Algorithmic Excursions: Topics in Computer Science II **Spring 2016**

Lecture 5 & 6 : ε-approximations, ε-nets from ε-approximations and the ε-net theorem

Lecturer: Kasturi Varadarajan *Scribe: Talal Riaz*

3.1 ε-approximations and ε-nets via ε-approximations

Theorem 3.1 *Suppose (X, R) has VC-dimension $d \geq 1$, and suppose that an ε is given where $0 \leq \varepsilon \leq 1$. Then (X, R) has an ε-approximation $A \subseteq X$ of size $O(\frac{d}{\varepsilon^2})\ln(d/\varepsilon)$.*

Proof: Assume that the size of the ground set is an integer power of 2 (**as a homework problem try to prove this theorem without this assumption**).

Let $X_0 = X$. Construct $X_1, \ldots, X_t$ as follows: For $0 \leq i \leq t-1$, let $X_{i+1} \subseteq X_i$ be such that:

- $|X_{i+1}| = |X_i|/2$
- X_{i+1} is an ε_{i+1}-approximation of $(X_i, R|_{X_i})$ where, (this can be done because of lemma 3.7 from the last lecture)

$$\varepsilon_{i+1} \leq \frac{\sqrt{|X_i| \cdot \ln(4 \cdot R|_{X_i})}}{|X_i|}$$

Since $(R|_{X_i}) \leq |X_i|^d$

$$\varepsilon_{i+1} \leq \frac{\sqrt{|X_i| \cdot \ln(4 \cdot |X_i|^d)}}{|X_i|}$$

For some constant c

$$\varepsilon_{i+1} \leq \sqrt{\frac{c \cdot d \cdot \ln(|X_i|)}{|X_i|}}$$

Notice that X_t is a δ-approximation for (X, R) where $\delta = \varepsilon_1 + \ldots \varepsilon_t$. Since the bounds on ε_i are geometrically increasing, we can bound the total error by a constant times the largest (last) error term. Thus, for absolute constants c, c''

$$\delta \leq c'' \cdot \sqrt{\frac{c \cdot d \cdot \ln(|X_{t-1}|)}{|X_{t-1}|}}$$

The RHS will be at most ε provided $|X_{t-1}| \geq \alpha \cdot \frac{d}{\varepsilon^2}\ln(d/\varepsilon)$ for some constant α. We simply choose t to be the first integer for which $X_t < \alpha \cdot \frac{d}{\varepsilon^2}\ln(d/\varepsilon)$. X_t is thus an $\varepsilon - approximation$. ■

We claim, however that we can do slightly better than this. The following claims are left as **homework exercises**.

Claim 3.2 *If A is an $\varepsilon/2$-approximation for (X, R) and A' is an $\varepsilon/2$-net for $(A, R|_A)$ then A' is an ε-net for (X, R).*

Claim 3.3 *(X, R) with VC-dimension d has an ε-net of size $O(d/\varepsilon \ln(d/\varepsilon))$*

3.2 ε-net theorem and the double sampling proof

Theorem 3.4 *(ε-**net theorem**, proved by Haussler-Welzl in '87) Suppose (X, R) has VC-dimension $2 \le d < \infty$, and suppose $0 < \varepsilon < 1/2$. Let N be a random sample of X of size $c \cdot d/\varepsilon \ln(d/\varepsilon)$ where c is a large constant. Then N is an ε-net with* $\Pr \ge 1/2$.

Preliminaries Let $r = 1/\varepsilon, s = c \cdot d \cdot r \cdot \ln(r)$. We may assume that each $r' \in R$ has $> \varepsilon|X|$ elements. Let E_0 denote the event that $\exists r' \in R$ such that $N \cap r' = \phi$. Our goal is to show that $\Pr[E_0] < 1/2$. Suppose that we pick another sample M of size s using the same sampling process. Let $k = s/2r$. Let E_1 denote the event that $\exists r' \in R$ such that $N \cap r' = \phi$ and $|M \cap r'| \ge k$. Since $E_1 \subseteq E_0$, this means that $\Pr(E_1) \le \Pr(E_0)$. We claim that

Claim 3.5 $\Pr[E_0] \le 2\Pr[E_1]$.

