CS 683: Advanced Algorithms	The ϵ-net Theorem
04/10/01	Scribes: Chaitanya Swamy, Leonid Meyerguz

1 The ϵ-net Theorem: Introduction

In the previous lecture, we looked at how large a random sample needs to be picked in order to ensure that the sample hits every "large" set in a given set system with high probability. More precisely, we stated the ϵ-net theorem, which gives the size of the random sample in terms of ϵ, the set system's VC-dimension and the desired probability bound. In this lecture, we present the proof of the theorem. Let's recall the statement of the theorem :

Theorem 1.1 (The ϵ-net Theorem) *Let $(U, \mathcal{R})$ be a set system with VC-dimension d. Then, a random sample $N \subseteq U$ of size $\frac{4}{\epsilon}\log\frac{2}{\delta} + \frac{8d}{\epsilon}\log\frac{8d}{\epsilon}$ is an ϵ-net for $(U, \mathcal{R})$ with probability at least $1 - \delta$.*

(Recall that a set S is an ϵ-net for $(U, \mathcal{R})$ if for all $A \in \mathcal{R}$ s.t. $|A| \geq \epsilon|U|$, $S \cap A \neq \emptyset$.)

Note that we clearly need to make use of the fact that $(U, \mathcal{R})$ has low(bounded) VC-dimension. Picking a set N at random guarantees that for any *one* set $A \in \mathcal{R}$ s.t. $|A| \geq \epsilon|U|$, $\Pr[A \cap N = \emptyset] \leq (1-\epsilon)^{|N|} \leq e^{-\epsilon|N|}$. However for N to be an ϵ-net we need to make sure that it hits *every* large set. Using the union bound the only bound we can obtain on the probability that N misses some large set in $\mathcal{R}$, is $|\mathcal{R}|e^{-\epsilon|N|}$. So to get a high probability guarantee the size of our ϵ-net would need to depend on $|\mathcal{R}|$ which in turn depends on $|U|$ (although by Lemma 2.1 we have reduced the dependence on $|U|$ from linear to logarithmic). Our goal is to bound the size of the ϵ-net in terms of just the VC-dimension.

2 Proving the ϵ-net Theorem

Before we proceed with the proof of the main theorem, we will find it useful to bound the size of a set system's collection of events $\mathcal{R}$ in terms of the size of its universe U and its VC-dimension. Note that all sets in $\mathcal{R}$ must be subsets of U, so we can certainly say that $|\mathcal{R}| \leq 2^{|U|}$. However, as the following lemma shows, we can come up with a better bound.

Lemma 2.1 *For any set system $(U, \mathcal{R})$ with $|U| = n$ and VC-dimension d, $|\mathcal{R}| \leq g(d, n)$, where $g(d, n) = \binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{d} \leq n^d$*

Proof. We proceed by induction on $n + d$. The trivial base case is when both $n = 0$ and $d = 0$, with $U = \emptyset$ and $\mathcal{R} = \{\emptyset\}$. Now, consider an arbitrary set system $(U, \mathcal{R})$, with VC-dimension d. We will inductively show that $|\mathcal{R}| \leq g(d, n)$.

Pick an arbitrary x in U, and let $U' = U - \{x\}$. Also, let $\mathcal{R}' = \{A \cap U' \mid A \in \mathcal{R}\}$ ie. $\mathcal{R}'$ is the projection of $\mathcal{R}$ onto U' (also denoted by $\mathcal{R}_{|\mathcal{U}}$). Consider the set system $(U', \mathcal{R}')$. Clearly, the VC-dimension of this new system is at most d (since if $\mathcal{R}'$ shatters a set of size $d+1$ so does $\mathcal{R}$). So by the induction hypothesis, $|\mathcal{R}'| \leq g(d, n-1)$.

Now, consider the difference between $|\mathcal{R}|$ and $|\mathcal{R}'|$. Note that sets A and A' in $\mathcal{R}$ map onto the the same set in $\mathcal{R}'$ only if $A = A' \cup \{x\}$ and $x \notin A'$. Let us define another set system $(U, \mathcal{R}'')$ where $\mathcal{R}'' = \{A' \mid A' \in \mathcal{R},\ x \notin A',\ A' \cup \{x\} \in \mathcal{R}\}$. Then, we can immediately infer that $|\mathcal{R}| = |\mathcal{R}'| + |\mathcal{R}''|$, since the sets in $\mathcal{R}''$ are exactly those that were counted only once in $\mathcal{R}'$.

We now claim that the VC-dimension of $(U', \mathcal{R}'')$ is at most $d-1$. To prove this claim, we will show that whenever $\mathcal{R}''$ shatters a set Y, $\mathcal{R}$ shatters the set $Y \cup \{x\}$. More specifically, we will show that for $\forall A \subseteq Y \cup \{x\},\ \exists S \in \mathcal{R}$ s.t. $S \cap (Y \cup \{x\}) = A$. We consider the following two cases:

1. $A \subseteq Y$. Since Y is shattered by $\mathcal{R}''$, $\exists S \in \mathcal{R}''$ s.t. $S \cap Y = A$. By the definition of $\mathcal{R}''$, $x \notin S,\ S \in \mathcal{R}$, so clearly $(S \cap (Y \cup \{x\}) = A$.

