4.5 V-C Dimension and Shatter Coefficient

The relationship between the V-C dimension and the shatter coefficient is discussed in this section. The following lemma, which is generally attributed to Vapnik-Chervonenkis (1971) and Sauer (1972), shows that the V-C dimension provides a useful bound for the shatter coefficient.

Lemma 4. *If $\mathcal{A}$ is a class of sets with V-C dimension $V_\mathcal{A}$, then for every n,*

$$s(\mathcal{A}, n) \le \sum_{i=0}^{V_\mathcal{A}} \binom{n}{i}.$$

The lemma shows the dichotomy of the behavior of the shatter coefficient. For $n \le V_\mathcal{A}$, $s(\mathcal{A}, n) = 2^n$, which is the same as the upper bound $\sum_{i=0}^{V_\mathcal{A}} \binom{n}{i} = \sum_{i=0}^{n} \binom{n}{i} = 2^n$. On the other hand, for $n > V_\mathcal{A}$, the upper bound, the sum of binomial coefficients up to $V_\mathcal{A}$, grows at most polynomially as indicated by the corollary below.

Corollary 1. *If $n > V_\mathcal{A}$, then*

$$s(\mathcal{A}, n) < (\frac{en}{V_\mathcal{A}})^{V_\mathcal{A}}.$$

Proof.

$$\begin{aligned}
\sum_{i=0}^{V_\mathcal{A}} \binom{n}{i} &\le \sum_{i=0}^{V_\mathcal{A}} \binom{n}{i} \left(\frac{n}{V_\mathcal{A}}\right)^{V_\mathcal{A}-i} = \left(\frac{n}{V_\mathcal{A}}\right)^{V_\mathcal{A}} \sum_{i=0}^{V_\mathcal{A}} \binom{n}{i} \left(\frac{V_\mathcal{A}}{n}\right)^{i} \\
&\le \left(\frac{n}{V_\mathcal{A}}\right)^{V_\mathcal{A}} \sum_{i=0}^{n} \binom{n}{i} \left(\frac{V_\mathcal{A}}{n}\right)^{i} = \left(\frac{n}{V_\mathcal{A}}\right)^{V_\mathcal{A}} \left(1 + \frac{V_\mathcal{A}}{n}\right)^{n} \\
&\le \left(\frac{n}{V_\mathcal{A}}\right)^{V_\mathcal{A}} e^{V_\mathcal{A}} = \left(\frac{en}{V_\mathcal{A}}\right)^{V_\mathcal{A}}.
\end{aligned}$$

For the last inequality, we have used the fact that $e^x \ge 1+x$, which implies $x \ge \log(1+x)$ for $x > -1$. Thus $\log(1+V_\mathcal{A}/n)^n = n\log(1+V_\mathcal{A}/n) \le n \cdot V_\mathcal{A}/n = V_\mathcal{A}$ and $(1+V_\mathcal{A}/n)^n \le e^{V_\mathcal{A}}$. □

Remark 8. If $\mathcal{A}$ has a finite V-C dimension, $s(\mathcal{A}, n)$ grows exponentially until $n = V_\mathcal{A}$ and then grows at a polynomial rate for $n > V_\mathcal{A}$. So, the growth function defined as $\log_2 s(\mathcal{A}, n)$ is linear up to the $V_\mathcal{A}$ and becomes logarithmic afterwards as shown in Figure 3.

Example 2.

(a) $\mathcal{A} = \{(-\infty, x] : x \in \mathbb{R}\}$ with $V_\mathcal{A} = 1$.

$$s(\mathcal{A}, n) \le \sum_{i=0}^{V_\mathcal{A}} \binom{n}{i} = \binom{n}{0} + \binom{n}{1} = 1 + n.$$

However, we have checked that $s(\mathcal{A}, n) = n + 1$ for this example.

