ՄԵՔ. ՈՒՍՈՒՑՄԱՆ ՄԱԹ. ՀԻՄՈՒՆՔՆԵՐԸ

Ա. ԴԱՆԱՆՅԱՆ
24.02.2022

(1) Հիշեցում

$$\underbrace{(X_1, Y_1), \ldots, (X_n, Y_n), (X_{n+1},}_{\text{դիտարկվող զույգեր}} \underbrace{Y_{n+1}}_{\text{անհայտ}}) \overset{iid}{\sim} \underbrace{P^*}_{\text{անհայտ}}$$

$g : \mathcal{X} \to \mathcal{Y}$ կանխատեսիչ

$$\underbrace{R(g) = \mathbb{E}\left[\ell(Y, g(X))\right]}_{\text{կանխատեսիչի ռիսկ}}, \text{ որտեղ } (X,Y) \sim P^*$$

$g^* \in \underset{g}{\operatorname{argmin}}\, R(g)$ - Բայեսի կանխատեսիչ

* Բինար դասակարգում՝ $\mathcal{Y} = \{0,1\}$, $\ell(y,y') = \mathbb{1}(y \neq y')$

$$R(g) = P(g(X) \neq Y) \;;\; g^*(x) = \mathbb{1}\left(\eta^*(x) \geq \tfrac{1}{2}\right).$$

* Նվազագույն քառակուսիների ռեգրեսիա՝ $\mathcal{Y} = \mathbb{R}$, $\ell(y,y') = (y-y')^2$

$$R(g) = \mathbb{E}\left[(g(X) - Y)^2\right] \;;\; g^*(x) = \eta^*(x) = \mathbb{E}\left[Y \mid X = x\right]:$$

3. Կապը բինար դասակարգման և խտության գնահատման միջև

Հիշեցնենք, որ մենք դիտարկում ենք միայն $\mathcal{X} \subset \mathbb{R}^p$ դեպքը: Դիտարկենք շատ հաճախ հանդիպող մի մասնավոր դեպք, երբ X պատահական մեծության պայմանական բաշխումը Y-ի նկատմամբ ունի խտության ֆունկցիա՝ $\forall A \in \mathcal{B}(\mathbb{R}^p)$

$$P(X \in A \mid Y = 0) = \int_A f_0(x)\,dx \quad P(X \in A \mid Y = 1) = \int_A f_1(x)\,dx :$$

Մեր նպատակն է հասկանալ թե ինչպես է g^*-ը կապված f_0, f_1 զույգի հետ:

<u>Թեորեմ 3</u> Կամայական $x \in \mathcal{X}$ -ի համար, տեղի ունի

$$g^*(x) = 1 \iff f_1(x) \cdot \pi_1 \geq f_0(x)\pi_0$$

որտեղ $\pi_0 = \mathbb{P}(Y=0)$ և $\pi_1 = \mathbb{P}(Y=1)$:

<u>Ապացույց</u> Նշանակենք ν-ով հետևյալ չափը $\mathbb{R}^p \times \{0,1\}$-ի վրա՝

$$\nu = \underbrace{\lambda_p}_{\text{Լեբեգի չափ}} \otimes \underbrace{(\delta_0 + \delta_1)}_{\text{երկու Դիրակի չափերի գումար}}$$

Բայեսի թեորեմը պնդում է, որ եթե (X,Y)-ը պատահական վեկտոր է $\mathcal{X} \times \mathcal{Y}$-ի վրա ($\mathcal{X} = \mathbb{R}^p$, $\mathcal{Y} = \{0,1\}$ մեր դեպքում), ապա (X,Y)-ի համատեղ բաշխման խտությունը գրվում է հետևյալ բանաձևով՝

$$f_{(X,Y)}(x,y) = f_{X|Y}(x|y) \times f_Y(y) = f_{Y|X}(y|x) \times f_X(x):$$

Հետևաբար

$$\eta^*(x) = \mathbb{E}[Y|X=x] = \int_{\mathcal{Y}} y\, f_{Y|X}(y|x)\, \nu(dy)$$

