Chapter 3

Approximations and Nets

Here is our scenario for this chapter. We are given a set A of points in $\mathbf{R}^d$ and a range r (which is simply a subset of $\mathbf{R}^d$). A is huge and probably not even known completely; similarly, r is not accessible explicitly. Still, we want to learn something about A and r.

The situation is familiar, definitely, if we don't insist on the geometric setting. For example, let A be the set of all voters, and let r be the set of those voters who plan to vote 'yes' at the next referendum on—I don't know, what. If we want to learn something about r, e.g. the ratio $\frac{|r|}{|A|}$, then we typically sample a subset $S \subseteq A$ of A and see what portion of S lies in r. We want to believe that

$$\frac{|r \cap S|}{|S|} \quad \text{approximates} \quad \frac{|r|}{|A|} \ ,$$

and statistics tells us to what extent this is justified.

No point in a large range. Let us start with a simple exercise. Assume that $|r \cap A| \geq \varepsilon |A|$ for some given ε, $0 < \varepsilon \leq 1$. What is the probability that a set S obtained by drawing s elements uniformly at random from A (with replacement) fails to intersect with r, i.e. $S \cap r = \emptyset$? For $p := \frac{|r \cap A|}{|A|}$ (note $p \geq \varepsilon$) we get[1]

$$\text{prob}(S \cap r = \emptyset) = (1-p)^s \leq (1-\varepsilon)^s \leq e^{-\varepsilon s} \ .$$

[1]We make use of the inequality $1 + x \leq e^x$ for all $x \in \mathbf{R}$.

That is, if $s = \frac{1}{\varepsilon}$ then this probability is at most $e^{-1} \approx 0.368$, and if we choose $s = \lambda\frac{1}{\varepsilon}$, then this probability decreases exponentially with λ: It is at most $e^{-\lambda}$.

For example, if $|A| = 10000$ and $|r \cap A| = 100$ (r contains 1% of the points in A), then a sample of 300 points is disjoint from r with probability at most $e^{-3} \approx 0.05$.

Smallest enclosing ball. Here is a potential use of this for one of our geometric problems. Suppose A is a set of n points in $\mathbf{R}^d$, and we want to compute the smallest enclosing ball of A. In fact, we are willing to accept some mistake, in that, for some given ε, we want a small ball that contains all but at most εn points from A. So let's choose a sample S of $\lambda\frac{1}{\varepsilon}$ points drawn uniformly (with replacement) from A and compute the smallest enclosing ball B of S. Now let $r := \mathbf{R}^d \setminus B$, the complement of B in $\mathbf{R}^d$, play the role of the range in the analysis above. Obviously $r \cap S = \emptyset$, so it is unlikely that $|r \cap A| \geq \varepsilon|A|$, since—if so—the probability of $S \cap r = \emptyset$ was at most $e^{-\lambda}$.

It is important to understand that this was complete nonsense!

For the probabilistic analysis above we have to first choose r and then draw the sample—and not, as done in the smallest ball example, first draw the sample and then choose r *based on the sample*. That cannot possibly work, since we could always choose r simply as the complement $\mathbf{R}^d \setminus S$—then clearly $r \cap S = \emptyset$ and $|r \cap A| \geq \varepsilon|A|$, unless $|S| > (1-\varepsilon)|A|$.

While you hopefully agree on this, you might find the counterargument with $r = \mathbf{R}^d \setminus S$ somewhat artificial, e.g. complements of balls cannot be that selective in 'extracting' points from A. It is exactly the purpose of this chapter to understand to what extent this advocated intuition is justified or not.

3.1 The Formal Set-Up

Range spaces and ε-nets. Here is the formal framework. Let X be a (possibly infinite) set and $\mathcal{R} \subseteq 2^X$. The pair $(X, \mathcal{R})$ is called a *range space*[2], with X its *points* and the elements of $\mathcal{R}$ its *ranges*.

[2] In order to avoid confusion: A range space is nothing else but a set system, sometimes also called hypergraph. It is the context, where we think of X as points and $\mathcal{R}$ as ranges in some geometric ambient space, that suggests the name at hand.

