23/03/21

Գլուխ II Վիճակագրական ուսուցման տեսություն

1) ՄԵԾ ԹՎԵՐԻ ՀԱՎԱՍԱՐԱՉԱՓ ՕՐԵՆՔ

Հիշեցում

$$Z_i = (X_i, Y_i) \overset{iid}{\sim} P^* \quad i = 1, \dots, n$$

$$\mathcal{F}_0 \subset \mathcal{F} = \{ f : \mathcal{X} \to \mathcal{Y} \text{ չափելի} \}$$

$$\hat{f}_n \in \underset{f \in \mathcal{F}_0}{\operatorname{argmin}} \underbrace{\frac{1}{n} \sum_{i=1}^{n} \underbrace{\ell(f(X_i), Y_i)}_{g(Z_i)}}_{\hat{L}_n(f)}$$

$$f_0^* \in \underset{f \in \mathcal{F}_0}{\operatorname{argmin}} \underbrace{\mathbb{E}[\ell(f(X), Y)]}_{L(f)}$$

$$\mathcal{G} = \{ g : \mathcal{Z} \to \mathbb{R} : \exists f \in \mathcal{F}_0 \text{ s.t. } g(z) = \ell(f(x), y) \}$$

THM 1

$$\underbrace{\mathbb{E}\left[L(\hat{f}_n) - L(f_0^*)\right]}_{\text{գնահատման սխալանք}} \leq 2\, \mathbb{E}\left[\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{i=1}^{n} g(Z_i) - \mathbb{E}[g(Z)] \right| \right]$$

$$\leq 4 \underbrace{\mathbb{E}\left[\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i g(Z_i) \right| \right]}_{R_n(\mathcal{G}) \text{ Ռադեմախերի բարդություն}}$$

2) ՎԱՊՆԻԿ-ՉԵՐՎՈՆԵՆԿԻՍԻ ՉԱՓՈՂԱԿԱՆՈՒԹՅՈՒՆ

Բինար դասակարգման դեպքում, Ռադեմախերի բարդությունը կարելի է սահմանափակել (վերևից) VC-չափողականությամբ.

$d_{VC}(\mathcal{G})$ = $\mathcal{G}$-ի Վապնիկ-Չերվոնենկիսի չափողականություն.

Սահմանենք $d_{VC}(\mathcal{G})$-ն.

Բինար դասակարգում՝ $\mathcal{Y} = \{-1, +1\}$ և

$$\ell(y, y') = -y \cdot y' = \begin{cases} +1 & \text{եթե } y \neq y' \\ -1 & \text{եթե } y = y' \end{cases}$$

Սահմ. Կանվանենք $\mathcal{F}_0$-ի աճի ֆունկցիա

$$\Pi_{\mathcal{F}_0}(m) = \max_{x_1, \dots, x_m} \operatorname{Card}\left(\{ (f(x_1), \dots, f(x_m)) : f \in \mathcal{F}_0 \}\right) \quad \square$$

Ակնհայտ է, որ $1 \leq \Pi_{\mathcal{F}_0}(m) \leq 2^m$:

Սահմ. Եթե $\exists m \in \mathbb{N}$ այնպիսին, որ $\Pi_{\mathcal{F}_0}(m) < 2^m$ ապա կասենք, որ $\mathcal{F}_0$-ն ունի վերջավոր VC-չափողականություն:

Սահմ. Եթե $\mathcal{F}_0$-ն ունի վերջավոր VC-չափողականություն, ապա նրա VC-չափողականությունը կանվանենք
$$d_{VC}(\mathcal{F}_0) = \max\{m\in\mathbb{N} : \Pi_{\mathcal{F}_0}(m)=2^m\}. \square$$

