Today

VC dimension

- Sauer-Shelah Lemma
- ε-nets

CS5.205.1
Toolkit in TCS
- Lecture #32
(14 June '21)
Instructor: Prahladh Harsha

Recap from last lecture.

$(X, \mathcal{S})$ - X - universe (not necessarily finite)

$\mathcal{S} \subseteq 2^X$ family of sets.

finite $A \subseteq X$.

$$\mathcal{S}|_A = \{ S \cap A \mid S \in \mathcal{S} \}.$$

A is shattered by $\mathcal{S}$ if $\mathcal{S}|_A = 2^A$

VC-dim $(\mathcal{S})$ = maximum size of a shattered set

Primal Shatter Coefficient

$$\pi_{\mathcal{S}}(m) = \max \{ |\mathcal{S}|_A| \mid A \subseteq X, |A| = m \}$$

For all $m \le d$, $\pi_{\mathcal{S}}(m) = 2^m$

What about $m > d$?

$\pi_{\mathcal{S}}(m) < 2^m$, but is it significantly smaller.

Lemma [VC-dimension lemma, Sauer-Shelah]

$(X, \mathcal{S})$ - set system w/ VC-dim d, then

$$\pi_{\mathcal{S}}(m) \le \binom{m}{\le d} = \binom{m}{0} + \binom{m}{1} + \dots + \binom{m}{d}$$

Pf: (Via Polynomial Method).

Let $A \subseteq X$, $|A| = m$.

Ambient space $\{0,1\}^A \cong \{0,1\}^m$

$\forall \; T \in \mathcal{S}|_A$, $\mathbb{1}_T \in \{0,1\}^A$

- indicator fn of T.

$$V = V(\mathcal{S}|_A) = \{\mathbb{1}_T \mid T \in \mathcal{S}|_A\}$$

$$\mathcal{F} = \{f : V \to \mathbb{R}\} \quad , \quad \dim(\mathcal{F}) = N = |\mathcal{S}|_A|$$

For each $T \in \mathcal{S}|_A$

$$P_T(x) = \prod_{i \in T} x_i \prod_{i \in A \setminus T} (1 - x_i)$$

$T_1, T_2 \in \mathcal{S}|_A$.

$$P_{T_1}(\mathbb{1}_{T_2}) = \begin{cases} 1 & \text{if } T_1 = T_2 \\ 0 & \text{if } T_1 \neq T_2 \end{cases}$$

$$P_{T_1}(\mathbb{1}_{T_2}) = \delta[T_1 = T_2]$$

Obs: $\{P_T \mid T \in \mathcal{Q}/_A\}$ are linearly independent

$\#\{P_T \mid T \in \mathcal{Q}/_A\} = \dim(\mathcal{F})$.

Hence $\{P_T \mid T \in \mathcal{Q}/_A\}$ form a basis for $\mathcal{F}$.

To complete the proof, we will show that $\forall T$

$$P_T \in \operatorname{Span}\{x_I \mid I \subseteq A;\ |I| \le d\}$$

then, $\mathcal{F} \subseteq \operatorname{Span}\{x_I \mid I \subseteq A,\ |I| \le d\}$

$$|\mathcal{Q}/_A| = \dim \mathcal{F} \le \binom{m}{\le d}.$$

Let $I \subseteq A$; $|I| = d+1$

$x_I = x_{i_1} x_{i_2} \ldots x_{i_{d+1}}$

$|I| > d \Rightarrow \exists\, B \subseteq I$ s.t. $B \notin \mathcal{Q}/_I$ $\Bigg\}$ Since VC-dim$(\mathcal{Q}) = d$ & $|I| > d$.

$$q_B(x) = \prod_{j \in B} x_j \prod_{j \in I \setminus B} (1 - x_j)$$

$q_B(\mathbb{1}_T) = 1 \iff \forall j \in B,\ j \in T$ & $\forall j \in I \setminus B,\ j \notin T$ $\iff T \cap I = B$

Hence $\forall\, 1\!\!1_T \in V$, $q_B(1\!\!1_T) = 0$.

Hence $q_B|_V = 0$

$$\prod_{j\in B} x_j \prod_{j\in I\setminus B}(1-x_j) = 0 \quad \text{for all } x\in V$$

$$\prod_{i\in I} x_i = \text{sum of lower deg monomials for all } x\in V.$$

Hence, every $f\in\mathcal{F}$ can be written as a sum of monomials of degree at most d.