Given $A \subseteq X$, finite, and $\varepsilon \in \mathbf{R}$, $0 \le \varepsilon \le 1$, a subset N of A is called an *ε-net of A (w.r.t. $\mathcal{R}$)* if

$$\text{for all } r \in \mathcal{R}\text{:} \quad |r \cap A| > \varepsilon |A| \quad \Rightarrow \quad r \cap N \neq \emptyset \ .$$

Examples. Typical examples of range spaces in this context are

- $(\mathbf{R}, \mathcal{H}_1)$ with $\mathcal{H}_1 := \{(-\infty, a] \mid a \in \mathbf{R}\} \cup \{[a, \infty) \mid a \in \mathbf{R}\}$ (*half-infinite intervals*), and
- $(\mathbf{R}, \mathcal{I})$ with $\mathcal{I} := \{[a, b] \mid a, b \in \mathbf{R}, a \le b\}$ (*intervals*),

and higher-dimensional counter-parts

- $(\mathbf{R}^d, \mathcal{H}_d)$ with $\mathcal{H}_d$ the closed *halfspaces* in $\mathbf{R}^d$ bounded by hyperplanes,
- $(\mathbf{R}^d, \mathcal{B}_d)$ with $\mathcal{B}_d$ the closed *balls* in $\mathbf{R}^d$,
- $(\mathbf{R}^d, \mathcal{S}_d)$ with $\mathcal{S}_d$ the d-dimensional *simplices* in $\mathbf{R}^d$, and
- $(\mathbf{R}^d, \mathcal{C}_d)$ with $\mathcal{C}_d$ the *convex sets* in $\mathbf{R}^d$.

ε-Nets w.r.t. $(\mathbf{R}, \mathcal{H}_1)$ are particularly simple to obtain. For $A \subseteq \mathbf{R}$, $S := \{\min A, \max A\}$ is an ε-net for every ε—it is even a 0-net. That is, there are ε-nets of size 2, independent from $|A|$ and ε.

The situation gets slightly more interesting for the range space $(\mathbf{R}, \mathcal{I})$ with intervals. Given ε and A with elements

$$a_1 < a_2 < \cdots < a_n \ ,$$

we observe that an ε-net must contain at least one element from any contiguous sequence $\{a_i, a_{i+1}, \ldots, a_{i+k-1}\}$ of $k > \varepsilon n$ (i.e. $k \ge \lfloor \varepsilon n \rfloor + 1$) elements in A. In fact, this is a necessary and sufficient condition for ε-nets w.r.t. intervals. Hence,

$$\{a_{\lfloor \varepsilon n \rfloor + 1}, a_{2\lfloor \varepsilon n \rfloor + 2}, \ldots\}$$

is an ε-net of size[3] $\left\lfloor \frac{n}{\lfloor \varepsilon n \rfloor + 1} \right\rfloor \le \left\lceil \frac{1}{\varepsilon} \right\rceil - 1$. So while the size of the net depends now on ε, it is still independent of $|A|$.

[3] The number L of elements in the set is the largest ℓ such that $\ell(\lfloor \varepsilon n \rfloor + 1) \le n$, hence $L = \left\lfloor \frac{n}{\lfloor \varepsilon n \rfloor + 1} \right\rfloor$. Since $\lfloor \varepsilon n \rfloor + 1 > \varepsilon n$, we have $\frac{n}{\lfloor \varepsilon n \rfloor + 1} < \frac{1}{\varepsilon}$, and so $L < \frac{1}{\varepsilon}$, i.e. $L \le \left\lceil \frac{1}{\varepsilon} \right\rceil - 1$.

Either almost all is needed or a constant suffices. Let us reveal the spectrum of possibilities right away, although its proof will have to await some preparatory steps.

Theorem 3.1.1 *Let $(X, \mathcal{R})$ be an infinite range space. Then one of the following two statements holds.*

(1) For every $n \in \mathbf{N}$ there is a set $A_n \subseteq X$ with $|A_n| = n$ such that for every ε, $0 \leq \varepsilon \leq 1$, an ε-net must have size at least $(1-\varepsilon)n$.

(2) There is a constant d depending on $(X, \mathcal{R})$, such that for every finite $A \subseteq X$ and every ε, $0 < \varepsilon \leq 1$, there is an ε-net of A w.r.t. $\mathcal{R}$ of size at most $O(\frac{d}{\varepsilon}\log\frac{1}{\varepsilon})$ (independent of the size of A).

