ε-net and VC-Dimension

497 - Randomized Algorithms

Sariel Har-Peled

November 14, 2002

The exposition here is based on [AS00].

1 VC Dimension

Definition 1.1 A *range space* S is a pair (X, R), where X is a (finite or infinite) set and R is a (finite or infinite) family of subsets of X. The elements of X are *points* and the elements of R are *ranges*. For $A \subseteq X$, $P_R(A) = \left\{ r \cap A \,\middle|\, r \in R \right\}$ is the *projection* of R on A.

If $P_R(A)$ contains all subsets of A (i.e., if A is finite, we have $|P_R(A)| = 2^{|A|}$) then A is *shattered* by R.

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

Let

$$g(d,n) = \sum_{i=0}^{d} \binom{n}{i}.$$

Note that for all $n, d \geq 1$, $g(d,n) = g(d,n-1) + g(d-1,n-1)$

Lemma 1.2 (Sauer's Lemma) *If (X,R) is a range space of* VC*-dimension d with $|X| = n$ points then $|R| \leq g(d,n)$.*

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

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

$$R_x = \left\{ r \setminus \{x\} \,\middle|\, x \in r, r \in R, r \setminus \{x\} \in R \right\}$$

and

$$R \setminus x = \left\{ r \setminus \{x\} \,\middle|\, r \in R \right\}.$$

Observe that $|R| = |R_x| + |R \setminus x|$ (Indeed, if r does not contain x than it is counted in R_x, if does contain x but $r \setminus x \notin R$, then it is also counted in R_x. The only remaining case is when both $r \setminus \{x\}$ and $r \cup \{x\}$ are in R, but then it is being counted once in R_x and once in $R \setminus x$.)

Observe that R_x has VC dimension $d-1$, as the largest set that can be shattered is of size $d-1$. Indeed, any set $A \subset X$ shattered by R_x, implies that $A \cup \{x\}$ is shattered in R.

Thus,

$$|R| = |R_x| + |R \setminus x| = g(n-1, d-1) + g(n-1, d) = g(d, n),$$

by induction. ∎

By applying Lemma 1.2, to a finite subset of X, we get:

Corollary 1.3 *If (X,R) is a rnage space of* VC*-dimension d then for every finitte subset A of X, we have $|P_R(A)| \leq g(d, |A|)$.*

Definition 1.4 Let (X,R) be a range space, and let A be a finite subset of X. For $0 \leq \varepsilon \leq 1$, a subset $B \subseteq A$, is an ε-*sample* for A if for any range $r \in R$, we have

$$\left| \frac{|A \cap r|}{|A|} - \frac{|B \cap r|}{|B|} \right| \leq \varepsilon.$$

Similarly, $N \subseteq A$ is an ε-*net* for A, if for any range $r \in R$, if $|r \cap A| \geq \varepsilon |A|$ implie that r contains at least one point of N (i.e., $r \cap N \neq \emptyset$).

Theorem 1.5 *There is a postive constant c such that if (X,R) is any range space of* VC*-dimension at most d, $A \subseteq X$ is a finite subset and $\varepsilon, \delta > 0$, then a random subset B of cardinality s of A wwhere s is at least the minimum between $|A|$ and*

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

is an ε-sample for A with probability at least $1-\delta$.

Theorem 1.6 *Let (X,R) be a range space of* VC*-dimension d, let A be a finite subset of X and suppose $0 < \varepsilon, \delta < 1$. Let N be a set obtained by m random independent draws from A, where*

$$m \geq \max\left(\frac{4}{\varepsilon} \log \frac{2}{\delta}, \frac{8d}{\varepsilon} \log \frac{8d}{\varepsilon} \right). \qquad (1)$$

Then N is an ε-net for A with probablity at least $1-\delta$.

1.1 Proof of Theorem 1.6

Let (X,R) be a range space of VC-dimension d, and let A be a subset of X of cardinality n. Suppose that m satisfiers Equation (1). Let $N = (x_1, \ldots, x_m)$ be the sample obtained by m independet samples from A (the elements of N are not necessarily distinct, and thats why we treat them as ordered set). Let E_1 be the probablity that N fails to be an ε-net. Namely,

$$E_1 = \left\{ \exists r \in R \;\middle|\; |r \cap A| \geq \varepsilon n, r \cap N = \emptyset \right\}.$$

(Namely, there exists a "heavy" range r that does not contain any point of N.) To complete the proof, we must show that $\mathbf{Pr}[E_1] \leq \delta$. Let $T = (y_1, \ldots, y_m)$ be another random sample generated in a similar fashion to N. Let E_2 be the event that N fails, but T "works", formally

$$E_2 = \left\{ \exists r \in R \,\middle|\, |r \cap A| \geq \varepsilon n, r \cap N = \emptyset, |r \cap T| \geq \frac{\varepsilon m}{2} \right\}.$$

($|r \cap T|$ denotes the number of elements of T belong to r.)

Intuitively, since $E_T\left[|r \cap T|\right] \geq \varepsilon m$, then for the range r that N fails for, we have with "good" probability that $|r \cap T| \geq \frac{\varepsilon n}{2}$. Namely, E_1 and E_2 have more or less the same probablity.

Claim 1.7 $\mathbf{Pr}[E_2] \leq \mathbf{Pr}[E_1] \leq 2\,\mathbf{Pr}[E_2]$.