Proof: Lets condition on N. It suffices to show that

$$\Pr[E_0|N] \le 2\Pr[E_1|N]$$

First suppose that N is an ε-net (this means that $N \cap r' \ne \phi\ \forall r' \in R$). Then, clearly $\Pr[E_0|N] = 0$ and $\Pr[E_1|N] = 0$. Thus $\Pr[E_0|N] \le 2\Pr[E_1|N]$.

Now, suppose that N is an ε-net (this means that $\exists r' \in R$ such that $N \cap r' = \phi$). Clearly, $\Pr[E_0|N] = 1$. We need to show that:

$\Pr[E_1|N] \ge \Pr[|M \cap r| \ge k] \ge 1/2$

Let $Y_i = 1$ if the i^{th} sample in M belongs to r' and 0 otherwise. Since $|r'| > \varepsilon|X|$, $E[Y_i] = \Pr[Y_i = 1] \ge 1/r = \varepsilon$.
Let $Y = \sum_{i=1}^{s} Y_i$. Note that $Y = |M \cap r'|$. By linearity of expectations, $E[Y] = \sum_{i=1}^{s} E[Y_i] \ge s/r = 2k$. **As a homework problem, argue using Chebyshev's inequality that** $\Pr[\mathbf{Y \le k}] < \mathbf{1/2}$. This means that $\Pr[E_1|N] \ge 1/2$. Thus $\Pr[E_0|N] \le 2\Pr[E_1|N]$. This completes the proof of Claim 6.1. ■

3.2.1 Double sampling:

Think of N and M as being produced in the following way:

- Pick a sample $A \subseteq X$ of size $2s$.
- Pick a random subset of size s from A to form N.

- Let $M = A \setminus N$

To show that $\Pr[E_1] \leq 1/4$, we fix A and show that the probability $\Pr[E_1|A] \leq 1/4$. More specifically, for each $r' \in R$, we will bound the following probability:

$$\alpha \equiv \Pr\left[N \cap r' = \phi \wedge |M \cap r'| \geq k \middle| A\right]$$

Suppose that $|A \cap r'| < k$, then α is simply 0. Now, suppose that $|A \cap r'| \geq k$. Then,

$$\alpha \leq \left(1 - \frac{k}{2s}\right) \cdot \left(1 - \frac{k}{2s-1}\right) \cdots \left(1 - \frac{k}{2s-s+1}\right)$$

$$\alpha \leq \left(1 - \frac{k}{2s}\right)^s \leq \exp\left(-\frac{k}{2s} \cdot s\right)$$

$$\alpha \leq \exp(-k/2) = \frac{1}{exp(\frac{c \cdot d}{4} \cdot \ln(r))} = \frac{1}{r^{cd/4}}$$

Now,

$$\Pr[E_1|A] = \Pr\left[\bigcup_{r' \in R} [N \cap r' = \phi \wedge |M \cap r'| \geq k|A]\right]$$

$$\Pr[E_1|A] = \Pr\left[\bigcup_{b \in R|A} \bigcup_{r' \in R: A \cap r' = b} [N \cap r' = \phi \wedge |M \cap r'| \geq k|A]\right]$$

Using the union bound,

$$\Pr[E_1|A] \leq \sum_{b \in R|A} \Pr\left[\bigcup_{r' \in R: A \cap r' = b} [N \cap r' = \phi \wedge |M \cap r'| \geq k|A]\right]$$

For a specific b, the terms inside the union just refer to the same event. Thus we can use the bound from above,

$$\Pr[E_1|A] \leq \sum_{b \in R|A} \frac{1}{r^{cd/4}} \leq \phi_d(2s) \cdot \frac{1}{r^{cd/4}}$$

$$\Pr[E_1|A] \leq \left(\frac{2s \cdot e}{d}\right)^d \cdot \frac{1}{r^{cd/4}}$$

$$\Pr[E_1|A] \leq (2cre\ln(r))^d \cdot \frac{1}{r^{cd/4}}$$

We can chose c to be large enough so that,

$$\Pr[E_1|A] \leq 1/4$$

This means that $\Pr[E_1] \leq 1/4$. From Claim 6.1, $\Pr[E_0] \leq 2\Pr[E_1] \leq 1/2$. This completes the proof of the ε-net theorem.

3.2.2 Paper topics:

Applications of ε-nets to set cover, load balancing and cuttings.