That is, either we have always ε-nets of size independent of $|A|$, or we have to do the trivial thing, namely choosing all but εn points for an ε-net.

Obviously, the range spaces $(\mathbf{R}, \mathcal{H}_1)$ and $(\mathbf{R}, \mathcal{I})$ fall into category (2) of the theorem.

For an example for (1), consider $(\mathbf{R}^2, \mathcal{C}_2)$. For any $n \in \mathbf{N}$, let A_n be a set on n points on a circle. For every $N \subseteq A_n$ there is a range $r \in \mathcal{C}_2$, namely the convex hull of $A_n \setminus N$, such that $A_n \cap r = A_n \setminus N$ (hence, $r \cap N = \emptyset$). Therefore, $N \subseteq A_n$ cannot be an ε-net of A_n w.r.t. $\mathcal{C}_2$ if $|A_n \setminus N| = n - |N| > \varepsilon n$. Consequently, an ε-net must contain at least $n - \varepsilon n = (1-\varepsilon)n$ points.[4]

So what distinguishes $(\mathbf{R}^2, \mathcal{C}_2)$ from $(\mathbf{R}, \mathcal{H}_1)$ and $(\mathbf{R}, \mathcal{I})$? And which of the two cases applies to the many other range spaces we have listed above? Will all of this eventually tell us something about our attempt of computing a small ball covering all but at most εn out of n given points? This and more should be clear by the end of this chapter.

What makes the difference: VC-dimension. Given a range space $(X, \mathcal{R})$ and $A \subseteq X$, we let

$$\mathcal{R}\big|_A := \{r \cap A \,|\, r \in \mathcal{R}\} ,$$

the *projection of $\mathcal{R}$ to A*.

Note that, for A a set of n points on a circle in the plane, $\mathcal{C}_2\big|_A = 2^A$; we get every subset of A by an intersection with a convex set. That is $|\mathcal{C}_2\big|_A| = 2^n$.

[4] If we were satisfied with any abstract example for category (1), we could have taken $(X, 2^X)$ for any infinite set X.

For A a set of n points in $\mathbf{R}$, we can easily see that[5] $|\mathcal{I}\big|_A| = \binom{n+1}{2} + 1 = O(n^2)$. (What about $|\mathcal{H}_1\big|_A|$?) Now comes the crucial definition.

Given a range space $(X, \mathcal{R})$, a subset A of X is *shattered by* $\mathcal{R}$ if $\mathcal{R}\big|_A = 2^A$. The *VC-dimension*[6] *of* $(X, \mathcal{R})$, $\mathrm{VCdim}(X, \mathcal{R})$, is the cardinality (possibly infinite) of the largest subset of X that is shattered by $\mathcal{R}$. If no set is shattered (i.e. not even the empty set which means that $\mathcal{R}$ is empty), we set the VC-dimension to -1.

We had just convinced ourselves that $(\mathbf{R}^2, \mathcal{C}_2)$ has arbitrarily large sets that can be shattered. Therefore, $\mathrm{VCdim}(\mathbf{R}^2, \mathcal{C}_2) = \infty$.

Consider now $(\mathbf{R}, \mathcal{I})$. Two points $A = \{a, b\}$ can be shattered, since for each of the 4 subsets, $\emptyset$, $\{a\}$, $\{b\}$, and $\{a, b\}$, of A, there is an interval that generates that subset by intersection with A. However, for $A = \{a, b, c\}$ with $a < b < c$ there is no interval that contains a and c but not b. Hence, $\mathrm{VCdim}(\mathbf{R}, \mathcal{I}) = 2$. (What is $\mathrm{VCdim}(\mathbf{R}, \mathcal{H}_1)$?)

The size of projections for finite VC-dimension.

Lemma 3.1.2 (Sauer's Lemma) *If $(X, \mathcal{R})$ is a range space of finite VC-dimension at most d, then*

$$|\mathcal{R}\big|_A| \leq \Phi_d(n) := \sum_{i=0}^{d} \binom{n}{i}$$

for all $A \subseteq X$ with $|A| = n$.

ProofFirst let us observe that $\Phi : \mathbf{N}_0 \cup \{-1\} \times \mathbf{N}_0 \to \mathbf{N}_0$ is defined by the recurrence[7]

$$\Phi_d(n) = \begin{cases} 0 & d = -1, \\ 1 & n = 0 \text{ and } d \geq 0, \text{ and} \\ \Phi_d(n-1) + \Phi_{d-1}(n-1) & \text{otherwise.} \end{cases}$$

---

[5]Given A as $a_1 < a_2 \cdots < a_n$ we can choose another $n+1$ points b_i, $0 \leq i \leq n$, such that

$$b_0 < a_1 < b_1 < a_2 < b_2 < \cdots b_{n-1} < a_n < b_n \ .$$

Each nonempty intersection of A with an interval can be uniquely written as $A \cap [b_i, b_j]$ for $0 \leq i < j \leq n$. This gives $\binom{n+1}{2}$ plus one for the empty set.

[6]'VC' in honor of the Russian statisticians V. N. Vapnik and A. Ya. Chervonenkis, who discovered the crucial role of this parameter in the late sixties.

[7]We recall that the binomial coefficients $\binom{n}{k}$ (with $k, n \in \mathbf{N}_0$) satisfy the recurrence $\binom{n}{k} = 0$ if $n < k$, $\binom{n}{0} = 1$, and $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$.

Second, we note that the VC-dimension cannot increase by passing from $(X, \mathcal{R})$ to a projection (A, R), $R := \mathcal{R}\big|_A$. Hence, it suffices to consider the finite range space (A, R)—which is of VC-dimension at most d—and show $|R| \leq \Phi_d(n)$ (since Φ is monotone in d).

Now we proceed to a proof by induction of this inequality. If $A = \emptyset$ or $R = \emptyset$ the statement is trivial. Otherwise, we consider the two 'derived' range spaces for some fixed $x \in A$:

$$(A \setminus \{x\}, R - x), \qquad \text{with } R - x := \{r \setminus \{x\} \,|\, r \in R\}$$

(note $R - x = R\Big|_{A \setminus \{x\}}$) and

$$(A \setminus \{x\}, R^{(x)}), \qquad \text{with } R^{(x)} := \{r \in R \,|\, x \notin r, r \cup \{x\} \in R\}.$$

Observe that the ranges in $R^{(x)}$ are exactly those ranges in $R - x$ that have two preimages under the map

$$R \ni \quad r \quad \mapsto \quad r \setminus \{x\} \quad \in R - x \ ,$$

all other ranges have a unique preimage. Consequently,

$$|R| = |R - x| + |R^{(x)}| \ .$$

We have $|R - x| \leq \Phi_d(n - 1)$. If $A' \subseteq A \setminus \{x\}$ is shattered by $R^{(x)}$, then $A' \cup \{x\}$ is shattered by R. Hence, $(A \setminus \{x\}, R^{(x)})$ has VC-dimension at most $d - 1$ and $|R^{(x)}| \leq \Phi_{d-1}(n - 1)$. Summing up, it follows that

$$|R| \leq \Phi_d(n - 1) + \Phi_{d-1}(n - 1) = \Phi_d(n)$$

which yields the assertion of the lemma. $\Box$

In order to see that the bound given in the lemma is tight, consider the range space

$$(X, \bigcup_{i=0}^{d} \binom{X}{i}) \ .$$

Obviously, a set of more than d elements cannot be shattered (hence, the VC-dimension is at most d), and for any finite $A \subseteq X$, the projection of the ranges to A is $\bigcup_{i=0}^{d} \binom{A}{d}$—with cardinality $\Phi_d(|A|)$.

We note that a rough, but for our purposes good enough estimate for Φ is given by[8]

$$\Phi_d(n) \leq n^d \quad \text{for } d \geq 2.$$

We have seen now that the maximum possible size of projections either grows exponentially (2^n in case of infinite VC-dimension) or it is bounded by a polynomial ($O(n^d)$ in case of finite VC-dimension d). The latter is the key to the existence of small ε-nets. Before shedding light on this, let us better understand when VC-dimension is finite.

