Math 547: Mathematical Foundations of Statistical Learning Theory Fall 2017

Lecture 22 — October 11

Instructor: Stas Minsker *Scribe: Mose Wintner*

22.1 VC Dimension

$$\mathbb{E}||P_n - P||_{\mathcal{C}} \le 2\mathbb{E}||R_n||_{\mathcal{C}} = \frac{2}{\sqrt{n}}\mathbb{E}_X\mathbb{E}_\varepsilon \sup_{C\in\mathcal{C}} |\frac{1}{\sqrt{n}}\sum \varepsilon_j I_C(X_j)| \le \frac{2}{\sqrt{n}}\mathbb{E}_X 12\sqrt{2}\int_0^\infty \sqrt{\log N(d_n,\mathcal{C},\varepsilon)}\, d\varepsilon.$$

The process

$$I_C \mapsto \frac{1}{\sqrt{n}}\sum \varepsilon_j I_C(X_j)$$

is sub-Gaussian with parameter $\frac{1}{n}\sum I_C^2(X_j)$. Then

$$\frac{1}{\sqrt{n}}\sum \varepsilon_j (I_{C_1}(X_j) - I_{C_2}(X_j))$$

is sub-Gaussian with parameter $P_n(I_{C_1} - I_{C_2})^2$. Given functions f, g define $d_n^2(f,g) = \sqrt{P_n(f-g)^2}$.

Goal: find a non-random bound for $N(d_n, \mathcal{C}, \varepsilon)$ when $\mathcal{C}$ is a VC class.

Remark 1. Given $f, g : S \to \mathbb{R}$, $d_n(f,g) \le ||f-g||$.

Definition 1. Given a class of funtions $\mathcal{F} = \{f : S \to \mathbb{R}\}$,

$$H(\mathcal{F},\varepsilon) = \sup_{Q\in P(S)} H(\mathcal{F}, L_2(Q), \varepsilon),$$

where $P(S)$ denotes the set of probability measures on S.

$$||f-g||^2_{L_2(Q)} = \int (f(x) - g(x))^2\, dQ(x).$$

If Q is a discrete measure then this integral becomes a sum.

Theorem 1. (Hassler) If $\mathcal{C}$ is a VC class and $\mathcal{F}$ is a collection of indicator functions indexed by $\mathcal{C}$, then $H(\mathcal{F},\varepsilon) \le 5V(\mathcal{C})\log(\frac{B}{\varepsilon})$ for all $\varepsilon \le 1$, where B is some numerical constant.

Corollary 1.

$$\mathbb{E}||P_n - P||_{\mathcal{C}} \le \frac{C}{\sqrt{n}}\int_0^2 \sqrt{5V(\mathcal{C})\log\frac{B}{\varepsilon}}\, d\varepsilon = C_1\sqrt{V(\mathcal{C})/n}.$$

For $C_1,\ C_2 \in \mathcal{C}$,

$$||I_{C_1} - I_{C_2}||^2_{L_2(Q)} = \int_S (I_{C_1} - I_{C_2})^2(x)\, dQ = Q(C_1 \Delta C_2).$$

The integrand is binary: 0 if x belongs to/avoids both sets and 1 if it belongs to only one. If $\mathcal{C}$ is a VC class, then so is the set of symmetric differences among pairs of sets in $\mathcal{C}$ because each is obtained by a finite number of intersections and complements. Let $\tilde{\mathcal{C}}$ denote the set of symmetric differences of sets in $\mathcal{C}$. Then $\tilde{\mathcal{C}}$ is VC and $m^{\tilde{\mathcal{C}}}(n) \leq 2^{2^2}(m^{\mathcal{C}}(n))^2$ by previous results. Hence

$$\mathbb{E}||P_n - P||_{\tilde{\mathcal{C}}} \leq K\sqrt{\frac{\log m^{\tilde{\mathcal{C}}}(n)}{n}} \leq K_1\sqrt{\frac{\log m^{\mathcal{C}}(n)}{n}}$$

for any probability measure P and any $n \geq 1$.

Since $\mathbb{E}||P_n - P||_{\tilde{\mathcal{C}}} \leq K_1\sqrt{\frac{\log m^{\mathcal{C}}(n)}{n}}$, we have that for any $C_1, C_2 \in \mathcal{C}$,

$$P(C_1 \Delta C_2) \leq P_n(C_1 \Delta C_2) + K_1\sqrt{\frac{\log m^{\mathcal{C}}(n)}{n}}.$$

Assume that P_n is supported on $X_1(\omega), \ldots, X_n(\omega)$. There exists $\mathcal{C}' \subset \mathcal{C}$ of cardinality at most $m^{\tilde{\mathcal{C}}}(n)$ such that

$$\mathcal{C}' \cap \{X_1(\omega), \ldots, X_n(\omega)\} = \tilde{\mathcal{C}} \cap \{X_1(\omega), \ldots, X_n(\omega)\}.$$

Then for all $C \in \mathcal{C}$ there exists $C' \in \mathcal{C}'$ such that $P_n(C \Delta C') = 0$. Therefore

$$P(C \Delta C') \leq K_1\sqrt{\frac{\log m^{\mathcal{C}}(n)}{n}},$$

which also means

$$||I_C - I_{C'}||_{L_2(P)} = \sqrt{P(C \Delta C')} \leq K_2 \left(\frac{\log m^{\mathcal{C}}(n)}{n}\right)^{1/n}.$$

Therefore C' is a $L_2(P)$ ε_n-net for $\mathcal{C}$, where

$$\varepsilon_n = K_2 \left(\frac{\log m^{\mathcal{C}}(n)}{n}\right)^{1/4}.$$

Now assume that $\varepsilon > 0$ is fixed and take $n = BV(\mathcal{C})/\varepsilon^5$ for B sufficiently large. Then $\varepsilon_n \leq \varepsilon$;

$$\varepsilon_n = K_2 \left(\frac{V(\mathcal{C})C \log \frac{BV(\mathcal{C})}{\varepsilon^5} \varepsilon^5}{BV(\mathcal{C})}\right)^{1/4} \lesssim \varepsilon.$$

For this choice of ε_n,

$$\operatorname{card}(\mathcal{C}') \leq \log m^{\mathcal{C}}(n) \leq V(\mathcal{C}) \log \frac{ne}{V(\mathcal{C})}$$

by the bound from a previous lecture. Plugging in our value of n,

$$\log \operatorname{card}(\mathcal{C}') \leq 5V(\mathcal{C}) \log \frac{Be}{\varepsilon},$$

so

$$H(\mathcal{C}, \varepsilon) \leq 5V(\mathcal{C}) \log \frac{Be}{\varepsilon}.$$