ϵ-net and ϵ-sample

Nader H. Bshouty
Department of Computer Science
Technion 32000
Israel
e-mail: bshouty@cs.technion.ac.il

Lynn Burroughs
Department of Computer Science
University of Calgary
Calgary, Alberta, Canada
lynnb@cpsc.ucalgary.ca

Abstract

Here we will give the proof of the ϵ-net and ϵ-sample theorem.

1 Preliminaries

Let F be a boolean function $F : X \to \{0,1\}$ and D a distribution on X. Let $\mathcal{U}$ be the uniform distribution. We will write $x \in_D X$ when we want to indicate that x is chosen from X according to the distribution D. Suppose we randomly and independently choose $S = \{x_1, \ldots, x_m\}$ from X, each x_i according to the distribution D. We will write E_X for $E_{x \in_D X}$. So for finite X we have

$$E_X[F(x)] = \sum_{x \in X} D(x)F(x),$$

and for infinite X we have ($D(x)$ is the distribution function)

$$E_X[F(x)] = \int F(x)dD(x).$$

We use E_S for $E_{x \in_{\mathcal{U}} S}$. So for a finite sample $S \subset X$ we have

$$E_S[F(x)] = \sum_{x \in S} \frac{F(x)}{|S|}.$$

We say that $\mathcal{S} = (X, C)$ is a *range space* if X is any set and C is a set of boolean functions $X \to \{0,1\}$. Each function in C can be also regarded as a subset of X. We will also call C *concept class.* For a boolean function $F \in C$ and a subset $A \subseteq X$ the *projection of F on A* is a boolean function $F_{|_A} : A \to \{0,1\}$ such that for every $x \in A$ we have $F_{|_A}(x) = F(x)$. For a subset $A \subseteq X$ we define the *projection of C on A* to be the set

$$P_C(A) = \{F_{|_A} \mid F \in C\}.$$

If $P_C(A)$ contains all the functions in 2^A then we say that A is *shattered*. The *Vapnik-Chervonenkis dimension* (or VC-dimension) of $\mathcal{S}$, denoted by $\text{VCdim}(\mathcal{S})$, is the maximum cardinality of a shattered subset of X.

Let (X, C) be a range space and D be a distribution on X. We say that a set of points $S \subseteq X$ is an *ϵ-net* if any $F \in C$ satisfies $E_X[F(x)] \geq \epsilon$ contains at least one positive point, i.e., a point y in S such that $F(y) = 1$. Notice that $E_S[F(x)] = 0$ if and only if S contains no positive point for F. Therefore, S is not an ϵ-net if

$$(\exists F \in C)\ E_X[F(x)] > \epsilon \text{ and } E_S[F(x)] = 0.$$

We say that S is ϵ-sample if

$$(\forall F \in C)\ |E_X[F(x)] - E_S[F(x)]| \leq \epsilon.$$

We will denote $d(r_1, r_2) = |r_1 - r_2|$. Therefore, S is not an ϵ-sample if

$$(\exists F \in C) d(E_X[F(x)], E_S[F(x)]) > \epsilon.$$

Notice that an ϵ-sample is an ϵ-net.

2 The Theorems

Let C be a concept class of boolean functions $F : X \to \{0, 1\}$. Suppose we randomly and independently choose $S = \{x_1, \ldots, x_m\}$ from X according to the distribution D. We have

Bernoulli For

$$m = \frac{1}{\epsilon} \ln \frac{1}{\delta}$$

we have

$$\Pr[E_X[F(x)] > \epsilon \text{ and } E_S[F(x)] = 0] \leq \delta.$$

Chernoff (Additive form) For

$$m = \frac{1}{2\epsilon^2} \ln \frac{2}{\delta}$$

we have

$$\Pr\left[|E_X[F(x)] - E_S[F(x)]| > \epsilon\right] \leq 2e^{-2\epsilon^2 m} = \delta.$$

Bernoulli For any finite concept class C and

$$m = \frac{1}{\epsilon}\left(\ln|C| + \ln\frac{1}{\delta}\right)$$

we have

$$\Pr\left[(\exists F \in C)\ E_X[F(x)] > \epsilon \text{ and } E_S[F(x)] = 0\right] \leq \delta.$$

Chernoff (Additive form) For any finite concept class C and

$$m = \frac{1}{2\epsilon^2}\left(\ln|C| + \ln\frac{2}{\delta}\right)$$

we have

$$\Pr\left[(\exists F \in C)\ |E_X[F(x)] - E_S[F(x)]| > \epsilon\right] \leq \delta.$$