3.2 VC-dimension of Halfspaces and Balls

Halfspaces. Let us investigate the VC-dimension of $(\mathbf{R}^2, \mathcal{H}_2)$. It is easily seen that three points in the plane can be shattered by halfplanes, as long a they do not lie on a common line. Hence, the VC-dimension is at least 3. Now consider 4 points. If three of them lie on a common line, there is no way to separate the middle point on this line from the other two by a halfplane. So let us assume that no three points lie on a line. Either three of them are vertices of a triangle that contains the fourth point—then we cannot possibly separate the fourth point from the remaining three points by a halfplane. Or the four points are vertices of a convex quadrilateral—then there is no way of separating the endpoints from a diagonal from the other two points. Consequently, four points cannot be shattered, and $\mathrm{VCdim}(\mathbf{R}^2, \mathcal{H}_2) = 3$ is established.

The above argument gets tedious in higher dimensions, if it works in a rigorous way at all. Fortunately, we can employ a classic.[9]

Lemma 3.2.1 (Radon's Lemma) *Let $d \in \mathbf{N}$. Every set A of $n \geq d + 2$ points in $\mathbf{R}^d$ can be partitioned as $A = A_1 \dot\cup A_2$ such that* $\mathrm{conv}(A_1) \cap \mathrm{conv}(A_2) \neq \emptyset$.

---

[8] A better estimate, at least for $d \geq 3$, is given by $\Phi_d(n) < \left(\frac{e\,n}{d}\right)^d$ for all $n, d \in \mathbf{N}$, $d \leq n$.

[9] Radon's Lemma forms a classical triumvirate with Carathéodory's Lemma and Helly's Lemma. You have heard about them in Section 1.5. Carathéodory's Lemma states that if point p lies in the convex hull of a set $A \subseteq \mathbf{R}^d$, then there is a subset $A' \subseteq A$ with $|A'| \leq d + 1$ and $p \in \mathrm{conv}(A')$. These three lemmas are not difficult to prove, but more often than not they constitute the key to proofs in convexity. As such, they are typical lemmas, although they are often referred to as theorems.

ProofFor $n \geq d+2$, n points $\{p_1, p_2, \ldots, p_n\}$ in $\mathbf{R}^d$ are affinely dependent, which—by definition—means that there are real coefficients λ_i, $i \in \{1, 2, \ldots, n\}$, not all 0, with

$$\sum_{i=1}^{n} \lambda_i = 0 \text{ and } \sum_{i=1}^{n} \lambda_i p_i = \mathbf{0} \ .$$

Now let $I_1 := \{i \,|\, \lambda_i > 0\}$ and $I_2 := \{i \,|\, \lambda_i \leq 0\}$, which both have to be non-empty. We have $\sum_{i \in I_1} \lambda_i = -\sum_{i \in I_2} \lambda_i$, and we let λ denote this value (which has to be positive). Now

$$p := \frac{1}{\lambda} \sum_{i \in I_1} \lambda_i p_i = -\frac{1}{\lambda} \sum_{i \in I_2} \lambda_i p_i$$

denotes a point that lies in the convex hulls of $A_1 := \{p_i \,|\, i \in I_1\}$ and $A_2 := \{p_i \,|\, i \in I_2\}$, respectively. Therefore, A_1 and A_2 provide a partition as required in the lemma. □

We get, as an easy implication, that a set A of at least $d+2$ points in $\mathbf{R}^d$ cannot be shattered by halfspaces. Indeed, let $A_1 \dot\cup A_2$ be a partition as guaranteed by Randon's Lemma. Now every halfspace containing A_1 must contain at least one point of A_2, hence $h \cap A = A_1$ is impossible for a halfspace h and thus A is not shattered by $\mathcal{H}_d$. Moreover, it is easily seen that the vertex set of a d-dimensional simplex (there are $d+1$ vertices) can be shattered by halfspaces (each subset of the vertices forms a face of the simplex and can thus be separated from the rest by a hyperplane). We summarize that

