第八讲

王伟文　暨南大学

1 Rademacher 复杂度上界

给定 n 个样本点 $x_1^n = (x_1, x_2, \ldots, x_n) \subset \mathbb{R}^d$, 考虑函数类 $\mathcal{F} \triangleq \{f \in \mathcal{F} | f : \mathbb{R}^d \to \mathbb{R}\}$, 集合 $\mathcal{F}(x_1^n) = \{(f(x)_1, \ldots, f(x_n)) : f \in \mathcal{F}\} \subset \mathbb{R}^n$ 的元素个数衡量了函数类 $\mathcal{F}$ 样本依赖 (sample-dependent) 的复杂度.

例 1.1. 考虑函数类 $\mathcal{F} \triangleq \{\mathtt{sgn}\left(w^T x + b\right) : w \in \mathbb{R}^d, b \in \mathbb{R}\}$, $F(x_1^n)$ 的元素个数至多为 2^n.

定义 1.1: 多项式判别 (Polynomial Discrimination)

称定义域 $\mathcal{X}$ 上的函数类 $\mathcal{F}$ 具有 $\nu(\geq 1)$ 阶多项式判别, 若对任意 $n \geq 1$ 及任意样本容量为 n 的数据点 $x_1^n = (x_1, \ldots, x_n)$ 均有

$$\mathtt{card}\left(\mathcal{F}(x_1^n)\right) \leq (n+1)^\nu.$$

其中 $\mathtt{card}(\cdot)$ 表示集合元素个数.

评论. 满足多项式判别的函数类样本依赖复杂度随着样本容量多项式增长.

定理 1.1

设函数类 $\mathcal{F}$ 满足 ν 阶多项式判别且 $\mathcal{F} \neq \{0\}$, 则对任意 $n \geq 1$ 及任意样本容量为 n 的数据点 $x_1^n = (x_1, \ldots, x_n)$ 均有

$$\mathbb{E}_\epsilon \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n \epsilon_i f(x_i) \right| \right] \leq 2\mathcal{D}(x_1^n) \sqrt{\frac{\nu \log(n+1)}{n}},$$

其中 $\mathcal{D}(x_1^n) \triangleq \sup_{f \in \mathcal{F}} \sqrt{\frac{\sum_{i=1}^n (f(x_i))^2}{n}}$, ϵ_i 为相互独立的 Rademacher 变量.

为证明上述定理, 我们需要利用 sub-Gaussian 变量最大值的上界引理.

引理 1.1: sub-Gaussian 变量最大值的上界

设 $\{X_i\}_{i\in[n]}$ 为一组均值为 0, 以 σ 为参数的 sub-Gaussian 变量, 则

$$\mathbb{E}[\max_{i\in[n]} X_i] \leq \sqrt{2\sigma^2 \log n} \quad \forall n \geq 1.$$

且

$$\mathbb{E}[\max_{i\in[n]} |X_i|] \leq \sqrt{2\sigma^2 \log(2n)} \leq 2\sqrt{\sigma^2 \log n} \quad \forall n \geq 2.$$

定理 1.1 的证明. 因为 ϵ_i 是以 1 为参数、相互独立的 sub-Gaussian 变量, 故

- $\frac{1}{n} f(x_i)\epsilon$ 是以 $\frac{f(x_i)}{n}$ 为参数且均值为 0 的 sub-Gaussian 变量
- $\sum\limits_{i=1}^{n} \frac{f(x_i)\epsilon_i}{n}$ 是以 $\sqrt{\sum\limits_{i=1}^{n} \frac{(f(x_i))^2}{n^2}}$ 为参数且均值为 0 的 sub-Gaussian 变量.

定义集合

$$\mathcal{S} \triangleq \left\{ \sum_{i=1}^{n} \frac{f(x_i)\epsilon_i}{n} : f \in \mathcal{F} \right\}$$

由引理 1.1

$$\begin{aligned}
\mathbb{E}_\epsilon \left[\sup_{f\in\mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{n} \epsilon_i f(x_i) \right| \right] &= \mathbb{E}\left[\sup_{X\in\mathcal{S}} |X| \right] \\
&\leq 2\sqrt{\sum_{i=1}^{n} \frac{(f(x_i))^2}{n^2}} \sqrt{\log \mathtt{card}(\mathcal{S})}.
\end{aligned}$$