Hence $|\mathcal{S}|_A| = \dim\mathcal{F} \le \binom{m}{\le d}$ $\boxtimes$

—

Epsilon-Nets (ε-nets)

$(X, \mathcal{S})$ - set system. X - finite

$\mu: X \to [0,1]$: $\sum_{x\in X}\mu(x) = 1$

ε-net $A \subseteq X$ - A has a representative from every heavy set

Formally

$A \subseteq X$, is an ε-net for $\mathcal{S}$ if

$\forall S \in \mathcal{S}, \quad \mu(S) \geq \varepsilon \Rightarrow S \cap A \neq \emptyset$

where $\mu(S) = \sum_{x \in S} \mu(x)$

Qn. Do there exist small ε-nets?

Weak ε-net Theorem:

For any set system $(X, \mathcal{S})$ & prob. measure μ & $\varepsilon \in (0,1)$, there exists an ε-net A of size $\leq \frac{1}{\varepsilon} \ln |\mathcal{S}|$

Pf: Construct a set $A \subseteq X$ by picking t elements from X independently accrg to dist μ (w/ repetitions, A - multiset)

$S \in \mathcal{S}, \quad \mu(S) \geq \varepsilon.$

$$\Pr_A [S \cap A = \emptyset] \leq (1-\varepsilon)^t < e^{-\varepsilon t}$$

$$\Pr_A [\exists S, \mu(S) \geq \varepsilon, S \cap A = \emptyset] < |\mathcal{S}| \cdot e^{-\varepsilon t}$$

≤ 1 if $t = \frac{1}{\varepsilon} \ln |\mathcal{S}|$

ie, $\Pr_A[A \text{ is not an } \varepsilon\text{-net}] < 1$

Hence $\exists$ ε-net A of size $t = \lceil \frac{1}{\varepsilon} \ln |\mathcal{S}| \rceil$ $\boxtimes$

Can there be smaller sized ε-nets?

YES, if VC-dim $(\mathcal{S})$ is small.

Theorem [ε-net theorem]

$\forall d > 1$, $(X, \mathcal{S})$ - set system w/ VC-dim $= d$.

& μ - prob dist on X & $\varepsilon \in (0,1)$.

then $\exists$ an ε-net A of $\mathcal{S}$ of size

$$|A| \leq \frac{d}{\varepsilon}\left(\ln \frac{1}{\varepsilon} + 2 \ln\ln \frac{1}{\varepsilon} + 6\right).$$

Pf: A - pick t elements from X independently according to μ.

$$E = \Pr_A\left[\exists S \in \mathcal{S}, \ \mu(S) \geq \varepsilon \ \& \ A \cap S = \phi\right]$$

B - pick $(T-t)$ elements from X independently according to μ.

Fix an $S \in \mathcal{S}_1$, $\mu(S) \geq \varepsilon$.

m_S - median of the number of elements in B.

ie, m_S - integer s.t.

$$\Pr_B[|B \cap S| < m_S] \leq \frac{1}{2} \leq \Pr_B[|B \cap S| \geq m_S]$$

$$\mathbb{E}_B[|B \cap S|] = (T-t)\mu(S)$$

(counting w/ mult)

$$|B \cap S| \sim \text{Binomial}(T-t, \mu(S))$$

$$\begin{aligned} m_S = \text{median} &\geq \text{mean} - 1 \\ &= (T-t)\mu(S) - 1 \\ &\geq (T-t)\varepsilon - 1 \end{aligned}$$

$$E = \Pr_A[\exists S \in \mathcal{S}_1\ \mu(S) \geq \varepsilon,\ |S \cap A| = 0]$$

$$\geq \Pr_{A,B}[\exists S \in \mathcal{S}_1\ \mu(S) \geq \varepsilon,\ |S \cap A| = 0,\ |S \cap B| \geq m_S]$$

$$E \leq \frac{\Pr_{A,B}[\exists S \in \mathcal{S}_1\ \mu(S) \geq \varepsilon,\ |S \cap A| = 0 \ \&\ |S \cap B| \geq m_S]}{\min_{S \in \mathcal{S}} \Pr[|B \cap S| \geq m_S]}$$

$$\leq 2 \cdot \Pr_{A,B}\left[\exists S \in \mathcal{S},\ \mu(S) \geq \varepsilon,\ \begin{array}{l} |S \cap A| = 0 \\ |S \cap B| \geq m_S \end{array}\right].$$

