VC-Dimension

PkGu

12/19/2021

In my note about the book "High Dimensional Probability", a bound for expected (over samples) excess risk $R(f_n^*) - R(f^*)$, where f_n^* minimizes the empirical risk inside a class $\mathcal{F}$ of Boolean functions and f^* minimizes the true risk, is introduced. Quote:

$$\mathbb{E}[R(f_n^*) - R(f^*)] \leq C \cdot \sqrt{\frac{vc(\mathcal{F})}{n}}$$

This inequality is in fact not far from the uniform bound for expected maximal error of empirical process (I like to call it "uniform LLN via VC-dim") (which is derived from Dudley's inequality, if we track further). Quote:

$$\mathbb{E}\left[\underset{f\in\mathcal{F}}{sup}\left|\frac{1}{n}\sum_{1}^{n} f(X_i) - \mathbb{E}f(X)\right|\right] \leq C \cdot \sqrt{\frac{vc(\mathcal{F})}{n}}$$

By the uniform LLN via VC-dim, there is (the same approach):

$$\mathbb{E}[|R_n(f_n^*) - R(f_n^*)|] \leq C \cdot \sqrt{\frac{vc(\mathcal{F})}{n}}$$

This inequality measures the difference between the "in-sample error" and the "generalization error", by which can assess the generalization behavior of the learning process.

Here another approach can lead to some concentration-style inequality doing the similar thing. I sketch it next. The reference of this part is the book "Foundation of data science" written by Blum, Hopcroft and Kannan.

The general setting is that we have a measure space (Ω, μ), and a class of Boolean functions $\mathcal{F}$. Given a finite subset $\omega_n = \{x_1, ..., x_n\} \in \Omega$, $\mathcal{F}(\omega_n)$ is defined to be $\{f(\omega_n) \in \{0,1\}^n | f \in \mathcal{F}\}$. The growth function $\pi_{\mathcal{F}}(n)$ is defined to be $\underset{\omega_n\in\Omega}{max}\ \#\{\mathcal{F}(\omega_n)\}$. Obviously $\pi_{\mathcal{F}}(n) \leq 2^n$ and especially when $n \leq vc(\mathcal{F})$, $\pi_{\mathcal{F}}(n) = 2^n$.

Here's another version of Sauer's lemma, yielding that after surpassing $vc(\mathcal{F})$, the growth function is bounded by a polynomial of order $vc(\mathcal{F})$:

$$\pi_{\mathcal{F}}(n) \leq \sum_{i=0}^{d} \binom{n}{i}$$

Here is the most important theorem in Page[150]: (forgive my not using Overleaf...)

Theorem. For any ϵ and δ in $(0,1)$, as long as

$$n \geq \frac{8}{\epsilon^2}\left[\log(2\pi_{\mathcal{F}}(2n)) + \log\frac{1}{\delta}\right]$$

for ω_n sampled from Ω with probability measure μ, with probability greater than $(1-\delta)$, we have

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