又因为存在映射 $M : \mathcal{F}(x_1^n) \to \mathcal{S}$, 故有 $\mathtt{card}(\mathcal{S}) \leq \mathtt{card}(\mathcal{F}(x_1^n))$. 综上

$$\begin{aligned}
\mathbb{E}_\epsilon \left[\sup_{f\in\mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{n} \epsilon_i f(x_i) \right| \right] &\leq 2\sqrt{\sum_{i=1}^{n} \frac{(f(x_i))^2}{n^2}} \sqrt{\log \mathtt{card}(\mathcal{F}(x_1^n))} \\
&\leq 2\sqrt{\sum_{i=1}^{n} \frac{(f(x_i))^2}{n^2}} \sqrt{\nu \log(n+1)} \\
&= 2\mathcal{D}(x_1^n) \sqrt{\frac{\nu \log(n+1)}{n}}.
\end{aligned}$$

□

评论. 若 $\forall f \in \mathcal{F}$ 有 $\|f\|_\infty \leq b$, 则有

$$\mathcal{R}(\mathcal{F}(x_1^n)) = \mathbb{E}_\epsilon \left[\sup_{f\in\mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{n} \epsilon_i f(x_i) \right| \right] \leq 2b\sqrt{\frac{\nu \log(n+1)}{n}}.$$

定理 1.2: 经典 Glivenko-Cantelli 定理

随机变量 X 的累积分布函数 $F(t)=\mathbb{P}[X\le t]$, $\hat{F}_n(t)$ 表示样本容量为 n 的 $F(t)$ 的经验估计量, 其样本 $X_i(i\in[n])$ 相互独立, 则

$$\mathbb{P}\left[\|\hat{F}_n-F\|_\infty \ge 4\sqrt{\frac{\log(n+1)}{n}}+\delta\right]\le e^{-\frac{n\delta^2}{2}}\quad \forall\delta\ge 0.$$

因此 $\|\hat{F}_n-F\|_\infty \xrightarrow{a.s.} 0$.

证明. 证明思路

- 考虑函数类 $\mathcal{F}=\{\mathbb{I}_{(-\infty,t]}(\cdot):t\in\mathbb{R}\}$, 该函数类为 1 —一致有界, 即 $\forall f\in\mathcal{F}$, $\|f\|_\infty\le 1$.
- $\|\hat{F}_n-F\|_\infty\le 2\mathcal{R}(\mathcal{F}(X_1^n))+\delta$ 至少以概率 $1-\exp(-\frac{n\delta^2}{2})$ 成立.
- $\mathcal{R}(\mathcal{F}(X_1^n))\le 2\sqrt{\frac{\nu\log(n+1)}{n}}$, 若 $\mathcal{F}$ 为 ν-阶多项式判别.
- 需要确定 ν 的值

接下来我们将证明

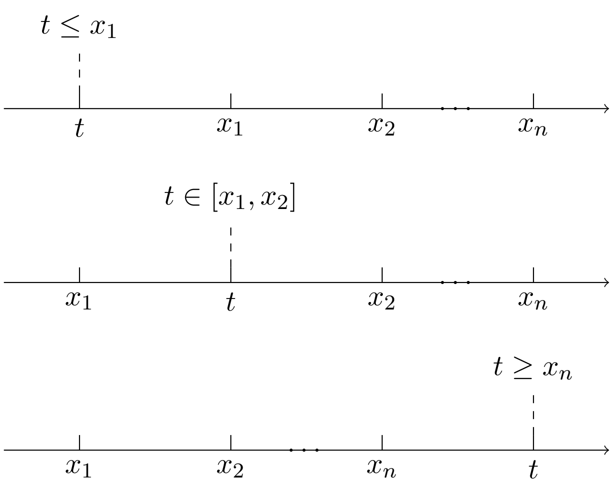
$$\mathtt{card}(\mathcal{F}(x_1^n))\le(n+1)\quad \forall n\ge 1及\forall\{x_i\}_{i\in[n]}\subset\mathbb{R}.$$

根据以上定义有

$$\mathcal{F}(x_1^n)\triangleq\{(\mathbb{I}_{(-\infty,t]}(x_1),\mathbb{I}_{(-\infty,t]}(x_2),\dots,\mathbb{I}_{(-\infty,t]}(x_n))\in\mathbb{R}^n:t\in\mathbb{R}\}$$

不失一般性, 假设 $x_1\le x_2\le\cdots\le x_n$(排序不改变元素数量), 对任意数据点 $\{x_i\}_{i\in[n]}$, 容易计算得到

$$\mathtt{card}(F(x_1^n))\le n+1$$



故 $\mathcal{F}$ 为 1-阶多项式判别. □

2 Vapnik-Chervonenkis(VC) 维

在本节中我们仅考虑二元函数类, 这类函数类的元素个数可以由人们所熟知的 Vapnik-Chervonenkis(VC) 维控制.