$$= \int_{\mathcal{Y}} y \frac{f_{(X,Y)}(x,y)}{f_X(x)} \nu(dy) = \frac{1}{f_X(x)} \int_{\mathcal{Y}} y\, f_{X|Y}(x,y) f_Y(y)\, \nu(dy)$$

$$= \frac{f_1(x)\pi_1}{f_X(x)}$$

Նույն հաշվարկը ցույց է տալիս, որ

$$1-\eta^*(x) = \int_{\mathcal{Y}} (1-y)\, f_{Y|X}(y|x)\, \nu(dy) = \frac{f_0(x)\pi_0}{f_X(x)} :$$

Հետևաբար,

$$g^*(x)=1 \iff \eta^*(x) \geq \frac{1}{2} \iff \eta^*(x) \geq 1-\eta^*(x)$$

$$\iff \frac{f_1(x)\pi_1}{\cancel{f_X(x)}} \geq \frac{f_0(x)\pi_0}{\cancel{f_X(x)}} \iff f_1(x)\pi_1 \geq f_0(x)\pi_0 . \blacksquare$$

Հիմնվելով այս թեորեմի վրա, կարելի է առաջարկել հետևյալ մոտեցումը՝ կանխատեսիչ կառուցելու համար:

a) Գնահատել π_0 և π_1-ը հետևյալ ձևով՝

$$\hat{\pi}_0 = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(Y_i=0) \qquad \hat{\pi}_1 = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(Y_i=1)$$

(ըստ մեծ թվերի օրենքի՝ $\hat{\pi}_j \xrightarrow[n\to\infty]{h.h.} \pi_j \quad \forall j\in\{0,1\}$)

b) Նշանակել $X_1^0,\dots,X_{n_0}^0$-ով այն X_i-երը, որոնց համապատասխանում են $Y_i=0$ պիտակներ:

c) Նշանակել $X_1^1,\dots,X_{n_1}^1$-ով այն X_i-երը, որոնց համապատասխանում են $Y_i=1$ պիտակներ:

d) $(X_1^0,\dots,X_{n_0}^0)$-ի հիման վրա գնահատել f_0-ն $\hat{f}_0$-ով:

e) $(X_1^1,\dots,X_{n_1}^1)$-ի հիման վրա գնահատել f_1-ը $\hat{f}_1$-ով:

f) Նշանակել $\hat{g}(x) = \mathbb{1}(\hat{f}_1(x)\hat{\pi}_1 \geq \hat{f}_0(x)\hat{\pi}_0)$:

Հարց Խնդրեն լուծել d) և e) կետերում նշված գնահատման խնդիրները: Այս հարցին կանդրադառնանք 3-րդ և 4-րդ դասախոսություններում:

4. Էմպիրիկ ռիսկի մինիմիզացում

Մեկ այլ տարածված մոտեցում՝ օգտագործվող կանխատեսիչ կառուցելու գործում, էմպիրիկ ռիսկի մինիմիզացումն է:

* Էմպիրիկ ռիսկ՝ $\hat{R}_n(g) = \underbrace{\frac{1}{n}\sum_{i=1}^{n} \ell(Y_i, g(X_i))}_{\text{training error}}$

  empirical risk

* ԷՌՄ (ERM)՝ ընտրում ենք թեկնածու կանխատեսիչների մի բազմություն՝ $\mathcal{G}$, և նշանակում

$$\hat{g}_n \in \arg\min_{g \in \mathcal{G}} \frac{1}{n}\sum_{i=1}^{n} \ell(Y_i, g(X_i)). \qquad \text{(ERM)}$$

Այս $\mathcal{G}$ բազմության համար ընտրությունը կարևոր է առնվազն 2 պատճառներով՝

1) (ERM) օպտիմիզացիայի խնդիրը պետք է հնարավոր լինի լուծել:

2) $\hat{g}_n$-ը պետք է ունենա փոքր ռիսկ:

Այս դասախոսությունների ընթացքում կկենտրոնանանք միայն երկրորդ կետի վրա: