

# Chapter 2

## $\varepsilon$ -nets and VC-dimension

### 2.1 Defining $\varepsilon$ -nets

A remarkably useful tool in modern computational geometry is the notion of  $\varepsilon$ -nets, which is used to power many randomized geometric constructions, as we shall see below. To define  $\varepsilon$ -nets, consider a set  $X$  with a probability measure  $\mu$  on  $X$ . Most often,  $\mu$  will be the uniform measure on the set  $X$ , which may be finite or infinite. Let  $F$  be a collection of subsets of  $X$ . Together, the pair  $(X, F)$  is called a *set system*. (When  $X$  is finite, a set system on  $X$  is sometimes referred to as a *hypergraph*.) Given  $0 \leq \varepsilon \leq 1$ , an  $\varepsilon$ -net for  $(X, F)$  is a subset of  $X$  that “hits” all the heavy sets in  $F$ , namely all the set in  $F$  that have measure at least  $\varepsilon$ . More formally,

**Definition 2.1.1.** *Given a set system  $(X, F)$  as above, a subset  $N \subseteq X$  is called an  $\varepsilon$ -net for  $(X, F)$  if  $N \cap S \neq \emptyset$  for all  $S \in F$  with  $\mu(S) \geq \varepsilon$ .*

For example, given a finite set system  $(X, F)$  with  $|X| = n$ ,  $\mu(S) = \frac{|S|}{n}$  for any  $S \in 2^X$ , and  $r \in \mathbb{N}^+$ , a  $(1/r)$ -net for  $(X, F)$  is a subset  $N$  that has a nonempty intersection with all sets of  $F$  that have at least  $n/r$  elements.

We will be interested in finding small  $\varepsilon$ -nets. Specifically, for constant  $\varepsilon$ , we will want to find  $\varepsilon$ -nets of constant size (!), namely, size that depends on  $\varepsilon$  but not on the size of the set system. This is not always possible, as easy constructions can suggest, but quite often it is. To describe an important class of set systems that admit small  $\varepsilon$ -nets we need the notion of VC-dimension.

### 2.2 Defining VC-dimension

The *Vapnik-Chervonenkis dimension*, or the VC-dimension is a numerical parameter of a set system that quantifies how “well behaved”, in a certain sense, the system is.

Given a set system  $(X, F)$  and a subset  $Y \subseteq X$ , the *restriction* of  $F$  on  $Y$  is the set  $F|_Y = \{S \cap Y : S \in F\}$ .

**Definition 2.2.1.** *Given a set system  $(X, F)$  as above, a subset  $A \subseteq X$  is said to be shattered by  $F$  if each of the subsets of  $A$  arises as an intersection  $A \cap S$  for some*

$S \in F$ , that is, if  $F|_Y = 2^A$ . The VC-dimension  $\dim(F)$  of  $F$  is the size of the largest subset of  $X$  that is shattered by  $F$ .

To see that the VC-dimension concept is not vacuous, consider two examples. Let  $X$  be the Euclidean plane  $\mathbb{R}^2$ ,  $F_1$  be the set of all convex polygons in the plane, and  $F_2$  be the set of halfplanes. The set system  $(X, F_1)$  has VC-dimension  $\infty$ . Indeed, consider an arbitrarily large finite set  $A \subseteq \mathbb{R}^2$  in convex position; any subset  $A' \in A$  can arise as an intersection of  $A$  with a convex polygon—just take  $\text{conv}(A')$ .

On the other hand, the set system  $(X, F_2)$  has the bounded VC-dimension 3. In fact, something more general holds.

**Proposition 2.2.2.** *Let  $R[x_1, x_2, \dots, x_d]_1$  denote the set of all linear functions in  $d$  variables, and let*

$$P_{d,1} = \{ \{x \in \mathbb{R}^d : p(x) \geq 0\} : p \in R[x_1, x_2, \dots, x_d]_1 \}.$$

*The VC-dimension of the set system  $(\mathbb{R}^d, P_{d,1})$  is  $d + 1$ .*

*Proof.* Consider a set  $A$  of  $d + 1$  points in  $\mathbb{R}^d$  in general position. It is easy to see that any subset of  $A$  can be separated from the rest by a halfspace. On the other hand, given a collection  $B$  of at least  $d + 2$  points in  $\mathbb{R}^d$ , Radon's theorem asserts that there exist two disjoint subsets  $B_1, B_2 \subset B$  such that

$$\text{conv}(B_1) \cap \text{conv}(B_2) \neq \emptyset.$$

By convexity, these two subsets cannot be separated by a hyperplane, so neither  $B_1$  nor  $B_2$  can arise as an intersection of  $B$  with a halfspace.  $\square$

## 2.3 The existence of small $\varepsilon$ -nets

The next theorem of Haussler and Welzl is the reason we went through the trouble of introducing  $\varepsilon$ -nets and VC-dimension. It says that for set systems with bounded VC-dimension, small  $\varepsilon$ -nets exist, and can in fact be found by simple random sampling.

**Theorem 2.3.1** ( $\varepsilon$ -net theorem). *Given a set system  $(X, F)$  with  $\dim(F) \leq d$ , such that  $d \geq 2$  and  $r \geq 2$  is a parameter, there exists a  $(1/r)$ -net for  $(X, F)$  of size at most  $Cdr \ln r$ , where  $C$  is an absolute constant.*

To prove this theorem we first need to establish two lemmas. The first concerns a so-called *shatter function* of a set system  $(X, F)$ . This is the function

$$\pi_F(m) = \max_{Y \subseteq X, |Y|=m} |F|_Y|.$$

To obtain intuition concerning this definition, note that  $\pi_F(|X|)$  is simply the size of  $F$ . The next lemma then states that the size of a set system with bounded VC-dimension is itself bounded.

**Lemma 2.3.2** (Shatter function lemma). *Given a set system  $(X, F)$  with  $\dim(F) \leq d$ , and any  $1 \leq m \leq |X|$ ,*

$$\pi_F(m) \leq \sum_{i=0}^d \binom{m}{i}.$$

This lemma implies that the size of  $(X, F)$  is  $O(n^d)$ , where  $n = |X|$ . In fact, a tighter bound can be obtained using estimates for binomial coefficients:

$$\pi_F(m) \leq \left(\frac{em}{d}\right)^d.$$

*Proof.* Note that for any  $Y \subseteq X$ , the VC-dimension of  $(Y, F|_Y)$  is at most  $d$ . (Any subset of  $Y$  that is shattered by  $F|_Y$  is also a subset of  $X$  that is shattered by  $F$ .) This implies that it suffices to show that the size of  $(X, F)$  is at most  $\sum_{i=0}^d \binom{n}{i}$ , where  $n = |X|$ . To this end, we use induction on  $d$ , and for a fixed  $d$  we do induction on  $n$ .

For the induction step, fix some  $x \in X$  and consider the set system  $(X \setminus \{x\}, F_1)$ , for  $F_1 = F|_{X \setminus \{x\}}$ . By the induction hypothesis,  $|F_1| \leq \sum_{i=0}^d \binom{n-1}{i}$ . Any two distinct sets  $A_1, A_2 \in F$  for which  $A_1 \cap F_1 \neq A_2 \cap F_1$  are counted as distinct sets of  $F_1$  and are thus considered in  $|F_1|$ . The only pairs of distinct sets of  $F$  that are not thus counted are pairs of sets  $A_1, A_2 \in F$ , such that  $A_1 \subseteq X \setminus \{x\}$  and  $A_2 = A_1 \cup \{x\}$ .

Consider the set system  $(X \setminus \{x\}, F_2)$ , defined as  $F_2 = \{A \in F_1 : A \in F \text{ and } A \cup \{x\} \in F\}$ . The discussion above implies  $|F| = |F_1| + |F_2|$ . Observe now that  $\dim(F_2) \leq d - 1$ , since if  $A \subseteq X \setminus \{x\}$  is shattered by  $F_2$  then  $A \cup \{x\}$  is shattered by  $F$ . By the induction hypothesis,  $|F_2| \leq \sum_{i=0}^{d-1} \binom{n-1}{i}$ . Therefore,

$$\begin{aligned} |F| &\leq \sum_{i=0}^{d-1} \binom{n-1}{i} + \sum_{i=0}^d \binom{n-1}{i} = 1 + \sum_{i=1}^d \left( \binom{n-1}{i-1} + \binom{n-1}{i} \right) = \\ &= \binom{n}{0} + \sum_{i=1}^d \binom{n}{i} = \sum_{i=0}^d \binom{n}{i}. \end{aligned}$$

□

We still need one more lemma before we commence the proof of the  $\varepsilon$ -net theorem.

**Lemma 2.3.3.** *Let  $X = \sum_{i=1}^n X_i$ , where the  $X_i$  are independent random variables, each attaining the value 1 with probability  $p$  and the value 0 with probability  $1 - p$ . Then*

$$P \left[ X \geq \frac{1}{2}np \right] \geq \frac{1}{2}$$

*provided that  $np \geq 8$ .*

*Proof.* The estimate in the lemma is very weak and much stronger ones can be obtained. This one is a consequence of Chebyshev's inequality that states that  $P[|X - E[X]| \geq k\sigma] \leq \frac{1}{k^2}$ . In our case,  $E[X] = np$  and  $\sigma \leq \sqrt{np}$ . We have

$$\begin{aligned} P \left[ X \geq \frac{1}{2}np \right] &= 1 - P \left[ X < \frac{1}{2}np \right] = 1 - P \left[ X - E[X] < -\frac{1}{2}np \right] \leq \\ 1 - P \left[ |X - E[X]| \geq \frac{1}{2}np \right] &\leq 1 - P \left[ |X - E[X]| \geq \frac{\sqrt{np}}{2} \sigma \right] \geq 1 - \frac{4}{np} \geq \frac{1}{2}. \end{aligned}$$

□

*Proof of the  $\varepsilon$ -net theorem.* First of all, note that we can assume that all sets of  $F$  have measure at least  $1/r$ , as the small sets do not change anything. Put  $s = Cdr \ln r$  and assume for the sake of clarity that it is an integer—the proof can be easily modified to remove this assumption. Let  $N$  denote a random sample of  $s$  elements from  $X$ , drawn with repetitions according to the probability distribution on  $X$ . (So  $N$  is regarded as a multiset, since it can have repeated elements.) Our goal is to show that  $N$  is a  $(1/r)$ -net with a positive probability—once we show that, it is easy to make this probability arbitrarily close to 1 by increasing the constant  $C$ . So let  $E_0$  be the “bad” event that  $N$  fails to be a  $(1/r)$ -net, namely, that there exists  $T \in F$  for which  $N \cap T = \emptyset$ .

We bound  $P[E_0]$  away from 1 by a “magic trick” that relates  $P[E_0]$  to the probability of another event, which we call  $E_1$ . Draw a second random sample of  $s$  elements from  $X$  and denote the resulting sequence by  $M$ .  $E_1$  is the event that there exists a set  $T \in F$  that is missed by  $N$  but is “heavily hit” by  $M$ . Specifically, put  $k = s/2r$ , again assuming that  $k$  is an integer, and let  $E_1$  be the event that there exists  $T \in F$  with  $N \cap T = \emptyset$  and  $|M \cap T| \geq k$ .

