Նկատառում $\hat{\theta}_n^{GD}$-ն և $\hat{\theta}_n^{SGD}$-ն չեն համընկնում (*)-ի լուծում $\hat{\theta}_n$-ի հետ, այլ միայն (լավագույն դեպքում) մոտարկում են այն:

$|L(\hat{\theta}_n^{GD}) - L(\hat{\theta}_n)|$-ը կոչվում է օպտիմիզացիայի սխալանք:

Ամփոփում Երեք տիպի սխալանք՝

1) մոտարկման՝ $L(f(\cdot,\theta^*)) - L(f^*)$
2) գնահատման՝ $L(f(\cdot,\hat{\theta}_n)) - L(f(\cdot,\theta^*))$
3) օպտիմիզացիայի՝ $|L(f(\cdot,\hat{\theta}_n^{GD})) - L(f(\cdot,\hat{\theta}_n))|$.

Մենք հիմնականում կուսումնասիրենք գնահատման սխալանքը:

ԳԼՈՒԽ II ԴԱՍԱԿԱՆ ՈՒՍՈՒՑՄԱՆ ՏԵՍՈՒԹՅՈՒՆ

1) Մեծ թվերի հավասարաչափ օրենք

Ունենք $(X_1,Y_1),\dots,(X_n,Y_n) \overset{iid}{\sim} \mathbb{P}^*$ ($\mathcal{X}\times\mathcal{Y}$-ի վրա)

iid = independent and identically distributed.

$\mathcal{F} = \{f : \mathcal{X} \to \mathcal{Y}$ չափելի$\}$.

$\mathcal{F}_0 \subset \mathcal{F}$ ($\mathcal{F}_0$-ն կարող է լինել ՆՑ-երի բազմությունը)

$l : \mathcal{Y}\times\mathcal{Y} \to \mathbb{R}$ չափելի ֆունկցիա, որը կանվանենք կորստի ֆունկցիա: Օրինակ՝

$l(y,y') = (y-y')^2$ կամ $l(y,y') = |y-y'|$:

Տեսական սխալանք կամ ռիսկ՝

$$L(f) = \mathbb{E}\left[\ell(f(X), Y)\right]$$
$$= \int_{\mathcal{X}\times\mathcal{Y}} \ell(f(x), y)\, d\mathbb{P}^*(x,y)$$

Էմպիրիկ ռիսկ՝

$$\hat{L}_n(f) = \frac{1}{n}\sum_{i=1}^{n} \ell(f(X_i, Y_i))$$
$$= \int_{\mathcal{X}\times\mathcal{Y}} \ell(f(x), y)\, d\hat{\mathbb{P}}_n(x,y)$$

որտեղ $\hat{\mathbb{P}}_n$-ը էմպիրիկ բաշխումն է՝

$$\hat{\mathbb{P}}_n(A) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left((X_i, Y_i)\in A\right):$$

Մեծ թվերի օրենք՝

$$\begin{cases} \hat{\mathbb{P}}_n(A) - \mathbb{P}^*(A) \xrightarrow[n\to\infty]{a.s.} 0 \\ \forall A \text{ չափելի } A\subset \mathcal{X}\times\mathcal{Y} \end{cases}$$

a.s. = almost surely

$$\begin{cases} \hat{L}_n(f) - L(f) \xrightarrow[n\to\infty]{a.s.} 0 \\ \forall f \in \mathcal{F} \end{cases}$$

Նկատառում ՄԹՕ-ից չի հետևում, որ

$$\sup_{f\in\mathcal{F}_0} \left|\hat{L}_n(f) - L(f)\right| \xrightarrow[n\to\infty]{a.s.} 0$$

այսինքն զուգամիտությունը հավասարաչափ չէ։ ▨

Մյուս կողմից, գնահատման սխալանքն է

$$L(\hat{f}_n) - L(f_0^*) \quad \text{որտեղ } f_0^* \in \operatorname*{argmin}_{f\in\mathcal{F}_0} L(f):$$

Հետևաբար

$$L(\hat{f}_n) - L(f_0^*) = L(\hat{f}_n) - \hat{L}_n(\hat{f}_n) \; +$$

$$+ \underbrace{\hat{L}_n(\hat{f}_n) - \hat{L}_n(f_0^*)}_{\leq 0} + \hat{L}_n(f_0^*) - L(f_0^*)$$

$$\leq 2 \sup_{f \in \mathcal{F}_0} \left| \hat{L}_n(f) - L(f) \right| \qquad (1)$$

Նշանակենք $Z_i = (X_i, Y_i)$ և

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

Որտեղից ունենք

$$\hat{L}_n(f) - L(f) = \frac{1}{n} \sum_{i=1}^{n} \left(g(Z_i) - \mathbb{E}[g(Z_i)] \right) \qquad (2)$$

հավասարությունը: Հետևաբար, (1) և (2)-ից ստանում ենք, որ

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

<u>Սահմ.</u> Եթե $Z_1, \dots, Z_n \overset{iid}{\sim} P^*$ և $\varepsilon_1, \dots, \varepsilon_n$ միմյանցից անկախ և $(Z_1, \dots, Z_n)$-ից անկախ են, այնպիսին, որ $P(\varepsilon_i = \pm 1) = 1/2 \ \forall i = 1, \dots, n$, ապա

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

մեծությունը կանվանենք Ռադեմախերի բարդություն:

Նկատառում՝ Եթե $\mathcal{G} \subset \mathcal{G}'$, ապա $R_n(\mathcal{G}) \subset R_n(\mathcal{G}')$:

Նաև $0 \leq R_n(\mathcal{G}) \leq \sup_{g \in \mathcal{G}} \max_{z \in \mathcal{Z}} |g(z)|$:

Թեորեմ 1. Եթե $Im(g) \subset [0,1] \quad \forall g \in \mathcal{G}$,

ապա

$$\frac{1}{2} R_n(\mathcal{G}) - \sqrt{\frac{\log 2}{2n}} \leqslant \mathbb{E}\left[\sup_{g\in\mathcal{G}} \left| \frac{1}{n}\sum_{i=1}^{n} g(Z_i) - \mathbb{E}[g(Z_i)] \right|\right]$$

$$\leqslant 2R_n(\mathcal{G})$$

Նկատառումներ

1) Եթե $R_n(\mathcal{G}) \geqslant 4\sqrt{\frac{\log 2}{2n}}$, ապա նախորդ թեորեմից հետևում է, որ

$$\frac{1}{4} R_n(\mathcal{G}) \leqslant \mathbb{E}\left[\sup_{g\in\mathcal{G}} \left| \frac{1}{n}\sum_{i=1}^{n} \left(g(Z_i) - \mathbb{E}[g(Z_i)]\right) \right|\right]$$

$$\leqslant 2R_n(\mathcal{G}):$$

Հետևաբար, մեծությունում գրված մեր արտահայտությունն ու Ռադեմախերի բարդությունը նույն կարգի են:

2) Ռադեմախերի բարդությունը ըստ n-ի նվազող է՝

$$R_n(\mathcal{G}) \geqslant R_{n+1}(\mathcal{G}) \quad \forall n \in \mathbb{N}: \qquad (3)$$

(3)-ը թողնում եմ որպես վարժություն: Իսկ մենք միայն կապացուցենք ավելի թույլ անհավասարությունը՝

$$R_n(\mathcal{G}) \geqslant R_{2n}(\mathcal{G}): \qquad (4)$$

(4)-ի ապացույց.

$$R_{2n}(\mathcal{G}) = \mathbb{E}\left[\frac{1}{2n}\sup_{g\in\mathcal{G}} \left| \sum_{i=1}^{2n} \varepsilon_i g(Z_i) \right|\right]$$

Ապացուցում

$$\sup_g \left|\sum_{i=1}^{2n} \varepsilon_i g(Z_i)\right| = \sup_g \left|\sum_{i=1}^{n} \varepsilon_i g(Z_i) + \sum_{i=n+1}^{2n} \varepsilon_i g(Z_i)\right|$$
$$\leqslant \sup_g \left(\left|\sum_{i=1}^{n} \varepsilon_i g(Z_i)\right| + \left|\sum_{i=n+1}^{2n} \varepsilon_i g(Z_i)\right| \right)$$
$$\leqslant \sup_g \left|\sum_{i=1}^{n} \varepsilon_i g(Z_i)\right| + \sup_g \left|\sum_{i=n+1}^{2n} \varepsilon_i g(Z_i)\right|$$

Որտեղից հետևում է, որ

$$R_{2n}(\mathcal{G}) \leqslant \frac{1}{2}\left\{ \mathbb{E}\left[\sup_g \left|\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i g(Z_i)\right|\right] + \mathbb{E}\left[\sup_g \left|\frac{1}{n}\sum_{i=n+1}^{2n} \varepsilon_i g(Z_i)\right|\right]\right\}$$
$$= \frac{1}{2}\left\{ R_n(\mathcal{G}) + R_n(\mathcal{G})\right\} = R_n(\mathcal{G}) : \blacksquare$$

Թեորեմ 1-ի համառոտ ապացույց

Դիցուք $Z_1', \dots, Z_n' \overset{iid}{\sim} P^*$ անկախ են $Z_1, \dots, Z_n \sim P^*$-ից:

$$\mathbb{E}\left[\sup_g \left|\frac{1}{n}\sum_{i=1}^{n}\left(g(Z_i) - \mathbb{E}[g(Z_i)]\right)\right|\right] =$$
$$= \mathbb{E}\left[\sup_g \left|\frac{1}{n}\sum_{i=1}^{n}\left(g(Z_i) - \mathbb{E}[g(Z_i')]\right)\right|\right]$$

(պայմանական մաթ. սպասում)

$$= \mathbb{E}\left[\sup_g \left|\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(g(Z_i) - g(Z_i')\right) \middle| Z_1, \dots, Z_n\right]\right|\right]$$

$|\mathbb{E}[X]| \leqslant \mathbb{E}|X|$

$$\leqslant \mathbb{E}\left[\sup_g \mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n}\left(g(Z_i) - g(Z_i')\right)\right| \middle| Z_1, \dots, Z_n\right]\right]$$

$\mathbb{E}[\sup_g X_g] \geqslant \sup_g \mathbb{E}[X_g]$

$$\leqslant \mathbb{E}\left[\mathbb{E}\left[\sup_g \left|\frac{1}{n}\sum_{i=1}^{n}\left(g(Z_i) - g(Z_i')\right)\right| \middle| Z_1, \dots, Z_n\right]\right]$$

$\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$

$$= \mathbb{E}\left[\sup_g \left|\frac{1}{n}\sum_{i=1}^{n}\left(g(Z_i) - g(Z_i')\right)\right|\right]$$
$$= \inf_{\sigma_i \in \{\pm 1\}} \mathbb{E}\left[\sup_g \left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i\left(g(Z_i) - g(Z_i')\right)\right|\right]$$
$$\leqslant \mathbb{E}\left[\sup_g \left|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\left(g(Z_i) - g(Z_i')\right)\right|\right] \leqslant 2R_n(\mathcal{G}) : \blacksquare$$

(g