Fix $S \in \mathcal{S}$, $\mu(S) \geq \varepsilon$.

$$\Pr_{A,B}\left[\,|S \cap A| = 0;\ |S \cap B| \geq m_S\right]$$

Change the experiment as follows

first pick C – T elements.

st $|C \cap S| \geq m_S$

& then set $\left.\begin{array}{l} A \leftarrow \binom{C}{t} \\ B \leftarrow C \setminus A \end{array}\right\}$

Let us suppose we have picked C st $|C \cap S| \geq m_S$.

$$\Pr_{A,B,C}\left[\,|A \cap S| = 0,\ |B \cap S| \geq m_S \;\middle/\; |C \cap S| \geq m_S\right]$$

$$\Pr_{A,B,C}\left[\,|A \cap S| = 0 \;\middle/\; |C \cap S| = k\right]$$

where $k \geq m_S$

$$= \frac{\binom{T-k}{t}}{\binom{T}{t}}$$

$$= \frac{t!\,(T-t)!\,(T-k)!}{T!\quad t!\,(T-k-t)!}$$

$$= \frac{\binom{T-t}{k}}{\binom{T}{k}} = \frac{(T-t)(T-t-1)\cdots(T-t-(k-1))}{T(T-1)\cdots(T-(k-1))}$$

$$\leq \left(1-\frac{t}{T}\right)^k \leq \left(1-\frac{t}{T}\right)^{m_s}$$

$$\Pr_{C,A,B}\left[A\cap S=\phi,\ B\cap S \geq m_s \,/\, C\cap S \geq m_s\right]$$

$$\leq \left(1-\frac{t}{T}\right)^{m_s}$$

$$\leq \left(1-\frac{t}{T}\right)^{m}$$

$$m_s = \min_{S\text{-heavy}} m_S$$

$$\geq (T-t)\varepsilon - 1$$

Due to Sauer-Shelah Lemma, the # of intersection patterns of C w/ $\mathcal{S}_r \leq \binom{T}{\leq d}$

$$\Pr\left[\exists S,\ \mu(S)\geq \varepsilon,\ A\cap S=\phi,\ B\cap S \geq m_s\right]$$

$$\leq \left(1-\frac{t}{T}\right)^m \sum_{i=0}^{d}\binom{T}{i} \cdots \alpha.$$

If $\alpha < 1$ then we are done ✓

If $t = \left\lceil \frac{d}{\varepsilon}\left(\ln\frac{1}{\varepsilon} + 2\ln\ln\frac{1}{\varepsilon} + 6\right)\right\rceil$

$\tau = \left\lceil \frac{\varepsilon}{d}\epsilon^2 \right\rceil$, then $\alpha < 1$

✓ $\boxtimes$

Role of VC-dimension in Boosting

X - universe.

H - class of Boolean fns on X.

$\mathcal{P}$ - dist on $X \times \{0,1\}$

h - hypothesis.

Generalization Error of h:

$$\varepsilon_h = \Pr_{(x,y)\sim\mathcal{D}}\left[h(x) \neq y\right].$$

$S = \{(x_1,y_1) \ldots , (x_N, y_N)\}$

$(x_t, y_t) \sim \mu$. (independently)

Empirical Error of h on S:

$$\hat{\varepsilon}_h = \frac{\{i \in [N] \mid h(x_i) \neq y_i\}}{N}$$

Theorem [Vapnik].

H - set of hypotheses w/ VC-dim $\leq d$.

then $\forall \delta$:

$$\Pr_S\left[\exists h \in H,\ |\hat{\varepsilon}(h) - \varepsilon(h)| > 2\sqrt{\frac{d\ln\left(\frac{2N}{d}\right) + \ln\left(\frac{9}{\delta}\right)}{N}}\right] \leq \delta.$$

H - class of hypotheses.

$\Theta_T(H)$ - class of hypotheses o/p by a T-round boosting alg.

$$\text{VC-dim}(H) \leq d \implies \text{VC-dim}(\Theta_T(H)) \leq 2(d+1)(T+1)\log_2(e(T+1))$$