$$\mathrm{VCdim}(\mathbf{R}^d, \mathcal{H}_d) = d + 1 \ .$$

Let us now consider the range space $(\mathbf{R}^d, \check{\mathcal{H}}_d)$, where $\check{\mathcal{H}}_d$ denotes the set of all closed halfspaces below non-vertical hyperplanes—we call these *lower halfspaces*. Since $\check{\mathcal{H}}_d \subseteq \mathcal{H}_d$, the VC-dimension of $(\mathbf{R}^d, \check{\mathcal{H}}_d)$ is at most $d+1$, but, in fact, it not too difficult to show

$$\mathrm{VCdim}(\mathbf{R}^d, \check{\mathcal{H}}_d) = d \ . \tag{3.1}$$

(Check this claim at least for $d = 2$.) This range space is a geometric example where the bound of Sauer's Lemma is attained. Indeed, for any set A of n points in $\mathbf{R}^d$ in general position[10], it can be shown that

$$|\check{\mathcal{H}}_d\big|_A| = \Phi_d(n) \ .$$

[10]No $i+2$ on a common i-flat for $i \in \{1, 2, \ldots, d-1\}$; in particular, no $d+1$ points on a common hyperplane.

Balls. It is easy to convince oneself that the VC-dimension of disks in the plane is 3: Three points not on a line can be shattered and four points cannot. Obviously not, if one of the points is in the convex hull of the other, and for four vertices of a convex quadrilateral, it is not possible for both diagonals to be separated from the endpoints of the respective other diagonal by a circle (if you try to draw a picture of this, you see that you get two circles that intersect four times, which we know is not true).

A more rigorous argument which works in all dimensions is looming with the help of (3.1), if we employ the following transformation called *lifting map*.

$$\begin{array}{rcl} \mathbf{R}^d & \longrightarrow & \mathbf{R}^{d+1} \\ (x_1, x_2, \ldots, x_d) = p & \mapsto & p^\wedge = (x_1, x_2, \ldots, x_d,\, {x_1}^2 + {x_2}^2 + \cdots + {x_d}^2) \end{array}$$

(For a geometric interpretation, this is a vertical projection of $\mathbf{R}^d$ to the unit paraboloid $x_{d+1} = {x_1}^2 + {x_2}^2 + \cdots + {x_d}^2$ in $\mathbf{R}^{d+1}$.) The remarkable property of this transformation is that it maps balls in $\mathbf{R}^d$ to halfspaces in $\mathbf{R}^{d+1}$ in the following sense.

Consider a ball $B_d(c, \rho)$ ($c = (c_1, c_2, \ldots, c_d) \in \mathbf{R}^d$ the center, and $\rho \in \mathbf{R}^+$ the radius). A point $p = (x_1, x_2, \ldots, x_d)$ lies in this ball iff

$$\begin{aligned} & \sum_{i=1}^{d} (x_i - c_i)^2 \le \rho^2 \Leftrightarrow \sum_{i=1}^{d} ({x_i}^2 - 2x_i c_i + {c_i}^2) \le \rho^2 \\ \Leftrightarrow \quad & \left(\sum_{i=1}^{d} (-2c_i) x_i \right) + ({x_1}^2 + {x_2}^2 + \cdots + {x_d}^2) \le \rho^2 - \sum_{i=1}^{d} {c_i}^2 \ ; \end{aligned}$$

this equivalently means that $p^\wedge$ lies below the non-vertical hyperplane (in $\mathbf{R}^{d+1}$)

$$\begin{aligned} & h = h(c, \rho) = \{x \in \mathbf{R}^{d+1} \mid \sum_{i=1}^{d+1} h_i x_i = h_{d+2}\} \ \text{ with} \\ & (h_1, h_2, \ldots, h_d, h_{d+1}, h_{d+2}) = \left((-2c_1), (-2c_2), \ldots, (-2c_d), 1, \rho^2 - \sum_{i=1}^{d} {c_i}^2 \right) \ . \end{aligned}$$

It follows that a set $A \subseteq \mathbf{R}^d$ is shattered by $\mathcal{B}_d$ (the set of closed balls in $\mathbf{R}_d$) iff $A^\wedge := \{p^\wedge \mid p \in A\}$ is shattered by $\check{\mathcal{H}}_{d+1}$. Assuming (3.1), this readily yields

$$\mathrm{VCdim}(\mathbf{R}^d, \mathcal{B}_d) = d + 1 \ .$$

The lifting map we have employed here is a special case of a more general paradigm called *linearization* which maps non-linear conditions to linear conditions in higher dimensions (see also the coming chapter on support vector machines).