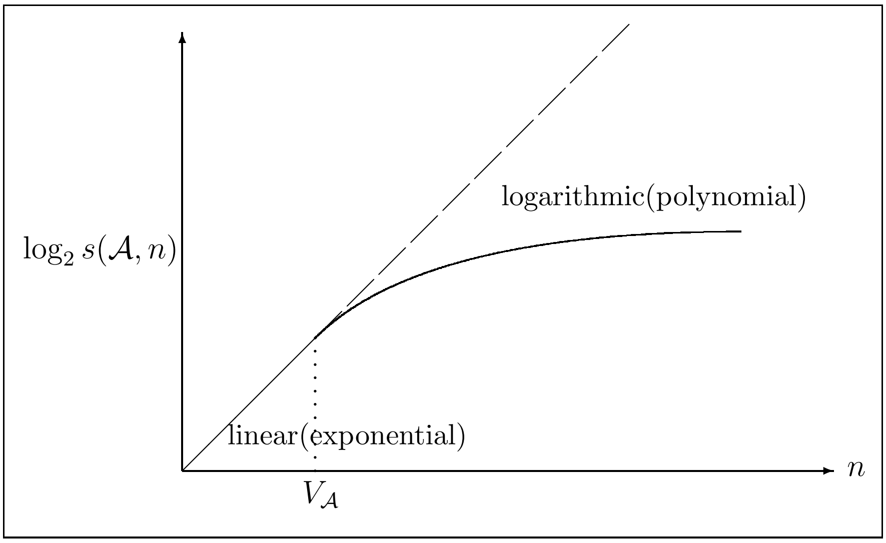

Figure 3: Growth function

(b) $\mathcal{A} = \{\{x \in \mathbb{R}^2 : \ \beta_0 + \beta' x \geq 0\} : \ \beta_0 \in \mathbb{R}, \beta \in \mathbb{R}^2\}$ with $V_{\mathcal{A}} = 3$.

$$s(\mathcal{A}, 4) = 14 < \sum_{i=0}^{3} \binom{4}{i} = 1 + 4 + 6 + 4 = 15.$$

Now we put together the properties of the shatter coefficient and the Vapnik-Chervonenkis inequality,

$$P\left(\sup_{f \in \mathcal{F}} |R_n(f) - R(f)| > \epsilon\right) \leq 8\, s(\mathcal{F}, n) \exp(-n\epsilon^2/32),$$

to obtain the following. Let $V_{\mathcal{A}}$ be the V-C dimension of the class of sets induced by the decision rules in $\mathcal{F}$, $\{f^{-1}\{1\} | f \in \mathcal{F}\}$.

Theorem 10. *If $V_{\mathcal{A}} < \infty$ then, for $n \geq V_{\mathcal{A}}$,*

$$P\left(\sup_{f \in \mathcal{F}} |R_n(f) - R(f)| > \epsilon\right) \leq 8 \left(\frac{en}{V_{\mathcal{A}}}\right)^{V_{\mathcal{A}}} \exp(-n\epsilon^2/32).$$

Some implications of the theorem are discussed. For simplicity, assume $V_{\mathcal{A}} > 2$. Then,

$$\left(\frac{en}{V_{\mathcal{A}}}\right)^{V_{\mathcal{A}}} \leq n^{V_{\mathcal{A}}}.$$

First of all, setting the bound $8n^{V_{\mathcal{A}}} \exp(-n\epsilon^2/32)$ to δ and solving for ϵ, we have the following PAC-style result: with probability at least $1 - \delta$, for all $f \in \mathcal{F}$,

$$|R_n(f) - R(f)| \leq 4\sqrt{\frac{2}{n}\left(V_{\mathcal{A}} \log n + \log \frac{8}{\delta}\right)}.$$

Furthermore, since $V_{\mathcal{A}} \log n/n \to 0$, $\sum_{n=V_{\mathcal{A}}}^{\infty} n^{V_{\mathcal{A}}} e^{-n\epsilon^2/32} < \infty$. Verify the statement by examining

$$n^{V_{\mathcal{A}}} \exp(-n\epsilon^2/32) = \exp\{V_{\mathcal{A}} \log n - n\epsilon^2/32\} = \exp\{-n(\epsilon^2/32 - V_{\mathcal{A}} \log n/n)\}.$$

Hence by the Borel-Cantelli lemma,

$$\sup_{f\in\mathcal{F}} |R_n(f) - R(f)| \to 0 \quad \text{w.p.1.}$$

From the fundamental inequality in Theorem 3,

$$R(f_n^*) - \inf_{f\in\mathcal{F}} R(f) \leq 2 \sup_{f\in\mathcal{F}} |R_n(f) - R(f)|.$$

This implies that

$$P\left(R(f_n^*) - \inf_{f\in\mathcal{F}} R(f) > \epsilon\right) \leq 8n^{V_{\mathcal{A}}} e^{-n\epsilon^2/128} \quad \text{for } n \geq V_{\mathcal{A}}.$$

As shown before, for a non-negative Z, $P(Z \geq \epsilon) \leq c\, e^{-2n\epsilon^2}$ for all $\epsilon > 0$, implies

$$EZ \leq \sqrt{\frac{\log(ce)}{2n}}.$$

Hence the estimation error,

$$ER(f_n^*) - \inf_{f\in\mathcal{F}} R(f) \leq \sqrt{\frac{\log(8n^{V_{\mathcal{A}}}e)}{n/128}} = 16\sqrt{\frac{\log(8e) + V_{\mathcal{A}} \log n}{2n}} = O\left(\sqrt{\frac{V_{\mathcal{A}} \log n}{n}}\right).$$

The main message of the V-C inequality is that if the V-C dimension of the class $\mathcal{F}$ is finite, then the estimation error converges to zero at the rate $O\left(\sqrt{\frac{V_{\mathcal{A}} \log n}{n}}\right)$ for all distributions of (X, Y). This result is distribution-free and obtained through a combinatorial characterization of the capacity of $\mathcal{F}$.