2. $x \in A$. Since $\mathcal{R}''$ shatters Y, $\exists S \in \mathcal{R}''$ s.t. $S \cap Y = A - \{x\}$. Then, $S \cup \{x\} \in \mathcal{R}$, and $(S \cup \{x\}) \cap (Y \cup \{x\}) = A \cup \{x\} = A$.

Since the VC-dimension of $(U', \mathcal{R}'')$ is at most $d-1$, by the induction hypothesis, $|\mathcal{R}''| \leq g(d-1, n-1)$. Therefore,

$$
\begin{aligned}
|\mathcal{R}| = |\mathcal{R}'| + |\mathcal{R}''| &\leq g(d, n-1) + g(d-1, n-1) \\
&= \sum_{i=0}^{d} \binom{n-1}{i} + \sum_{i=0}^{d-1} \binom{n-1}{i} = \binom{n-1}{0} + \sum_{i=1}^{d} \binom{n-1}{i} + \binom{n-1}{i-1} \\
&= \binom{n}{0} + \sum_{i=1}^{d} \binom{n}{i} = \sum_{i=0}^{d} \binom{n}{i} = g(d, n)
\end{aligned}
$$

Finally, simple induction shows that $g(d, n) \leq n^d$.

$$
\begin{aligned}
g(d+1, n) = g(d, n) + \binom{n}{d+1} &\leq n^d + \frac{n!}{(d+1)!(n-d-1)!} && (1) \\
&\leq n^d + \frac{n^{d+1}}{(d+1)!} && (2) \\
&\leq n^{d+1} && (3)
\end{aligned}
$$

Above, (1) follows from the induction hypothesis, (2) follows by upper-bounding the quotient in (1), and (3) follows because both halves of the sum are smaller than $\frac{n^{d+1}}{2}$ for $n \geq 2, d \geq 1$. This concludes the proof of our lemma. ■

Corollary 2.2 *If $X \subseteq U$, then $(X, \mathcal{R}_{|X})$ has VC-dimension $\leq d$ which implies $|\mathcal{R}_{|X}| \leq |X|^d$.*

We now proceed with the proof of the ϵ-net theorem. Let $m = \frac{4}{\epsilon}\log\frac{2}{\delta} + \frac{8d}{\epsilon}\log\frac{8d}{\epsilon}$. We want to show that if N is a random set of size m, the probability that N does not hit some large set in $\mathcal{R}$ is at most δ. Consider the event $E_1 = \{\exists A \in \mathcal{R} \text{ s.t. } A \cap N = \emptyset,\ |A| \geq \epsilon n\}$. We have to show that $\Pr[E_1] \leq \delta$. We introduce another event, E_2, bound $\Pr[E_1]$ in terms of $\Pr[E_2]$, and then in turn bound $\Pr[E_2]$. Suppose we pick another random sample, T, also of size m. Note that both N and T may have multiple copies of an element if the element is picked more than once. Let E_2 be the event that N misses some set A in $\mathcal{R}$, $|A| \geq \epsilon n$, but T has a "large" intersection with the set A. More formally, define $E_2 = \{\exists A \in \mathcal{R} \text{ s.t. } A \cap N = \emptyset,\ |A| \geq \epsilon n,\ |A \cap T| \geq \frac{\epsilon m}{2}\}$.

Lemma 2.3 $\Pr[E_2] \geq \frac{1}{2}\Pr[E_1]$

Proof. Clearly $E_2 \subseteq E_1$, since if E_2 happens then E_1 definitely happens. Since,

$$\Pr[E_2|E_1] = \frac{\Pr[E_2 \cap E_1]}{\Pr[E_1]} = \frac{\Pr[E_2]}{\Pr[E_1]},$$

we only need show that $\Pr[E_2|E_1] \geq \frac{1}{2}$. Let $T = \{y_1, \ldots, y_m\}$ (where y_i and y_j may be copies of the same element). Since E_1 happens there is some set A s.t. $|A| \geq \epsilon n,\ A \cap N = \emptyset$. Therefore $\Pr[E_2|E_1]$ is at least the probability that for this A, $|A \cap T| \geq \frac{\epsilon m}{2}$.

Let $|A| = \epsilon n$ (if $|A| \geq \epsilon n$ consider some ϵn size subset of it). Let X_i be the random variable(rv.) such that $X_i = 1$ if $y_i \in A$, and 0 otherwise. Let $X = \sum_{i=1}^{m} X_i$. Clearly, $X = |A \cap T|$. $\mathrm{E}X_i = \Pr[X_i = 1] = |A|/|U| = \epsilon$, so $\mathrm{E}X = \epsilon m$. To bound the deviation from the expectation we use use Chebyshev's inequality : $\Pr[|X - E(X)| > t] \leq \frac{\mathrm{Var}(X)}{t^2}$.