Proof: Clearly, $E_2 \subseteq E_1$, and thus $\mathbf{Pr}[E_2] \leq \mathbf{Pr}[E_1]$. As for the other part, note that $\mathbf{Pr}\left[E_2 \,\middle|\, E_1\right] = \mathbf{Pr}[E_2 \cap E_1] / \mathbf{Pr}[E_1] = \mathbf{Pr}[E_2] / \mathbf{Pr}[E_1]$. It is thus enough to show that $\mathbf{Pr}\left[E_2 \,\middle|\, E_1\right] \geq 1/2$.

Assume that E_1 occur. There is $r \in R$, such that $|r \cap A| > \varepsilon n$ and $r \cap N = \emptyset$. The required problity is at least the probablity that for this spacific r, we have $|r \cap T| \geq \frac{\varepsilon n}{2}$. However, $|r \cap T|$ is a binomial variable with expectation εm, and variance $\varepsilon(1-\varepsilon)m \leq \varepsilon m$. Thus, by Cheby's inequality,

$$\mathbf{Pr}\left[|r \cap T| < \frac{\varepsilon m}{2}\right] \leq \mathbf{Pr}\left[||r \cap T| - \varepsilon m| > \frac{\varepsilon m}{2}\right] \mathbf{Pr}\left[||r \cap T| - \varepsilon m| > \frac{\sqrt{\varepsilon m}}{2}\sqrt{\varepsilon m}\right] \leq \frac{4}{\varepsilon m} \leq \frac{1}{2},$$

by Equation (1). Thus, $\mathbf{Pr}[E_2] / \mathbf{Pr}[E_1] = \mathbf{Pr}\left[|r \cap T| \geq \frac{\varepsilon n}{2}\right] = 1 - \mathbf{Pr}\left[|r \cap T| < \frac{\varepsilon m}{2}\right] \geq \frac{1}{2}$. ■

Thus, it is enough to bound the probablity of E_2. Let

$$E_2' = \left\{ \exists r \in R \,\middle|\, r \cap N = \emptyset, |r \cap T| \geq \frac{\varepsilon m}{2} \right\},$$

Clearly, $E_2 \subseteq E_2'$. Thus, bounding the probablity of E_2' is enough to prove the theorem. Note however, that a shocking thing happend! We no longer have A as participating in our event. Namely, we turned bounding an event that dependends on a global quantity, into bounding a quantity that depends only on local quantity/experiment. This is the crucial idea in this proof.

Claim 1.8 $\mathbf{Pr}[E_2] \leq \mathbf{Pr}[E_2'] \leq g(d, 2m) 2^{-em/2}$.

Proof: We imagine that we sample the elements of $N \cup T$ together, by picking $Z = (z_1, \ldots, z_{2m})$ independetly from A. Next, we randomly decide the m elements of Z that go into N, and remaining elements go into T. Clearly,

$$\mathbf{Pr}\left[E_2'\right] = \sum_Z \mathbf{Pr}\left[E_2' \,\middle|\, Z\right] \mathbf{Pr}[Z].$$

Thus, from this point on, we fix the set Z, and we bound $\mathbf{Pr}\left[E_2' \,\middle|\, Z\right]$.

It is now enough to consider the ranges in the projection space $P_Z(R)$. By Lemma 1.2, we have $|P_Z(r)| \leq g(d, 2m)$.

Let us fix any $r \in \mathcal{P}_Z(R)$, and consider the event

$$E_r = \left\{ |r \cap T| > \frac{\varepsilon m}{2} \text{ and } r \cap N = \emptyset \right\}.$$

For $p = |r \cap (N \cup T)|$, we have

$$\begin{aligned}
\mathbf{Pr}[E_r] &\leq \mathbf{Pr}\left[r \cap N = \emptyset \,\middle|\, |r \cap (N \cup T)| > \frac{\varepsilon m}{2} \right] = \frac{\binom{2m-p}{m}}{\binom{2m}{m}} \\
&= \frac{(2m-p)(2m-p-1)\cdots(m-p+1)}{2m(2m-1)\cdots(m+1)} \\
&= \frac{m(m-1)\cdots(m-p+1)}{2m(2m-1)\cdots(2m-p+1)} \leq 2^{-p} \leq 2^{-\varepsilon m/2}.
\end{aligned}$$

Thus,

$$\mathbf{Pr}\left[E_2' \,\middle|\, Z \right] \leq \sum_{r \in P_Z(R)} \mathbf{Pr}[E_r] \leq |P_Z(R)|\, 2^{-\varepsilon m/2} = g(d, 2m) 2^{-\varepsilon m/2},$$

implying that $\mathbf{Pr}[E_2'] \leq g(d, 2m) 2^{-\varepsilon m/2}$. ■

Proof of Theorem 1.6. By Lemma 1.7 and Lemma 1.8, we have $\mathbf{Pr}[E_1] \leq 2g(d, 2m) 2^{-\varepsilon m/2}$. It is thus remains to verify that if m satisfies Equation (1), then $2g(d, 2m) 2^{-\varepsilon m/2} \leq \delta$. One can verify that this inequality is implied by Equation (1). However, we omit the details, as this is tedious. ■

References

[AS00] N. Alon and J. H. Spencer. *The probabilistic method*. Wiley Inter-Science, 2nd edition, 2000.