We have

ϵ-Net ([HW]) There is a constant c_{Net} such that for any concept class C and

$$m = \frac{c_{Net}}{\epsilon}\left(\text{VCdim}(C)\log\frac{1}{\epsilon} + \log\frac{1}{\delta}\right)$$

we have

$$\Pr\left[(\exists F \in C)\ E_X[F(x)] > \epsilon \text{ and } E_S[F(x)] = 0\right] \leq \delta.$$

ϵ-Sample ([VC]) There is a constant c_{VC} such that for any concept class C and

$$m = \frac{c_{VC}}{\epsilon^2}\left(\text{VCdim}(C)\log\frac{\text{VCdim}(C)}{\epsilon} + \log\frac{1}{\delta}\right)$$

we have

$$\Pr\left[(\exists F \in C)\ |E_X[F(x)] - E_S[F(x)]| > \epsilon\right] \leq \delta.$$

Define

$$g(d, n) = \sum_{i=1}^{d}\binom{n}{i}.$$

Exercise. Use the inequality $g(d, 2m) \leq (2m)^d$ to show that the following Lemma implies the proof of the ϵ-net result.

Lemma. Let (X, C) be a range space of VC-dimension d. Let D be a distribution over X. Let S be a sequence of points obtained by m random independent draws from X according to the distribution D where

$$2g(d, 2m)e^{-\frac{\epsilon m}{4}} \leq \delta,$$

and $m \geq 8/\epsilon$. Then with probability at least $1 - \delta$ we have that S is an ϵ-net for X.

Proof: Let C_ϵ be the set of all $F \in C$ with $E_X[F(x)] \geq \epsilon$. Define the random variable

$$A = [(\exists F \in C_\epsilon)\ E_S[F(x)] = 0]. \tag{1}$$

That is, $A = 1$ if the statement in the square brackets is true and 0 otherwise. Notice that $E_S[F(x)] = 0$ means that no point y in S is positive for F. We will write $\Pr_S[A]$ for $\Pr_S[A = 1]$. To prove the lemma we need to prove that

$$\Pr_S[A] \leq \delta.$$

Now the difficulty here is that the number of elements in C_ϵ may be infinite.

The approach we will take here is the following: Notice that $\Pr_S[A] = E_S[A]$. Now we change the probability space to an equivalent one as follows. Instead of choosing m points in X according to the distribution D we choose $2m$ points W from X according to the distribution D and then uniformly choose m points N from W. Obviously, this is the same probability space and therefore

$$\Pr_S[A] = \Pr_{W,N}[A].$$

Notice that here (and in the sequel) we are using the same event A for two different probability spaces. What we actually mean here is: $\Pr_S[A_S] = \Pr_{W,N}[A_N]$ where A_S is the event defined in (1) and A_N is the same event where we replace S by N.

Now we use the following beautiful result in probability. Let B be an event. Then

$$E_S[B] = E_{W,N}[B] = E_W[E_N[B|W]].$$

The inner expectation is $E_N[B|W]$ is the expectation of the event B when W is a fixed set. Now, it is easier to handle this expectation because W is finite (not like X) and the set $\{F_{|W} \mid F \in C\}$ is also finite. For the proof we will choose B to be the event

$$B = [(\exists F \in C_\epsilon)\ E_N[F(x)] = 0 \text{ and } E_W[F(x)] \geq \epsilon/4].$$

Notice that B is A with the extra condition that $E_W[F(x)] \geq \epsilon/4$. When $F \in C_\epsilon$ the probability that a random point in X is positive for F is greater than or equal to ϵ. So for $F \in C_\epsilon$ the condition $E_W[F(x)] \geq \epsilon/4$ is true with high probability. Therefore we expect that the probability of A to be close to the probability of B. We added the condition $E_W[F(x)] \geq \epsilon/4$ to obtain the property that is similar to $F \in C_\epsilon$ (which is $E_X[F] \geq \epsilon$) over the finite sub-domain W. We now formally prove this

Claim 1: We have

$$\Pr_S[A] \leq 2 \Pr_{W,N}[B].$$

Proof of Claim 1: We have

$$\Pr_{W,N}[\bar{B}|A] =$$

$$\Pr[(\forall F \in C_\epsilon)\ E_N[F(x)] > 0 \text{ or } E_W[F(x)] \le \epsilon/4 \ \mid\ (\exists F \in C_\epsilon)\ E_N[F(x)] = 0].$$