Now,  $P[E_1] \leq P[E_0]$ , since  $E_1$  required  $E_0$ . We show that  $P[E_1] \geq \frac{1}{2}P[E_0]$ . For this, we first bound  $P[E_1|N]$ . If  $N$  is a  $(1/r)$ -net, then  $P[E_1] = P[E_0] = 0$ . Otherwise, fix a set  $T_N$  for which  $N \cap T = \emptyset$  and note that  $P[E_1|N] \geq P[|M \cap T_N| \geq k]$ . Now, the quantity  $|M \cap T_N|$  can be viewed as a sum of  $s$  Bernoulli trials, each having success probability  $1/r$ . Thus, by the above lemma, the probability that this sum is at least  $k = s/2r = \frac{1}{2}\frac{s}{r}$  is at least  $\frac{1}{2}$ , namely if  $N$  is not a  $(1/r)$ -net then  $P[E_1|N] \geq \frac{1}{2}$ . In general then,  $P[E_1|N] \geq \frac{1}{2}P[E_0|N]$  for all  $N$ , which implies  $P[E_1] \geq \frac{1}{2}P[E_0]$ .

Now it’s time for the second part of the magic trick, which involves showing that  $P[E_1]$  is in fact small, strictly smaller than  $\frac{1}{2}$ . Given the above inequality, this will imply that  $P[E_0]$  is less than 1.

Take a random sample  $A$  of  $2s$  elements from  $X$ , and then choose  $s$  elements of  $A$  uniformly at random (without repetitions this time), denote this collection by  $N$ , and denote the rest of  $A$  by  $M$ . These  $N$  and  $M$  have the same distribution as above. Let’s analyze  $P[E_1|A]$ .

Let  $A$  be fixed. Fix a particular  $T \in F$  and consider the probability that “ $E_1$  holds for this particular  $T$ ”, namely,  $P_T = P[N \cap T = \emptyset, |M \cap T| \geq k|A]$ . If  $|A \cap T| < k$  then  $P_T = 0$ . Otherwise we use the fact that  $P_T \leq P[N \cap T = \emptyset|A]$ . This is the probability that a random sample of the  $s$  elements of  $N$  out of the  $2s$  elements of  $A$  avoids the at least  $k$  elements of  $A \cap T$ . Thus

$$P_T \leq \frac{\binom{2s-k}{s}}{\binom{2s}{s}} = \frac{\frac{(2s-k)!}{(s-k)!}}{\frac{(2s)!}{s!}} = \frac{(2s-k)(2s-k-1)\cdots(s-k+1)}{(2s)(2s-1)\cdots(s+1)} \leq \left(\frac{2s-k}{2s}\right)^s = \left(1 - \frac{k}{2s}\right)^s \leq e^{-(k/2s)s} = e^{-k/2} = e^{-(Cd \ln r)/r} = r^{-Cd/4}.$$

This estimates  $P_T$  for a fixed  $T \in F$ . To estimate  $P[E_1|A]$  in general we finally use, for the first time in the proof, the VC-dimension of  $X$ . The shatter function lemma

implies that the number of possible sets  $A \cap T$ , for  $T \in F$  is bounded, and the event “ $N \cap T = \emptyset, |M \cap T| \geq k$ ” depends only on  $A \cap T$ . We combine this with the above estimate via the union bound:

$$P[E_1|A] \leq \sum_{i=0}^d \binom{2s}{i} r^{-Cd/4} \leq \left(\frac{2es}{d}\right)^d r^{-Cd/4} = \left(2Cer^{1-\frac{C}{4}} \ln r\right)^d < \frac{1}{2}$$

when  $d, r \geq 2$  and  $C$  is chosen to be sufficiently large. Since this holds for all  $A$  we have  $P[E_1] < \frac{1}{2}$ , which implies  $P[E_0] < 1$ , concluding the proof.  $\square$

## 2.4 $\varepsilon$ -samples

I'd like to briefly mention a notion related to  $\varepsilon$ -nets that will also be useful.

**Definition 2.4.1.** *Given a set system  $(X, F)$ , a subset  $N \subseteq X$  is called an  $\varepsilon$ -sample for  $(X, F)$  if*

$$\left| \frac{|N \cap S|}{|N|} - \mu(S) \right| \leq \varepsilon$$

for all  $S \in F$ .

