ and $\mathbf{r}' \in \mathcal{R}'$, respectively. Thus, the number of different subsets of B realized by $\widehat{\mathsf{S}}$ is bounded by $\mathcal{G}_{\delta}(n)\mathcal{G}_{\delta'}(n)$. Thus, $2^n \le n^{\delta} n^{\delta'}$, for $\delta, \delta' > 1$. We conclude that $n \le (\delta + \delta') \lg n$, which implies that $n = O((\delta + \delta') \log(\delta + \delta'))$, by Lemma 38.2.4(C). ■

Interestingly, one can prove a considerably more general result with tighter bounds. The required computations are somewhat more painful.

Theorem 38.2.6. *Let $\mathsf{S}_1 = (X, \mathcal{R}^1), \ldots, \mathsf{S}_k = (X, \mathcal{R}^k)$ be range spaces with* VC *dimension $\delta_1, \ldots, \delta_k$, respectively. Next, let $f(\mathbf{r}_1, \ldots, \mathbf{r}_k)$ be a function that maps any k-tuple of sets $\mathbf{r}_1 \in \mathcal{R}^1, \ldots, \mathbf{r}_k \in \mathcal{R}^k$ into a subset of X. Here, the function f is restricted to be defined by a sequence of set operations like complement, intersection and union. Consider the range set*

$$\mathcal{R}' = \left\{ f(\mathbf{r}_1, \ldots, \mathbf{r}_k) \,\middle|\, \mathbf{r}_1 \in \mathcal{R}_1, \ldots, \mathbf{r}_k \in \mathcal{R}_k \right\}$$

and the associated range space $\mathsf{T} = (X, \mathcal{R}')$. Then, the VC *dimension of T is bounded by $O(k\delta \lg k)$, where $\delta = \max_i \delta_i$.*

Proof: Assume a set $Y \subseteq \mathsf{X}$ of size t is being shattered by $\mathcal{R}'$, and observe that

$$\begin{aligned} \left|\mathcal{R}'_{|Y}\right| &\le \left| \left\{ (\mathbf{r}_1, \ldots, \mathbf{r}_k) \,\middle|\, \mathbf{r}_1 \in \mathcal{R}^1_{|Y}, \ldots, \mathbf{r}_k \in \mathcal{R}^k_{|Y} \right\} \right| \le \left|\mathcal{R}^1_{|Y}\right| \cdots \left|\mathcal{R}^k_{|Y}\right| \le \mathcal{G}_{\delta_1}(t) \cdot \mathcal{G}_{\delta_2}(t) \cdots \mathcal{G}_{\delta_k}(t) \\ &\le \left(\mathcal{G}_\delta(t)\right)^k \le \left(2\left(te/\delta\right)^\delta\right)^k, \end{aligned}$$

by Lemma 38.2.1 and Lemma 38.2.2. On the other hand, since Y is being shattered by $\mathcal{R}'$, this implies that $\left|\mathcal{R}'_{|Y}\right| = 2^t$. Thus, we have the inequality $2^t \le \left(2(te/\delta)^\delta\right)^k$, which implies $t \le k(1 + \delta \lg(te/\delta))$. Assume that $t \ge e$ and $\delta \lg(te/\delta) \ge 1$ since otherwise the claim is trivial, and observe that $t \le k(1 + \delta \lg(te/\delta)) \le 3k\delta \lg(t/\delta)$. Setting $x = t/\delta$, we have

$$\frac{t}{\delta} \le 3k \frac{\ln(t/\delta)}{\ln 2} \le 6k \ln \frac{t}{\delta} \implies \frac{x}{\ln x} \le 6k \implies x \le 2 \cdot 6k \ln(6k) \implies x \le 12k \ln(6k),$$

by Lemma 38.2.4(C). We conclude that $t \le 12\delta k \ln(6k)$, as claimed. ■

Corollary 38.2.7. *Let $\mathsf{S} = (X, \mathcal{R})$ and $\mathsf{T} = (X, \mathcal{R}')$ be two range spaces of* VC *dimension δ and δ', respectively, where $\delta, \delta' > 1$. Let $\widehat{\mathcal{R}} = \left\{ \mathbf{r} \cap \mathbf{r}' \,\middle|\, \mathbf{r} \in \mathcal{R}, \mathbf{r}' \in \mathcal{R}' \right\}$. Then, for the range space $\widehat{\mathsf{S}} = (X, \widehat{\mathcal{R}})$, we have that $\dim_{\mathsf{VC}}(\widehat{\mathsf{S}}) = O(\delta + \delta')$.*