Let $F_0 \in C_\epsilon$ such that $E_N[F_0(x)] = 0$. Then the above probability is

$$\begin{aligned}
\Pr_{W,N}[\bar{B}|A] &\le \Pr[E_N[F_0(x)] > 0 \text{ or } E_W[F_0(x)] \le \epsilon/4] && \text{Since } E_N[F_0(x)] = 0 \\
&= \Pr[E_W[F_0(x)] < \epsilon/4] && \text{Since } |W| = 2|N| \\
&\le \Pr[E_{W\setminus N}[F_0(x)] \le \epsilon/2] && \text{Since } F_0 \in C_\epsilon. \\
&\le \frac{1}{2}
\end{aligned}$$

Exercise. Prove the latter inequality using Chebyschev and using the condition $m \ge 8/\epsilon$.

Now

$$\begin{aligned}
\Pr_{W,N}[B] &= \Pr_{W,N}[A \text{ and } B] \\
&= \Pr_{W,N}[B|A] \Pr_{W,N}[A] \\
&= (1 - \Pr_{W,N}[\bar{B}|A]) \Pr_S[A] \\
&\ge \frac{\Pr_S[A]}{2}. \square
\end{aligned}$$

Now we prove

Claim 2: We have

$$\Pr_{W,N}[B] \le g(d, 2m) e^{-\frac{\epsilon m}{4}}.$$

Proof of Claim 2: For each $F \in C_\epsilon$ let

$$B_F = [E_N[F(x)] = 0 \text{ and } E_W[F(x)] \ge \epsilon/4].$$

Then

$$B = \bigvee_{F \in C_\epsilon} B_F.$$

Now if we fix $F \in C_\epsilon$ we have

$$E_{W,N}[B_F] = E_W[E_N[B_F|W]].$$

Now

$$\begin{aligned}
E_N[B_F|W] &= \Pr[B_F|W] \\
&= \Pr[E_N[F(x)] = 0 \text{ and } E_W[F(x)] \ge \epsilon/4 \mid W] \\
&\le \Pr[E_N[F(x)] = 0 \mid W,\ \ E_W[F(x)] \ge \epsilon/4] \\
&\le \left(1 - \frac{\epsilon}{4}\right)^m \le e^{-\frac{\epsilon m}{4}}.
\end{aligned}$$

We can regard $B_F|W$ as the event

$$B_F|W = [E_N[F_{|_W}(x)] = 0 \text{ and } E_W[F_{|_W}(x)] \geq \epsilon/4].$$

Now if $F_{|_W} = F'_{|_W}$ then the events $B_F|W$ and $B_{F'}|W$ are the same events. By Sauer lemma the number of different events is at most

$$|\{F_{|_W} \mid F \in C\}| \leq |P_W(C)| \leq g(d, 2m).$$

Therefore

$$\begin{aligned} \Pr[B] &= E_{W,N}\left[\bigvee_{F\in C_\epsilon} B_F\right] \\ &\leq E_W\left[E_N\left[\bigvee_{F\in C_\epsilon} B_F \mid W\right]\right] \\ &\leq g(d,2m)E_W[E_N[B_F|W]] \\ &\leq g(d,2m)e^{-\frac{\epsilon m}{4}}.\square \end{aligned}$$

Exercise. Show that the following Lemma implies the proof of the ϵ-sample result.

Lemma. Let (X, C) be a range space of VC-dimension d. Let D be a distribution over X. Let S be a sequence of points obtained by m random independent draws from X according to the distribution D where

$$2g(d, 2m)e^{-\frac{\epsilon^2 m}{2}} \leq \delta,$$

and $m \geq 2\ln 2/\epsilon^2$. Then with probability at least $1 - \delta$ we have that S is an ϵ-sample for X.