The notion of an  $\varepsilon$ -sample is stronger than that of an  $\varepsilon$ -net, since an  $\varepsilon$ -sample not only hits all heavy sets, but also represents them proportionally.  $\varepsilon$ -samples are in fact what Vapnik and Chervonenkis were after when they wrote their now-famous VC-dimension paper, and they proved that small  $\varepsilon$ -samples exist and can be found by random sampling for set systems with constant VC-dimension. The proof is similar to the Haussler-Welzl proof of the  $\varepsilon$ -net theorem. (Actually, the Haussler-Welzl proof is modeled after the original of Vapnik and Chervonenkis.)

**Theorem 2.4.2** ( $\varepsilon$ -sample theorem). *Given a set system  $(X, F)$  with  $\dim(F) \leq d$ , such that  $d \geq 2$  and  $r \geq 2$  is a parameter, there exists a  $(1/r)$ -sample for  $(X, F)$  of size at most  $Cdr^2 \ln r$ , where  $C$  is an absolute constant.*

## 2.5 Bounding the VC-dimension

The  $\varepsilon$ -net theorem provides a great incentive to bound the VC-dimension of various set systems, as this now implies that we can construct small  $\varepsilon$ -nets. We will soon see algorithmic applications of this, but let's begin with the VC-dimension bounds. We have already seen that the set system defined by halfspaces in  $\mathbb{R}^d$  has VC-dimension  $d + 1$ . We now generalize this result to polynomials.

**Theorem 2.5.1.** *Let  $R[x_1, x_2, \dots, x_d]_{\leq D}$  denote the set of all real polynomials in  $d$  variables of degree at most  $D$ , and let*

$$P_{d,D} = \{ \{x \in \mathbb{R}^d : p(x) \geq 0\} : p \in R[x_1, x_2, \dots, x_d]_{\leq D} \}.$$

*The VC-dimension of the set system  $(\mathbb{R}^d, P_{d,D})$  is at most  $\binom{d+D}{d}$ .*

*Proof.* The proof reduces the case of polynomials to the case of hyperplanes in a higher-dimensional space using *linearization*. Let  $M$  be the set of all possible nonconstant monomials of degree at most  $D$  in  $x_1, x_2, \dots, x_d$ . For example, when  $d = D = 2$  we have  $M = \{x_1, x_2, x_1x_2, x_1^2, x_2^2\}$ . Note that  $|M| = \binom{d+D}{d} - 1$ , since monomials of  $M$  correspond to placements of  $D$  identical balls in  $d + 1$  bins. The bins are the  $d$  coordinates, plus an “extra” one, which is used by monomials of degree strictly less than  $D$ . There is one configuration that corresponds to a constant monomial, and this is the one configuration that is subtracted from  $\binom{d+D}{d}$ . Denote  $m = |M|$ , and let the coordinates in  $\mathbb{R}^m$  be indexed by the monomials of  $M$ . The linearization we use is the mapping  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^m$ , defined by  $\varphi(x)_\mu = \mu(x)$ . For example, when  $d = D = 2$ , the map is

$$\varphi(x_1, x_2) = (x_1, x_2, x_1x_2, x_1^2, x_2^2).$$

Now, if  $A \in \mathbb{R}^d$  is shattered by  $P_{d,D}$ , then  $\varphi(A)$  is shattered by half-spaces in  $\mathbb{R}^m$ . Indeed, consider  $B \subseteq A$ , and let  $p \in P_{d,D}$  be a polynomial that is positive over  $B$  and negative over  $A \setminus B$ . Denote  $p = a_0 + \sum_{\mu \in M} a_\mu \mu$ . Now consider the halfspace  $h_p$  in  $\mathbb{R}^m$ , defined as  $\{y \in \mathbb{R}^m : a_0 + \sum_{\mu \in M} a_\mu y_\mu \geq 0\}$ . For example, if  $p = 2 + 3x_2 - 5x_1x_2 + x_1^2$ , then  $h_p = \{y \in \mathbb{R}^m : 2 + 3y_2 - 5y_3 + y_4 \geq 0\}$ . Then  $h_p \cap \varphi(A) = \varphi(B)$ , and in general  $\varphi(A)$  is shattered by halfspaces in  $\mathbb{R}^m$ . Thus  $\dim(P_{d,D}) \leq m + 1 = \binom{d+D}{d}$ .  $\square$

Now that we have bounded the VC-dimension of polynomials, we can extend the result even further, to the domain of *semialgebraic sets*. A semialgebraic set is a set definable by a Boolean combination of polynomial inequalities. More formally, a set  $A \subseteq \mathbb{R}^d$  is called semialgebraic if there are polynomials  $p_1, p_2, \dots, p_k \in R[x_1, x_2, \dots, x_d]_{\leq D}$ , for some  $D$ , and a Boolean formula  $F(X_1, X_2, \dots, X_k)$ , such that

$$A = \{x \in \mathbb{R}^d : F(p_1 \geq 0, p_2 \geq 0, \dots, p_k \geq 0)\}.$$

We can bound the VC-dimension of a set system defined by semialgebraic sets using the following general result.

**Theorem 2.5.2.** *Let  $F(X_1, X_2, \dots, X_k)$  be a set-theoretic formula involving the operations of union, intersection, and subtraction. Let  $(X, S)$  be a set system with bounded  $\dim(S) = d$ . Let*

$$T = \{F(s_1, s_2, \dots, s_k) : s_1, s_2, \dots, s_k \in S\}.$$

*Then  $\dim(T) = O(dk \log(dk))$ .*

*Proof.* Let  $A \subseteq X$  be an  $n$ -point set. By induction on the structure of  $F$ ,

$$F(s_1, s_2, \dots, s_k) \cap A = F(s_1 \cap A, s_2 \cap A, \dots, s_k \cap A).$$

In particular,  $F(s_1, s_2, \dots, s_k) \cap A$  depends only on the individual intersections  $s_i \cap A$ . Thus  $\pi_T(n) \leq \pi_S(n)^k$ . The shatter function lemma implies  $\pi_S(n) \leq \sum_{i=0}^d \binom{n}{i}$ . If  $A$  is shattered by  $T$ , then  $\pi_T(n) = 2^n$ . Thus

$$2^n \leq \left( \sum_{i=0}^d \binom{n}{i} \right)^k \leq \left( \frac{en}{d} \right)^{dk}.$$

Using elementary calculus, this implies  $n = O(dk \log(dk))$ .  $\square$

Now let  $H_d$  be the set of all hyperplanes in  $\mathbb{R}^d$ , and let

$$P_{d,1}^* = \{\{h \in H_d : h \text{ above } x\} : x \in \mathbb{R}^d\} \cup \{\{h \in H_d : h \text{ below } x\} : x \in \mathbb{R}^d\}.$$

Consider the set system  $(H_d, P_{d,1}^*)$ . By duality, the VC-dimension of this set system is  $d + 1$ . Now consider the set system

$$\Gamma_d = \left( H_d, \{\{h \in H_d : h \cap \gamma \neq \emptyset\} : \gamma \text{ is a line segment in } \mathbb{R}^d\} \right).$$

It is easy to see that  $\Gamma_d$  is contained in the set system

$$\left( H_d, \{(s_1 \cap s_2) \cup (s_3 \cap s_4) : s_1, s_2, s_3, s_4 \in P_{d,1}^*\} \right).$$

By above theorem, the VC-dimension of  $\Gamma_d$  is  $O(d \log d)$ . Now let

$$\Delta_d = \left( H_d, \{\{h \in H_d : h \cap \delta \neq \emptyset\} : \delta \text{ is a } d\text{-simplex in } \mathbb{R}^d\} \right).$$

A hyperplane intersects a simplex if and only if it intersects at least one of its edges. Thus  $\Delta_d$  is part of a set system obtained from  $P_{d,1}^*$  using a set theoretic formula on  $O(d)$  variables. Therefore, its VC-dimension is  $O(d^3 \log d)$ .