Corollary 38.2.8. *Any finite sequence of combining range spaces with finite* VC *dimension (by intersecting, complementing, or taking their union) results in a range space with a finite* VC *dimension.*

38.3. On ε-nets and ε-sampling

38.3.1. ε-nets and ε-samples

Definition 38.3.1 (ε-sample). Let $\mathsf{S} = (\mathsf{X}, \mathcal{R})$ be a range space, and let x be a finite subset of X. For $0 \le \varepsilon \le 1$, a subset $C \subseteq \mathsf{x}$ is an ***ε-sample*** for x if for any range $\mathbf{r} \in \mathcal{R}$, we have

$$\left| \overline{m}(\mathbf{r}) - \overline{s}(\mathbf{r}) \right| \le \varepsilon,$$

where $\overline{m}(\mathbf{r}) = |\mathsf{x} \cap \mathbf{r}| / |\mathsf{x}|$ is the measure of $\mathbf{r}$ (see Definition 38.1.2) and $\overline{s}(\mathbf{r}) = |C \cap \mathbf{r}| / |C|$ is the estimate of $\mathbf{r}$ (see Definition 38.1.3). (Here C might be a multi-set, and as such $|C \cap \mathbf{r}|$ is counted with multiplicity.)

As such, an ε-sample is a subset of the ground set x that "captures" the range space up to an error of ε. Specifically, to estimate the fraction of the ground set covered by a range $\mathbf{r}$, it is sufficient to count the points of C that fall inside $\mathbf{r}$.

If X is a finite set, we will abuse notation slightly and refer to C as an ***ε-sample*** for S.

To see the usage of such a sample, consider $\mathsf{x} = \mathsf{X}$ to be, say, the population of a country (i.e., an element of X is a citizen). A range in $\mathcal{R}$ is the set of all people in the country that answer yes to a question (i.e., would you vote for party Y?, would you buy a bridge from me?, questions like that). An ε-sample of this range space enables us to estimate reliably (up to an error of ε) the answers for all these questions, by just asking the people in the sample.

The natural question of course is how to find such a subset of small (or minimal) size.

Theorem 38.3.2 (ε-sample theorem, [VC71]). *There is a positive constant c such that if $(X, \mathcal{R})$ is any range space with VC dimension at most δ, $\mathsf{x} \subseteq X$ is a finite subset and $\varepsilon, \varphi > 0$, then a random subset $C \subseteq \mathsf{x}$ of cardinality*

$$s = \frac{c}{\varepsilon^2}\left(\delta \log \frac{\delta}{\varepsilon} + \log \frac{1}{\varphi}\right)$$

is an ε-sample for x with probability at least $1 - \varphi$.

(In the above theorem, if $s > |\mathsf{x}|$, then we can just take all of x to be the ε-sample.)

Sometimes it is sufficient to have (hopefully smaller) samples with a weaker property – if a range is "heavy", then there is an element in our sample that is in this range.

Definition 38.3.3 (ε-net). A set $N \subseteq \mathsf{x}$ is an ***ε-net*** for x if for any range $\mathbf{r} \in \mathcal{R}$, if $\overline{m}(\mathbf{r}) \geq \varepsilon$ (i.e., $|\mathbf{r} \cap \mathsf{x}| \geq \varepsilon\,|\mathsf{x}|$), then $\mathbf{r}$ contains at least one point of N (i.e., $\mathbf{r} \cap N \neq \emptyset$).

Theorem 38.3.4 (ε-net theorem, [HW87]). *Let $(X, \mathcal{R})$ be a range space of VC dimension δ, let x be a finite subset of X, and suppose that $0 < \varepsilon \leq 1$ and $\varphi < 1$. Let N be a set obtained by m random independent draws from x, where*

$$m \geq \max\left(\frac{4}{\varepsilon} \lg \frac{4}{\varphi}, \frac{8\delta}{\varepsilon} \lg \frac{16}{\varepsilon}\right). \tag{38.3}$$

Then N is an ε-net for x with probability at least $1 - \varphi$.

(We remind the reader that $\lg = \log_2$.)

The proofs of the above theorems are somewhat involved and we first turn our attention to some applications before presenting the proofs.

Remark 38.3.5. The above two theorems also hold for spaces with shattering dimension at most δ, in which case the sample size is slightly larger. Specifically, for Theorem 38.3.4, the sample size needed is $O\left(\frac{1}{\varepsilon} \lg \frac{1}{\varphi} + \frac{\delta}{\varepsilon} \lg \frac{\delta}{\varepsilon}\right)$.

Remark 38.3.6. The ε-net theorem is a relatively easy consequence (up to constants) of the ε-sample theorem – see bibliographical notes for details.

