HDP
(22) VC dimension

@ 滕佳烨

前情回顾

The upper bound of $\mathbb{E} \sup_{t \in T} X_t$

1. Dudley's inequality

$$\mathbb{E} \sup_{t \in T} X_t \leq CK \int_0^\infty \sqrt{\log \mathcal{N}(T, d, \varepsilon)}\, d\varepsilon.$$

2. Generic Chaining Bound with γ_2 (tighter but harder to calculate)

$$\mathbb{E} \sup_{t \in T} X_t \leq CK \gamma_2(T, d).$$

The lower bound of $\mathbb{E} \sup_{t \in T} X_t$

1. Sudakov's minoration inequality

$$\mathbb{E} \sup_{t \in T} X_t \geq c\varepsilon \sqrt{\log \mathcal{N}(T, d, \varepsilon)}.$$

An application: VC dimension

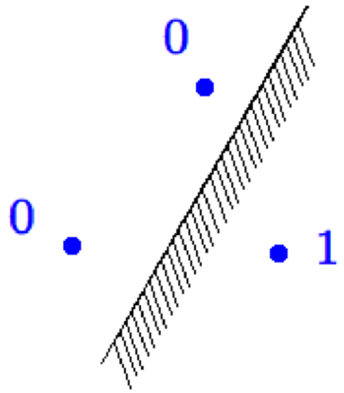
Definition 8.3.1 (VC dimension)**.** Consider a class $\mathcal{F}$ of Boolean functions on some domain Ω. We say that a subset $\Lambda \subseteq \Omega$ is *shattered* by $\mathcal{F}$ if any function $g : \Lambda \to \{0,1\}$ can be obtained by restricting some function $f \in \mathcal{F}$ onto Λ. The *VC dimension* of $\mathcal{F}$, denoted $\mathrm{vc}(\mathcal{F})$, is the largest cardinality[1] of a subset $\Lambda \subseteq \Omega$ shattered by $\mathcal{F}$.

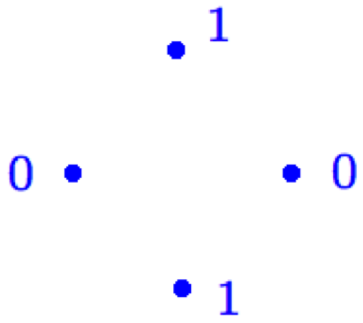
Hint: 一个函数族最多**可以完全**“拟合”多少点

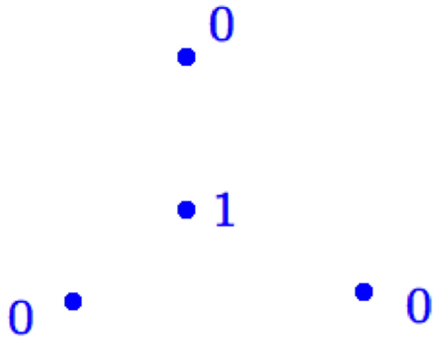
可以：点的位置存在即可

完全：固定点以后label要任意变化

例如：线性函数族可以拟合任意n+1个点







Pajor's Lemma

Lemma 8.3.13 (Pajor's Lemma). *Let $\mathcal{F}$ be a class of Boolean functions on a finite set Ω. Then*

$$|\mathcal{F}| \leq \left|\{\Lambda \subseteq \Omega : \Lambda \textit{ is shattered by } \mathcal{F}\}\right|.$$

We include the empty set $\Lambda = \emptyset$ in the counting on the right side.

$$|\mathcal{F}| \geq 2^{\mathrm{vc}(\mathcal{F})}.$$

Hint: induction.

Remark: although there are many Boolean functions (infinite), we first consider it on finite set, which may produce a small cardinality (at most $2^{|\Omega|}$).

For linear classifiers, there are many hyperplanes in the space, however, when considering two points, the hyperplanes can only produce like 0,1 or 1,0.

Sauer-Shelah Lemma

Theorem 8.3.16 (Sauer-Shelah Lemma). *Let $\mathcal{F}$ be a class of Boolean functions on an n-point set Ω. Then*

$$|\mathcal{F}| \leq \sum_{k=0}^{d} \binom{n}{k} \leq \left(\frac{en}{d}\right)^d$$

where $d = \mathrm{vc}(\mathcal{F})$.

Hint: apply pajor's Lemma with knowledge $|\Lambda| \leq d$.
Intuition: the complexity of $|\mathcal{F}|$ is bounded using VC dim.
Limitation: there can be only finite $|\mathcal{F}|$ (on the finite dataset).

Covering numbers via VC dim

Theorem 8.3.18 (Covering numbers via VC dimension). *Let $\mathcal{F}$ be a class of Boolean functions on a probability space (Ω, Σ, μ). Then, for every $\varepsilon \in (0,1)$, we have*

$$\mathcal{N}(\mathcal{F}, L^2(\mu), \varepsilon) \leq \left(\frac{2}{\varepsilon}\right)^{Cd}$$

where $d = \mathrm{vc}(\mathcal{F})$.

Intuition: the covering number of $\mathcal{F}$ is bounded using VC dim (c.f., dim).
Hint: reduce the domain Ω to Ω_n (n points subset) which requires that with a constant prob, two ϵ-separated functions on Ω (related to packing numbers, also covering numbers of $\mathcal{F}$) are still separated well on Ω_n (related to $|\mathcal{F}_n|$)

VC dim and generalization

Theorem 8.3.23 (Empirical processes via VC dimension)**.** *Let $\mathcal{F}$ be a class of Boolean functions on a probability space (Ω, Σ, μ) with finite VC dimension* $\mathrm{vc}(\mathcal{F}) \geq 1$*. Let $X, X_1, X_2, \ldots, X_n$ be independent random points in Ω distributed according to the law μ. Then*

$$\mathbb{E} \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{n} f(X_i) - \mathbb{E} f(X) \right| \leq C \sqrt{\frac{\mathrm{vc}(\mathcal{F})}{n}}. \tag{8.29}$$

Hint:

Rademacher Complexity ->(Dudley’s ineq) Covering number -> VC dim

Take-away Messages

1. VC dim
   - A way to model the complexity of function class
   - Covering number can be bounded by VC dim $(1/\epsilon)^d$
   - Empirical process can be bounded by VC dim ($\sqrt{d/n}$)

@ 滕佳烨

Thanks!