Since the X_i's are independent and identically distributed, $\mathrm{Var}(X) = m\mathrm{Var}(X_i)$. $\mathrm{Var}(X_i) = \mathrm{E}X_i^2 - (\mathrm{E}X_i)^2 = \epsilon - \epsilon^2$ since for the 0-1 rv. $X_i,\ X_i = X_i^2$. So $\mathrm{Var}(X) = m\epsilon(1-\epsilon)$.

Now, $\Pr[X < \frac{\epsilon m}{2}] \leq \Pr[|X - \mathrm{E}X| > \frac{\epsilon m}{2}] \leq \frac{4m\epsilon(1-\epsilon)}{\epsilon^2 m^2} = \frac{4(1-\epsilon)}{\epsilon m} \leq \frac{4}{\epsilon m} \leq \frac{1}{2}$ (since $m \geq \frac{8d}{\epsilon} \geq \frac{8}{\epsilon}$).

So $\Pr[E_2|E_1] \geq \Pr[X \geq \frac{\epsilon m}{2}] \geq \frac{1}{2}$ proving the lemma. ■

The big picture is that to bound $\Pr[E_1]$ we can equivalently bound $\Pr[E_2]$ and use Lemma 2.3. An alternative useful way of viewing the random experiment is as follows : pick a set $Z \subseteq U$ of size $2m$ at random and then pick an m element subset of Z at random to be N, letting $T = Z - N$. Then intuitively $\Pr[E_2]$ should be "small" since it is unlikely that some $A \in \mathcal{R}$ having a large intersection with T, and hence with Z, misses N completely, N being picked randomly from Z. The following lemma makes this more precise.

Lemma 2.4 $\Pr[E_2] \leq g(d, 2m)2^{-\frac{\epsilon m}{2}}$

Proof. We consider the alternate view of the random experiment described above. The two views are equivalent since the probability that an element $x \in U$ is included in N is $\frac{m}{n}$ in both cases. Let Z be the $2m$ random element set picked in step 1 and let $T = Z - N$ (so $Z = N \cup T$). Let $A \in \mathcal{R}$. Define $E_A = \{|A \cap T| \geq \frac{\epsilon m}{2},\ A \cap N = \emptyset,\ |A| \geq \epsilon n\}$.

If $|A \cap Z| = |A \cap N| + |A \cap T| < \frac{\epsilon m}{2}$, E_A does not happen. So,

$$
\begin{aligned}
\Pr[E_A] &= \Pr[N \text{ chosen s.t. } N \cap A = \emptyset,\ |A \cap (N \cup T)| \geq \frac{\epsilon m}{2}] \\
&\leq \frac{\binom{2m - \frac{\epsilon m}{2}}{m}}{\binom{2m}{m}} = \frac{(2m - \frac{\epsilon m}{2}) \cdots (m+1) \cdot m \cdots (m - \frac{\epsilon m}{2} + 1)}{2m \cdots (2m - \frac{\epsilon m}{2} + 1) \cdot (2m - \frac{\epsilon m}{2}) \cdots (m+1)} \\
&= \frac{m \cdots (m - \frac{\epsilon m}{2} + 1)}{2m \cdots (2m - \frac{\epsilon m}{2} + 1)} \leq 2^{-\frac{\epsilon m}{2}}.
\end{aligned}
$$

Now for two sets $A, A' \in \mathcal{R}$ s.t. $|A|, |A'| \geq \epsilon n$ and $A \cap (N \cup T) = A' \cap (N \cup T)$, the events $E_A, E_{A'}$ are the same. Also the number of sets $A \in \mathcal{R}$ s.t. $A \cap (N \cup T)$ is unique is at most $|\{A \cap (N \cup T) | A \in \mathcal{R}\}| = |\mathcal{R}_{|\mathcal{Z}}| \leq g(d, 2m)$ by Corollary 2.2 . So,

$$
\begin{aligned}
\Pr[E_2] &\leq \bigcup_{A | A \cap (N \cup T) \text{ is unique}} \Pr[E_A] \leq |\mathcal{R}_{|\mathcal{Z}}| 2^{-\frac{\epsilon m}{2}} \\
&\leq g(d, 2m) 2^{-\frac{\epsilon m}{2}},
\end{aligned}
$$

proving the lemma. ■

Finally we just need to show that our choice of m gives $\Pr[E_1] \leq 2g(d, 2m)2^{-\frac{\epsilon m}{2}} \leq \delta$. Since $g(d, 2m) \leq (2m)^d$, it suffices to show that $2(2m)^d \leq \delta 2^{\frac{\epsilon m}{2}}$ ie. $\frac{\epsilon m}{2} \geq \log \frac{2}{\delta} + d \log 2m$. Since $m \geq \frac{4}{\epsilon} \log \frac{2}{\delta}$, $\frac{\epsilon m}{4} \geq \log \frac{2}{\delta}$. So we only need show that $\frac{\epsilon m}{4} \geq d \log(2m)$. It's easy to check that for $m \geq \frac{8d}{\epsilon} \log \frac{8d}{\epsilon}$ the above inequality holds.

So $\Pr[E_1] \leq \delta$ which implies that N is an ϵ-net with probability at least $1 - \delta$ proving the theorem.