定义 2.1: 打散 (Shattering)

考虑二值函数类 $\mathcal{F}$, 给定 n 个数据点集 $x_1^n = (x_1, \ldots, x_n)$, 称它们被 $\mathcal{F}$ 打散若

$$\mathtt{card}(\mathcal{F}(x_1^n)) = 2^n$$

成立.

定义 2.2: Vapnik-Chervonenkis(VC) 维

二值函数类型 $\mathcal{F}$ 的 Vapnik-Chervonenkis(VC) 维是 $\mathcal{F}$ 能打散的最大的数据点数量, 记为 $\nu(F)$. 换言之, 存在 $\nu(\mathcal{F})$ 个数据点 $x_1^{\nu(\mathcal{F})}$ 能够被 $\mathcal{F}$ 打散, 对于任意 $\nu(\mathcal{F}) + 1$ 个数据点, 均不能被 $\mathcal{F}$ 打散.

- 若 $\mathcal{F}$ 为有限值, 则称二值函数类 $\mathcal{F}$ 为 VC 类.

例 2.1. 函数类 $\mathcal{F} \triangleq \{\mathbb{I}_{(-\infty,a)}(\cdot) : a \in \mathbb{R}\}$, 给定一个数据点 x_1, $\mathtt{card}(F(x_1)) = 2$, 对任意给定的两个数据点的集合 $x_1^2 = (x_1, x_2)$, 不妨设 $x_1 < x_2$, 此时 $\mathtt{card}(F(x_1^2)) = 3 < 2^2$. 因此 $\nu(\mathcal{F}) = 1$.

例 2.2. 函数类 $\mathcal{F} \triangleq \{\mathbb{I}_{(a,b]}(\cdot) : a, b \in \mathbb{R}$ 且 $a < b\}$, 任意给定的两个数据点的集合 $x_1^2 = (x_1, x_2)$, $\mathtt{card}(\mathcal{F}(x_1^2)) = 4$, 对于任意 3 个数据点集合 $x_1^3 = (x_1, x_2, x_3)$, 不妨设 $x_1 < x_2 < x_3$, 容易知道 $(1, 0, 1) \notin \mathcal{F}(x_1^3)$, 故 $\mathtt{card}(\mathcal{F}(x_1^3)) < 2^3$, 因此 $\nu(\mathcal{F}) = 2$.

命题 2.1

若二值函数类 $\mathcal{F}_1 \subset \mathcal{F}_2$, 则 $\nu(\mathcal{F}_1) \le \nu(\mathcal{F}_2)$.

引理 2.1: Sauer-Shelah 引理

若二值函数类 $\mathcal{F}$ 有 $\nu(\mathcal{F}) < \infty$, 对于任意样本容量为 n 的数据点 $P = (x_1, x_2, \ldots, x_n)$ 满足 $n \ge \nu(\mathcal{F})$, 有

$$\mathtt{card}(\mathcal{F}(P)) \le \sum_{i=0}^{\nu(\mathcal{F})} \binom{n}{i} \le (n+1)^{\nu(\mathcal{F})}.$$

3 控制 VC 维

定理 3.1

设 $\mathscr{G}$ 为函数 $g:\mathbb{R}^d \to \mathbb{R}$ 的向量空间, 满足 $\mathtt{dim}(\mathscr{G}) < \infty$. 定义集合

$$\mathcal{S}(\mathscr{G}) \triangleq \{\{x \in \mathbb{R}^d : g(x) \le 0\} : g \in \mathscr{G}\},$$

则函数类

$$\mathcal{F} \triangleq \{\mathbb{I}_{\mathsf{S}}(\cdot) : \mathsf{S} \in \mathcal{S}(\mathscr{G})\}$$

的 VC 维至多是 $\mathtt{dim}(\mathscr{G})$.

证明. 这里我们只要证明 $\mathcal{F}$ 不能打散任意 $n := \mathtt{dim}(\mathscr{G}) + 1$ 个数据点.