Եթե $Card(\mathcal{F}_0)>1$, ապա $d_{VC}(\mathcal{F}_0)\geq 1$

Լեմմ. (Սաուերի լեմմա) Եթե $d_{VC}(\mathcal{F}_0)=d_0$ ապա
$$\Pi_{\mathcal{F}_0}(m)\leq \sum_{i=0}^{d_0}\binom{m}{i}\leq \left(\frac{3m}{d_0}\right)^{d_0}$$
կամայական $m\in\mathbb{N}$-ի համար:
(Ցուցում՝ ինդուկցիա ըստ m-ի:) ■

Դիտարկում. Աճի ֆունկցիան՝ $m\mapsto \Pi_{\mathcal{F}_0}(m)$, էքսպոնենցիալ է $[0, d_{VC}(\mathcal{F}_0)]$ հատվածում, բայց բազմանդամային $[d_{VC}(\mathcal{F}_0), +\infty)$-ում:

$$d_{VC}(\mathcal{F}_0)=d_0<+\infty \Rightarrow \begin{cases}\Pi_{\mathcal{F}_0}(m)=2^m, & m\leq d_0\\ \Pi_{\mathcal{F}_0}(m)\leq\left(\frac{3}{d_0}\right)^{d_0}\cdot m^{d_0}\end{cases}$$

Օրինակներ

① $\mathcal{F}_0=\{\mathbb{1}_{[a;+\infty)} : a\in\mathbb{R}\}$

* $\Pi_{\mathcal{F}_0}(1)=2$ քանի որ եթե $x=0$, և $f_1=\mathbb{1}_{[-1,+\infty)}$, $f_2=\mathbb{1}_{[1,+\infty)}$, ապա $Card\{f_1(x), f_2(x)\}=Card\{1,0\}=2$
* $\Pi_{\mathcal{F}_0}(2)=3<2^2=4$:

Իրոք, դիտարկենք $x_1<x_2$: Բոլոր $f\in\mathcal{F}_0$-երի համար $f(x_1)\leq f(x_2)$: Հետևաբար, $\{(f(x_1),f(x_2)) : f\in\mathcal{F}_0\}$ բազմությունը չի պարունակում $(1,0)$ վեկտորը: Ակնհայտ է, որ այն պարունակում է $(0,0)$, $(0,1)$ և $(1,1)$ վեկտորները:

Հետևաբար՝ $d_{VC}(\mathcal{F}_0)=1$.

② $\mathcal{F}_0=\{\mathbb{1}_{[a;+\infty)}:a\in\mathbb{R}\}\cup\{\mathbb{1}_{(-\infty,a]}:a\in\mathbb{R}\}$

* $\Pi_{\mathcal{F}_0}(2)=4$. Բավական է վերցնել $x_1=0$, $x_2=1$ $f_1=\mathbb{1}_{(-\infty,-1]}$, $f_2=\mathbb{1}_{[1,\infty)}$ $f_3=\mathbb{1}_{(-\infty,0]}$ $f_4=\mathbb{1}_{(-\infty,1]}$

* $\Pi_{\mathcal{F}_0}(3) = 6 < 2^3$ ։ իրոք, $\mathcal{F}_0$-ի բոլոր կետերը ճնշված են։ Իսկապես, $(f(x_1), f(x_2), f(x_3))$-ը չի կարող լինել $(1,0,1)$ կամ $(0,1,0)$, եթե $x_1 < x_2 < x_3$ և $f \in \mathcal{F}_0$։

Հետևությունը՝ $d_{VC}(\mathcal{F}_0) = 2$։

(3) Վարժություն՝ $\mathcal{X} = \mathbb{R}^2$ $\mathcal{F}_0 = \{\mathbb{1}_{[0,+\infty)}(a^T x) : a \in \mathbb{R}^2\}$

Ցույց տալ, որ $d_{VC}(\mathcal{F}_0) = 2$։

Թեորեմ 2 Եթե $\mathcal{F}_0 \subset \{f\ \mathcal{X} \to \{-1,+1\}$ չափելի$\}$ և $d_{VC}(\mathcal{F}_0) = d_0 < +\infty$, ապա

$$R_n(\mathcal{F}_0) \leq \sqrt{\frac{2 d_0 \log(6n/d_0)}{n}}.$$

Հետևաբար նաև

$$\mathbb{E}\left[L(\hat{f}_n) - L(f_0^*)\right] \leq 6\sqrt{\frac{d_0 \log(6n/d_0)}{n}}.$$