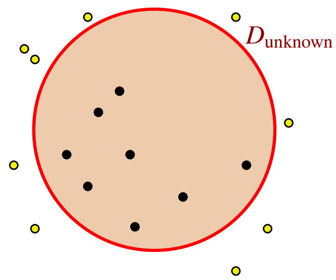
38.3.2. Some applications

We mention two (easy) applications of these theorems, which (hopefully) demonstrate their power.

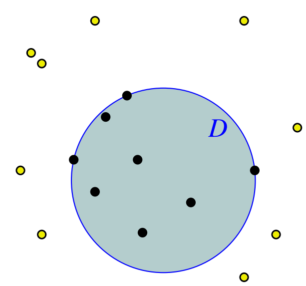
38.3.2.1. Range searching

So, consider a (very large) set of points P in the plane. We would like to be able to quickly decide how many points are included inside a query rectangle. Let us assume that we allow ourselves 1% error. What Theorem 38.3.2 tells us is that there is a subset of *constant size* (that depends only on ε) that can be used to perform this estimation, and it works for *all* query rectangles (we used here the fact that rectangles in the plane have finite VC dimension). In fact, a random sample of this size works with constant probability.

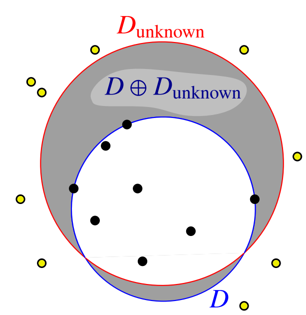
38.3.2.2. Learning a concept



Assume that we have a function f defined in the plane that returns '1' inside an (unknown) disk D_{unknown} and '0' outside it. There is some distribution $\mathcal{D}$ defined over the plane, and we pick points from this distribution. Furthermore, we can compute the function for these labels (i.e., we can compute f for certain values, but it is expensive). For a mystery value $\varepsilon > 0$, to be explained shortly, Theorem 38.3.4 tells us to pick (roughly) $O((1/\varepsilon)\log(1/\varepsilon))$ random points in a sample R from this distribution and to compute the labels for the samples. This is demonstrated in the figure on the right, where black dots are the sample points for which $f(\cdot)$ returned 1.



So, now we have positive examples and negative examples. We would like to find a hypothesis that agrees with all the samples we have and that hopefully is close to the true unknown disk underlying the function f. To this end, compute the smallest disk D that contains the sample labeled by '1' and does not contain any of the '0' points, and let $g : \mathbb{R}^2 \to \{0, 1\}$ be the function g that returns '1' inside the disk and '0' otherwise. We claim that g classifies correctly all but an ε-fraction of the points (i.e., the probability of misclassifying a point picked according to the given distribution is smaller than ε); that is, $\mathbf{Pr}_{\mathsf{p}\in\mathcal{D}}\left[f(\mathsf{p}) \neq g(\mathsf{p})\right] \leq \varepsilon$.



Geometrically, the region where g and f disagree is all the points in the symmetric difference between the two disks. That is, $\mathcal{E} = D \oplus D_{\text{unknown}}$; see the figure on the right.

Thus, consider the range space S having the plane as the ground set and the symmetric difference between any two disks as its ranges. By Corollary 38.2.8, this range space has finite VC dimension. Now, consider the (unknown) disk D' that induces f and the region $\mathbf{r} = D_{\text{unknown}} \oplus D$. Clearly, the learned classifier g returns incorrect answers only for points picked inside $\mathbf{r}$.

Thus, the probability of a mistake in the classification is the measure of $\mathbf{r}$ under the distribution $\mathcal{D}$. So, if $\mathbb{P}_{\mathcal{D}}[\mathbf{r}] > \varepsilon$ (i.e., the probability that a sample point falls inside $\mathbf{r}$), then by the ε-net theorem (i.e., Theorem 38.3.4) the set R is an ε-net for S (ignore for the time being the possibility that the random sample fails to be an ε-net) and as such, R contains a point u inside $\mathbf{r}$. But, it is not possible for g (which classifies correctly all the sampled points of R) to make a mistake on u, a contradiction, because by construction, the range $\mathbf{r}$ is where g misclassifies points. We conclude that $\mathbb{P}_{\mathcal{D}}\left[\mathbf{r}\right] \leq \varepsilon$, as desired.

