Գլուխ 2 Դասական ուսուցման տեսությունը

1) Մեծ թվերի հավասարաչափ օրենք.

$$\mathbb{E}[L(\hat{f}_n) - L(f_0^*)] \le 2R_n(\mathcal{G}):$$

2) VC-չափողականություն

Եթե $\mathcal{Y} = \{-1, +1\}$, $d_{VC}(\mathcal{G}) = d_0 < +\infty$ և $n \ge d_0$, ապա $R_n(\mathcal{F}_0) \le \sqrt{\frac{2 d_0 \log(n/d_0)}{n}}$:

3) ՆՑ-ի VC-չափողականությունը

Հիշեցնենք, որ $(L, d_0, d_1, \dots, d_L)$ կառուցվածքով նեյրոնային ցանցը (ՆՑ) ունի հետևյալ տեսքը՝

$$f(x;\theta) = \sigma_L(W_L \cdots \sigma_1(W_1 x)\dots):$$

Այսուհետ կենթադրենք, որ $\sigma_1, \dots, \sigma_{L-1}$-ը սկալյար ակտիվացիաներ են՝ $\exists \sigma : \mathbb{R} \to \mathbb{R}$ s.t. $\sigma_\ell(x) = (\sigma(x_1), \dots, \sigma(x_d))$ $\forall x \in \mathbb{R}^d$:

Ավելին, կասենք, որ σ-ն կտոր առ կտոր բազմանդամային է, եթե $\exists I_1, \dots, I_K$ ինտերվալների բազմություն

այնպիսին, որ

$$\sigma(x) = \sum_{k=1}^{K} \mathbb{1}(x \in I_k)\, p_k(x),$$

որտեղ $p_1, \dots, p_K$-ն բազմանդամներ են:

Օրինակ՝ $\sigma(x) = \mathbb{1}_{[0;+\infty)}(x)$; $\sigma(x) = \max(0;x)$:

Թեորեմ 4 Նշանակենք $\mathcal{F}_{L,\sigma,p}$-ով բոլոր այն ՆՑ-երի բազմությունը, որոնք ունեն L խորություն, σ-սկալյար ակտիվացում, $d_L = 1$ և $\sigma_L(x) = \mathrm{sign}(x)$: Ճիշտ են հետևյալ պնդումները՝

1. Եթե σ-ն կտոր առ կտոր հաստատուն է, ապա $d_{L,\sigma,p} := d_{VC}(\mathcal{F}_{L,\sigma,p}) = \tilde{\Theta}(p)$:
2. Եթե σ-ն կտոր առ կտոր գծային է, ապա $d_{L,\sigma,p} = \tilde{\Theta}(L \cdot p)$:
3. Եթե σ-ն կտոր առ կտոր բազմանդամային է, ապա $d_{L,\sigma,p} = \tilde{\Theta}(L^2 p)$:

Հիշեցում՝ գրում ենք, որ $a_n = \tilde{O}(b_n)$, եթե $\exists\, C > 0$ և $a > 0$ s.t. $a_n \le C\, b_n \cdot (\log n)^a$

Գրում ենք $a_n = \tilde{\Theta}(b_n)$, եթե $a_n = \tilde{O}(b_n)$ & $b_n = \tilde{O}(a_n)$:

<u>Հետևանք</u>՝ եթե F_0-ն ReLU ակտիվացմամբ, L-խորությամբ և p-պարամետրով ՆՑ է, ապա

$$R_n(F_0) = \tilde{O}\left(\sqrt{\frac{L\cdot p}{n}}\right):$$

Նկատենք, որ այս արդյունքից հետևում է, որ ընդհանրացման սխալանքը ձգտում է 0-ի, եթե $\frac{L\cdot p}{n} \xrightarrow[n\to\infty]{} 0$:

Համանմանորեն, եթե $p \geq n$, ապա $R_n(F_0)$-ն չի ձգտում 0-ի: Նշենք սակայն, որ դրանից չի հետևում, որ

$$\mathbb{E}\left[L(\hat{f}_n) - L(f_0^*)\right] \underset{n\to\infty}{\not\to} 0:$$

4) <u>Իրական արժեք ընդունող ՆՑ</u>

Դուրս գանք բինար դասակարգման խնդիրներից և դիտարկենք այն դեպքը, երբ

- $\mathcal{Y} = [-1, 1]$
- $y \mapsto \ell(y, y')$ ֆունկցիան Լիպշից է $\forall y' \in \mathcal{Y}$

<u>Թեորեմ 5</u> եթե $y \mapsto \ell(y,y')$-ը Լիպշից է, ապա $R_n(\mathcal{G}) \leq 2L_0 \cdot R_n(F_0)$
(L_0-ն Լիպշիցի հաստատունն է, այսինքն $|\ell(y_1,y') - \ell(y_2,y')| \leq L_0 |y_1 - y_2| \ \forall y_1, y_2, y'$)

Թ.5-ը հաճախ անվանում են Լեդու-Թալագրանի սեղմման անհավասարություն:

Օրինակ՝

* $\ell(y,y') = |y-y'| \Rightarrow L_0 = 1$
* $\ell(y,y') = \log(1+e^{-y\cdot y'}) \Rightarrow L_0 = 1$

Նշանակումներ՝

* եթե $a, b \in \mathbb{R}^d$, ապա $a^T b = \langle a, b\rangle = \sum_{i=1}^{d} a_i b_i$ և $\|a\|_2 = (a_1^2 + \ldots + a_d^2)^{1/2}$, $\|a\|_1 = |a_1| + \ldots + |a_d|$:
* եթե W-ն $d_1 \times d_2$ մատրից է, ապա $\|W\|_F = \left(\sum_{i,j} w_{ij}^2\right)^{1/2}$ կոչվում է Ֆրոբենիուսի նորմ:

Հետաքրքիր է, որ այս դեպքում ՆՑ-երի համար կարելի է ստանալ ավելի ճշգրիտ գնահատականներ:

Թեորեմ 6 Եթե $L = 2$ և $\sigma : \mathbb{R} \to [-1,+1]$ այնպիսին է, որ $\sigma(0) = 0$ և $|\sigma(u) - \sigma(v)| \leq |u - v|$ $\forall u, v \in \mathbb{R}$, ապա $\mathcal{X} = [-1,1]^d$ և

$$\mathcal{F}_B = \left\{ f(x,\theta) = \sum_{i=1}^{k} v_i \sigma(w_i^T x) : \|v\|_1 \leq 1, \|w_i\|_1 \leq B \ \forall i \right\}$$

դեպքում, տեղի ունի

$$R_n(\mathcal{F}_B) \leq B \sqrt{\frac{2 \log 2d}{n}} :$$

Եզրակացություն Եթե B-ն փոքր է, ապա $R_n(\mathcal{F}_B)$-ն կլինի փոքր, նույնիսկ եթե k-ն (պարամետրերի քանակը) շատ մեծ է:

Եզրակացություն Թ.6-ը կիրառելի է ReLU ՆՑ-երի համար, բայց կիրառելի չէ 0-1 ակտիվացման դեպքում:

Թեորեմ 7 Եթե σ_i-երը սկալյար ակտիվացումներ են և σ-ն 1-համասեռ է՝ $\sigma(\alpha x) = \alpha \cdot \sigma(x)$ $\forall x \in \mathbb{R}$ և $\forall \alpha \geq 0$, ապա

$$\mathcal{F}_L(B) = \left\{ f(\cdot,\theta) : \|W_1\|_F \leq B \ \& \ldots \|W_L\|_F \leq B \right\}$$

Նեյրոնային ցանցերի դեպքում՝ տեղի ունի

$$R_n(\mathcal{F}_L(B)) \leq C \times \frac{\sqrt{L} \cdot B^L}{\sqrt{n}}$$

անհավասարությունը, որտեղ $C < +\infty$ հաստատուն է:

Եզրակացություն Նույնիսկ խորը ReLU ՆՑ-երի դեպքում, Ռադեմախերի բարդությունը կձգտի 0-ի, եթե բոլոր կշիռները՝ W_ℓ-երը, փոքր են:

Մասնավոր դեպքում, եթե $\|W_\ell\|_F \leq 1$ $\forall \ell = 1, \ldots, L$ ապա $R_n(\mathcal{F}_L(1)) \leq C \times \sqrt{\frac{L}{n}}$:

Հաճախ հանդիպող սխալ

Դիտարկենք $\widehat{B} = \max\limits_{l=1,\dots,L} \|\widehat{W}_l\|_F$,

որտեղ $\widehat{\theta}_n=(\widehat{W}_1,\dots,\widehat{W}_L)$-ը էմպիրիկ սխալանքը մինիմիզացնող պարամետրերի վեկտորն է՝

$$(1) \quad \widehat{\theta}_n \in \arg\min_{\theta} \frac{1}{n}\sum_{i=1}^{n} l\big(f(X_i;\theta);Y_i\big):$$

Սխալ պնդում՝ տեղի ունի

$$\mathbb{E}\big[L(\widehat{f}_n)-L(f_0^*)\big] \leq 4C \times \frac{\sqrt{L}\times \widehat{B}^L}{\sqrt{n}}$$

անհավասարությունը, որտեղ $\widehat{f}_n = f(\cdot,\widehat{\theta}_n)$:

Սխալ ապացույց Դիտարկենք $\widetilde{\theta}_n$-ով

$$(2) \quad \widetilde{\theta}_n \in \arg\min_{\theta:\ \|W_l\|_1\leq \widehat{B}} \frac{1}{n}\sum_{i=1}^{n} l\big(f(X_i,\theta);Y_i\big):$$

Ակնհայտ է, որ $\widetilde{\theta}_n$-ը նաև (1)-ի լուծում է; ինչպես նաև $\widehat{\theta}_n$-ը (2)-ի լուծում է: Ուրեմն

$$\mathbb{E}\big[L(\widehat{f}_n)-L(f_0^*)\big] = \mathbb{E}\big[L(\widetilde{f}_n)-L(f_0^*)\big]$$

$$\leq 4R_n(\mathcal{F}_L(\widehat{B})) \leq 4C\,\frac{\sqrt{L}\times\widehat{B}^L}{\sqrt{n}}$$

սխալ, քանի որ $\widehat{B}$-ն պատահական է