Դիտողություն Ավելի բարդ մեթոդներով, կարելի է ցույց տալ, որ

$$R_n(\mathcal{F}_0) \leq C\sqrt{\frac{d_0}{n}}.$$

Թեորեմ 2-ի ապացույցը

Դիտարկենք $A = \{(f(x_1), \dots, f(x_n)) : x_1, \dots, x_n \in \mathcal{X},\ f \in \mathcal{F}_0\}$

Բնականաբար $A \subset \{-1; +1\}^n$։

$$R_n(\mathcal{F}_0) = \mathbb{E}\left[\sup_{f \in \mathcal{F}_0} \left|\frac{1}{n}\sum_{i=1}^n \varepsilon_i f(X_i)\right|\right]$$

$$\leq \mathbb{E}\left[\sup_{a \in A} \left|\frac{1}{n}\sum_{i=1}^n a_i \varepsilon_i\right|\right]$$

որտեղ $a = (a_1, \dots, a_n) \in A \subset \{-1, +1\}^n$։

համաձայն Յենսենի անհավասարության՝

$$R_n(\mathcal{F}_0) = \mathbb{E}\left[\log \exp\left(\lambda \sup_{a \in A}\left|\frac{1}{n}\sum_{i=1}^n a_i \varepsilon_i\right|\right)\right]\frac{1}{\lambda}$$

$$\leq \left(\log \mathbb{E}\left[\exp\left(\sup_{a \in A}\left|\frac{\lambda}{n}\sum_{i=1}^n a_i \varepsilon_i\right|\right)\right]\right)\frac{1}{\lambda}$$

$$\leq \left(\log \mathbb{E}\left[\sup_{a \in A} \exp\left|\frac{\lambda}{n}\sum_{i=1}^n a_i \varepsilon_i\right|\right]\right)\frac{1}{\lambda}$$

$$\leq \left(\log \sum_{a \in A \cup (-A)} \mathbb{E}\left[\exp\left(\frac{\lambda}{n}\sum_{i=1}^n a_i \varepsilon_i\right)\right]\right)\frac{1}{\lambda}$$

$$= \frac{1}{\lambda}\log\left(\sum_a \prod_{i=1}^n \mathbb{E}\left[\exp\left(\frac{\lambda}{n} a_i \cdot \varepsilon_i\right)\right]\right)$$

Հոֆդինգի անհավասարությունը՝

$$\mathbb{E}\left[\exp\left(\frac{\lambda a_i}{n}\cdot\varepsilon_i\right)\right]\leqslant \exp\left(\frac{1}{2}\left(\frac{\lambda a_i}{n}\right)^2\right)$$

$$=\exp\left(\lambda^2/2n^2\right)$$

Հետևաբար՝

$$R_n(\mathcal{F}_0)\leqslant\frac{1}{\lambda}\log\left(\sum_{a\in A\cup(-A)}\prod_{i=1}^{n}\exp\left(\frac{\lambda^2}{2n^2}\right)\right)$$

$$=\frac{1}{\lambda}\log\left(2\cdot \mathrm{Card}(A)\times e^{\lambda^2/2n}\right)$$

$$=\frac{\log 2\,\mathrm{Card}(A)}{\lambda}+\frac{\lambda}{2n}$$

Վերցնելով $\lambda=\sqrt{2n\log 2\,\mathrm{Card}(A)}$

ստանում ենք

$$R_n(\mathcal{F}_0)\leqslant\sqrt{\frac{2\log(2\,\mathrm{Card}(A))}{n}}\ :$$

Ապացույցը վերջացնելու համար, նկատենք, որ $\mathrm{Card}(A)=\Pi_{\mathcal{F}_0}(n)\leqslant\left(\frac{3n}{d_0}\right)^{d_0}$: ☒