Little lies. The careful reader might be tearing his or her hair out because of the above description. First, Theorem 38.3.4 might fail, and the above conclusion might not hold. This is of course true, and in real applications one might use a much larger sample to guarantee that the probability of failure is so small that it can be practically ignored. A more serious issue is that Theorem 38.3.4 is defined only for finite sets. Nowhere does it speak about a continuous distribution. Intuitively, one can approximate a continuous distribution to an arbitrary precision using a huge sample and apply the theorem to this sample as our ground set. A formal proof is more tedious and requires extending the proof of Theorem 38.3.4 to continuous distributions. This is straightforward and we will ignore this topic altogether.

38.4. A better bound on the growth function

In this section, we prove Lemma 38.2.2. Since the proof is straightforward but tedious, the reader can safely skip reading this section.

Lemma 38.4.1. *For any positive integer n, the following hold.*

(i) $(1 + 1/n)^n \leq e$.

(ii) $(1 - 1/n)^{n-1} \geq e^{-1}$.

(iii) $n! \geq (n/e)^n$.

(iv) *For any $k \leq n$, we have* $\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{ne}{k}\right)^k$.

Proof: (i) Indeed, $1 + 1/n \leq \exp(1/n)$, since $1 + x \leq e^x$, for $x \geq 0$. As such $(1 + 1/n)^n \leq \exp(n(1/n)) = e$.

(ii) Rewriting the inequality, we have that we need to prove $\left(\frac{n-1}{n}\right)^{n-1} \geq \frac{1}{e}$. This is equivalent to proving $e \geq \left(\frac{n}{n-1}\right)^{n-1} = \left(1 + \frac{1}{n-1}\right)^{n-1}$, which is our friend from (i).

(iii) Indeed,

$$\frac{n^n}{n!} \leq \sum_{i=0}^{\infty} \frac{n^i}{i!} = e^n,$$

by the Taylor expansion of $e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$. This implies that $(n/e)^n \leq n!$, as required.

(iv) Indeed, for any $k \leq n$, we have $\frac{n}{k} \leq \frac{n-1}{k-1}$, as can be easily verified. As such, $\frac{n}{k} \leq \frac{n-i}{k-i}$, for $1 \leq i \leq k-1$. As such,

$$\left(\frac{n}{k}\right)^k \leq \frac{n}{k} \cdot \frac{n-1}{k-1} \cdots \frac{n-k+1}{1} = \binom{n}{k}.$$

As for the other direction, by (iii), we have $\displaystyle \binom{n}{k} \leq \frac{n^k}{k!} \leq \frac{n^k}{\left(\frac{k}{e}\right)^k} = \left(\frac{ne}{k}\right)^k$. ∎

Lemma 38.2.2 restated. *For $n \geq 2\delta$ and $\delta \geq 1$, we have $\left(\frac{n}{\delta}\right)^\delta \leq \mathcal{G}_\delta(n) \leq 2\left(\frac{ne}{\delta}\right)^\delta$, where $\mathcal{G}_\delta(n) = \sum_{i=0}^{\delta} \binom{n}{i}$.*

Proof: Note that by Lemma 38.4.1 (iv), we have $\mathcal{G}_\delta(n) = \sum_{i=0}^{\delta} \binom{n}{i} \leq 1 + \sum_{i=1}^{\delta} \left(\frac{ne}{i}\right)^i$. This series behaves like a geometric series with constant larger than 2, since

$$\left(\frac{ne}{i}\right)^i \bigg/ \left(\frac{ne}{i-1}\right)^{i-1} = \frac{ne}{i}\left(\frac{i-1}{i}\right)^{i-1} = \frac{ne}{i}\left(1 - \frac{1}{i}\right)^{i-1} \geq \frac{ne}{i}\frac{1}{e} = \frac{n}{i} \geq \frac{n}{\delta} \geq 2,$$

by Lemma 38.4.1. As such, this series is bounded by twice the largest element in the series, implying the claim. ∎

38.5. Some required definitions

Definition 38.5.1 (Convex hull). The ***convex hull*** of a set $\mathsf{R} \subseteq \mathbb{R}^d$ is the set of all convex combinations of points of R; that is,

$$\mathcal{CH}(\mathsf{R}) = \left\{ \sum_{i=0}^{m} \alpha_i \mathsf{v}_i \;\middle|\; \forall i \;\; \mathsf{v}_i \in \mathsf{R}, \alpha_i \geq 0, \text{ and } \sum_{j=1}^{m} \alpha_i = 1 \right\}.$$

References

[HW87] D. Haussler and E. Welzl. *ε-nets and simplex range queries*. *Discrete Comput. Geom.*, 2: 127–151, 1987.

[VC71] V. N. Vapnik and A. Y. Chervonenkis. *On the uniform convergence of relative frequencies of events to their probabilities. Theory Probab. Appl.*, 16: 264–280, 1971.