因为 $\mathscr{G}$ 为函数向量空间, 故对任意函数 $g \in \mathscr{G}$, 存在 $\mathtt{dim}(\mathscr{G})$ 个基函数 $\{\varphi_k\}$ 及向量 $c \in \mathbb{R}^{\mathtt{dim}(\mathscr{G})}$ 使得

$$g = \sum_{k=1}^{\mathtt{dim}(\mathscr{G})} c_k \varphi_k$$

给定任意 n 个数据点集合 $x_1^n = (x_1, x_2, \ldots, x_n)$, 定义矩阵 $\Phi = [\varphi_k(x_i)]_{ik} \in \mathbb{R}^{n \times \mathtt{dim}(\mathscr{G})}$.

因为 $\mathtt{dim}(\mathscr{G}) = n - 1$, 故 $\mathtt{rank}(\Phi) \le n - 1$, 因此存在非零向量 $\gamma \in \mathbb{R}^n$ 使得

$$\Phi^T \gamma = 0$$

又因为

$$\gamma^T \Phi c = \langle \gamma, (g(x_i)) \rangle = 0 \quad \forall g \in \mathscr{G}.$$

不失一般性, 不妨假设 γ 至少存在一个分量大于 0, 则上式可改写为

$$\sum_{i:\gamma_i \le 0} -\gamma_i g(x_i) = \sum_{i:\gamma_i > 0} \gamma_i g(x_i) \quad \forall g \in \mathscr{G}.$$

此恒等式表明 $\forall g \in \mathscr{G}$,

$$\mathtt{sgn}(r_i) = \mathtt{sgn}(g(x_i)) \quad \forall i = 1, 2, \ldots, n$$

不成立.

定义向量 v

$$v_i = \begin{cases} 1, \gamma_i \le 0 \\ 0, \gamma_i > 0 \end{cases}$$

综上所述 $v \notin \mathcal{F}(x_1^n)$.

故 $\mathtt{card}(\mathcal{F}(x_1^n)) < 2^n$, 因此 $\mathcal{F}$ 不能打散任意 n 个数据点. □

例 3.1. 线性函数. 定义线性函数类 $\mathscr{L} \triangleq \{a^T x + b : a \in \mathbb{R}^d, b \in \mathbb{R}\}$, $\texttt{dim}(\mathscr{L}) = d + 1$. 给定 (a, b) 定义半空间 $\mathcal{H}_{a,b} = \{x \in \mathbb{R}^d : a^T x + b \leq 0\}$, 对于函数类

$$\mathcal{F} \triangleq \{\mathbb{I}_{\mathcal{H}_{a,b}}(\cdot) : a \in \mathbb{R}^d, b \in \mathbb{R}\}.$$

由定理 3.1, $\nu(F) \leq d + 1$.

例 3.2. $\mathbb{R}^d$ 中的球体. 考虑球体 $\mathcal{S}_{a,b} = \{x \in \mathbb{R}^d : \|x - a\| \leq b\}$, 其中 $(a, b) \in \mathbb{R}^d \times \mathbb{R}$. 设

$$f_{a,b}(x) = \|x\|^2 - 2\sum_{i=1}^{d} a_i x_i + \|a\|^2 - b^2$$

故 $\mathcal{S}_{a,b} = \{x \in \mathbb{R}^d : f_{a,b}(x) \leq 0\}$

又因为 $f_{a,b}$ 可以改写为

$$f_{a,b}(x) = (1, x_1, x_2, \ldots, x_d, \|x\|^2) \cdot (\|a\|^2 - b^2, a_1, a_2, \ldots, a_d, 1)^T$$

故 $\mathscr{S} = \{f_{a,b} : a \in \mathbb{R}^d, b \in \mathbb{R}\} \subset \mathscr{L} = \{\langle c, \cdot \rangle : c \in \mathbb{R}^{d+2}\}$, 记 $\mathcal{S}_c = \{x \in \mathbb{R}^{d+2} : \langle c, x \rangle \leq 0\}$.

定义函数类

$$\mathcal{F} \triangleq \{\mathbb{I}_{\mathcal{S}_{a,b}}(\cdot) : a \in \mathbb{R}^d, b \in \mathbb{R}\}.$$

$$\mathcal{L} \triangleq \{\mathbb{I}_{\mathcal{S}_c}(\cdot) : c \in \mathbb{R}^{d+2}\}$$

由定理 3.2, $\nu(\mathcal{F}) \leq \nu(\mathcal{L}) \leq \texttt{dim}(\mathscr{L}) = d + 2$.