We have clarified the VC-dimension for all examples of range spaces we have listed in the beginning of Section 3.1 except for the one involving simplices. Before we elaborate on this, let us prove a first bound on the size of ε-nets when the VC-dimension is finite.

3.3 Small ε-Nets, an Easy Warm-up Version

Theorem 3.3.1 *Let $n, d \in \mathbf{N}$, $d \geq 2$, $\varepsilon \in \mathbf{R}^+$. Let $(X, \mathcal{R})$ be a range space of VC-dimension d. If $A \subseteq X$ with $|A| = n$, then there exists an ε-net N of A w.r.t. $\mathcal{R}$ with $|N| \leq \lceil \frac{d \ln n}{\varepsilon} \rceil$.*

ProofWe restrict our attention to the finite projected range space (A, R), $R := \mathcal{R}\big|_A$, for which we know $|R| \leq \Phi_d(n) \leq n^d$. It suffices to show that there is a set $N \subseteq A$ with $|N| \leq \frac{d \ln n}{\varepsilon}$ which contains an element from each $r \in R_\varepsilon := \{r \in R \,|\, |r| > \varepsilon n\}$.

Suppose, for some $s \in \mathbf{N}$ (to be determined), we let N be a set obtained by drawing s elements uniformly at random from A with replacement. For each $r \in R_\varepsilon$, we know that $\text{prob}(r \cap N = \emptyset) < (1 - \varepsilon)^s \leq e^{-\varepsilon s}$. Therefore,

$$\begin{aligned} \text{prob}(N \text{ is not } \varepsilon\text{-net of } A) &= \text{prob}(\exists r \in R_\varepsilon : r \cap N = \emptyset) \\ &= \text{prob}(\bigvee_{r \in R_\varepsilon} (r \cap N = \emptyset)) \\ &\leq \sum_{r \in R_\varepsilon} \text{prob}(r \cap N = \emptyset) < |R_\varepsilon| e^{-\varepsilon s} \leq n^d e^{-\varepsilon s} \ . \end{aligned}$$

It follows that if s is chosen so that $n^d e^{-\varepsilon s} \leq 1$, then $\text{prob}(N \text{ is not } \varepsilon\text{-net of } A) < 1$ and there remains a positive probability for the event that N is an ε-net of A. Now

$$n^d e^{-\varepsilon s} \leq 1 \quad \Leftrightarrow \quad n^d \leq e^{\varepsilon s} \quad \Leftrightarrow \quad d \ln n \leq \varepsilon s.$$

That is, for $s = \lceil \frac{d \ln n}{\varepsilon} \rceil$, the probability of obtaining an ε-net is positive, and therefore an ε-net of that size has to exist.[11] □

[11] This line of argument "If an experiment produces a certain object with positive prob-

If we are willing to invest a little more in the size of the random sample N, then the probability of being an ε-net grows dramatically. More specifically, for $s = \lceil \frac{d \ln n + \lambda}{\varepsilon} \rceil$, we have

$$n^d e^{-\varepsilon s} \leq n^d e^{-d \ln n - \lambda} = e^{-\lambda} ,$$

and, therefore, a sample of that size is an ε-net with probability at least $1 - e^{-\lambda}$.

We realize that we need $\frac{d \ln n}{\varepsilon}$ sample size to compensate for the (at most) n^d subsets of A which we have to hit—it suffices to ensure positive success probability. The extra $\frac{\lambda}{\varepsilon}$ allows us to boost the success probability.

Also note that if A were shattered by $\mathcal{R}$, then $R = 2^A$ and $|R| = 2^n$. In the line of the proof above, that would require us to choose s to be roughly $\frac{n \ln 2}{\varepsilon}$, a useless estimate which even exceeds n unless ε is large (at least $\ln 2 \approx 0.69$).

---

ability, then it has to exist", as trivial as it is, admittedly needs some time to digest. It is called *The Probabilistic Method*, and was used and developed to an amazing extent by the famous Hungarian mathematician Paul Erdős starting in the thirties.