Math 547: Mathematical Foundations of Statistical Learning Theory Fall 2017

Lecture 24 — October 16

Instructor: Stas Minsker *Scribe: Mose Wintner*

24.1 Symmetric difference metric

$\mathcal{F} = \{f : S \to \mathbb{R}\}$, $F(x) = \sup_{f\in\mathcal{F}} |f(x)|$ the "envelope".
$\Gamma_{\mathcal{F}} = \{\Gamma_f : f \in \mathcal{F}\}$ a VC class. Want to bound $\sup_Q N(\mathcal{F}, L_2(Q), \varepsilon)$.

Last time we showed that

$$\int_S |f - g|\, dQ = \frac{(Q \times \Lambda)(\Gamma_f \Delta \Gamma_g)}{2\int F\, dQ} 2 \int F\, dQ.$$

Assume that $\sup_x F(x) \le M$. Then

$$\begin{aligned}
\int_| f - g|\, dQ &\le \varepsilon \\
\Leftrightarrow \mu(\Gamma_f \Delta \Gamma_g) &\le \frac{\varepsilon}{2\int F\, dQ} = \varepsilon' \\
\Rightarrow N(\mathcal{F}, L_1(Q), \varepsilon) &\le N(\Gamma_{\mathcal{F}}, \mu, \varepsilon') \\
&\le 5V(\Gamma_{\mathcal{F}}) \log \frac{2B||F||_{L_1(Q)}}{\varepsilon} \\
&\le 5V(\Gamma_f) \log \frac{2BM}{\varepsilon}.
\end{aligned}$$

To obtain the bound in $L_2(Q)$, note that

$$||f - g||^2_{L_2(Q)} = \int_S (f - g)^2\, dQ \le |f - g| \cdot 2F\, dQ.$$

If this is $\le \varepsilon^2$, then $||f - g||_{L_2(Q)} \le \varepsilon$. If $||f - g||_{L_1(Q)} \le \frac{\varepsilon^2}{2M}$, which implies

$$N(\mathcal{F}, L_2(Q), \varepsilon) \le N(\mathcal{F}, L_1(Q), \varepsilon^2/2M) \le 5V(\Gamma_{\mathcal{F}}) \log \frac{2BM \cdot 2M}{\varepsilon^2}.$$

We proved in a previous lecture something like

$$\mathbb{E}||P_n - P||_{\mathcal{C}} \le K \sqrt{\sup_{C\in\mathcal{C}} P(C)} \sqrt{1/n} \vee \sqrt{1/n}.$$

Theorem 1. Let $\mathcal{C}$ be a VC class of VC dimension V. Let $\mathcal{A}$ be a class of distributions for (X, Y) for which there exists $C \in \mathcal{C}$ such that $Y = 1$ if and only if $X \in C$. Then

$$\inf_{g_n \in S: \to \{\pm 1\}} \sup_{\mathcal{A}} P(Y \ne g_n(X)) \ge \frac{V - 1}{2en} \left(1 - \frac{1}{n}\right)$$

Proof: There exist $\{x_1, \dots, x_V\} \subset S$ which is shattered by C. Consider the following family $\mathcal{A}'$ of distributions. Consider the following family of distributions: $X = x_i$ with probability $1/n$ for $i = 1, \dots, V-1$, $X = x_V$ with probability $1 - (V-1)/n$; and $Y = f_b(X) = b_i$ if $X = x_i$ for $i = 1, \dots, V-1$ and -1 if $i = V$, where $b = (b_1, \dots b_{V-1}) \in \{\pm 1\}^{V-1}$. This gives a family of 2^{V-1} distributions. Note that $\inf_{C \in \mathcal{C}} P(Y \neq g_C(X)) = 0$ where $g_C(X) = (-1)^{X \notin C}$. Then $\mathcal{A}' \subset \mathcal{A}$. Hence

$$\sup_{\mathcal{A}} P(Y \neq g_n(X)) \geq \sup_{\mathcal{A}'} P(Y \neq g_n(x)) = \sup_{b} P(f_b(X) \neq g_n(X)) \geq 2^{1-V} \sum_{b \in \{\pm 1\}^{V-1}} P(f_b(X) \neq g_n(X))$$

by estimating the minimax risk by the Bayes risk from below (i.e. assuming b is random, estimating supremum from below by the average). This is then equal to $P = (f_B(X) \neq g_n(X))$, where $B \sim U\{\pm 1\}^{V-1}$ a discrete uniform. This is equal then to

$$\int_{g_n(X) \neq \eta(X)} |\eta(X)| \, d\Pi.$$

Note that