Proof: Define the random variable

$$A = [(\exists F \in C)\ d(E_X[F(x)], E_S[F(x)]) \geq \epsilon].$$

To prove the lemma we need to prove that

$$\Pr_S[A] \leq \delta.$$

Now we change the probability space to an equivalent one as follows. Instead of choosing m points in X according to the distribution D we choose $2m$ points W from X according to the distribution D and then uniformly choose m points N from W. Obviously, this is the same probability space and therefore

$$\Pr_S[A] = \Pr_{W,N}[A].$$

Let B be an event. Then

$$E_S[B] = E_{W,N}[B] = E_W[E_N[B|W]].$$

For the proof we will choose B to be the event

$$B = [(\exists F \in C)\ d(E_X[F(x)], E_N[F(x)]) \geq \epsilon \text{ and } d(E_W[F(x)], E_N[F(x)]) \geq \epsilon/2].$$

We now can prove

Claim 3: We have

$$\Pr_S[A] \le 2 \Pr_{W,N}[B].$$

Proof of Claim 3: Suppose A is true and let $F_0 \in C$ such that $d(E_X[F_0(x)], E_N[F_0(x)]) \ge \epsilon$. Then

$$\begin{aligned}
\Pr_{W,N}[\bar{B}|A] &\le \Pr[d(E_W[F_0(x)], E_N[F_0(x)]) < \epsilon/2] \\
&\le \Pr[d(E_W[F_0(x)], E_X[F_0(x)]) > \epsilon/2] \\
&\le \frac{1}{2}.
\end{aligned}$$

Exercise. Prove the latter inequality using Chernoff bound and using the condition $m \ge 2\ln 2/\epsilon^2$.

Now as in the ϵ-net proof we have

$$\Pr_{W,N}[B] \ge \frac{\Pr_S[A]}{2}. \square$$

Now we prove

Claim 4: We have

$$\Pr_{W,N}[B] \le g(d, 2m) e^{-\frac{\epsilon^2 m}{2}}.$$

Proof of Claim 4: Let

$$C_\epsilon = \{F \in C \mid d(E_X[F(x)], E_N[F(x)]) \ge \epsilon\}.$$

$$B_F = [d(E_W[F(x)], E_N[F(x)]) \ge \epsilon/2].$$

Then

$$B = \bigvee_{F \in C_\epsilon} B_F.$$

Now if we fix $F \in C_\epsilon$ we have

$$E_{W,N}[B_F] = E_W[E_N[B_F|W]].$$

Now for a fix F and by Chernoff bound we have

$$\begin{aligned}
E_N[B_F|W] &= \Pr[B_F|W] \\
&= \Pr[d(E_W[F(x)], E_N[F(x)]) \ge \epsilon/2 \mid W] \\
&\le e^{-2\left(\frac{\epsilon}{2}\right)^2 m} = e^{-\frac{\epsilon^2 m}{2}}.
\end{aligned}$$

Now if $F_{|W} = F'_{|W}$ then the events $B_F|W$ and $B_{F'}|W$ are the same events. By Sauer lemma the number of different events is at most

$$|\{F_{|W} \mid F \in C\}| \le |P_W(C)| \le g(d, 2m).$$

Therefore,

$$\begin{aligned}
\Pr[B] &= E_{W,N}\left[\bigvee_{F\in C_\epsilon} B_F\right] \\
&\leq E_W\left[E_N\left[\bigvee_{F\in C_\epsilon} B_F \mid W\right]\right] \\
&\leq g(d,2m)E_W[E_N[B_F|W]] \\
&\leq g(d,2m)e^{-\frac{\epsilon^2 m}{2}}.\Box
\end{aligned}$$

Minimal Expectation For any concept class C and

$$m = \min\left(\frac{c_{VC}}{\epsilon^2}\left(\text{VCdim}(C)\log\frac{\text{VCdim}(C)}{\epsilon} + \log\frac{1}{\delta}\right)\ ,\ \frac{1}{2\epsilon^2}\left(\ln|C| + \ln\frac{1}{\delta}\right)\right)$$

we have

$$\Pr\left[\left|\min_{F\in C} E_X[F(x)] - \min_{F\in C} E_S[F(x)]\right| \geq \epsilon\right] \leq \delta.$$

Proof of the Minimal Expectation. We use the Chernoff and Vapnik Chervonenkis bounds. Let $G, H \in C$ such that

$$E_X[G(x)] = \min_{F\in C} E_X[F(x)]\ ,\ E_S[H(x)] = \min_{F\in C} E_S[F(x)].$$

Then with probability at least $1-\delta$ we have

$$E_S[H(x)] \leq E_S[G(x)] \leq E_X[G(x)] + \epsilon \leq E_X[H(x)] + \epsilon \leq E_S[H(x)] + 2\epsilon$$

Which implies that $|E_X[G(x)] - E_S[H(x)]| \leq \epsilon.\Box$

References

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

[VC] V. N. Vapnik, A. Y. Chervonenkis, On the uniform convergence of relative frequencies of events to their probabilities. *theory of Probability and its Applications*, 16(2): 264-280, 1971.