ՄԵՐ ՈՒՍՈՒՑՄԱՆ ՄԱԹ. ՀԻՄՈՒՆՔՆԵՐԸ

Ա. ԴԱՂԱԼՅԱՆ
22.02.2022

1) Վերահսկվող ուսուցում (supervised learning)

Սկսենք խնդրի մաթեմատիկական ձևակերպումից: Ենթադրենք մեր տրամադրության տակ ունենք զույգեր՝

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n), \quad n \in \mathbb{N},$$

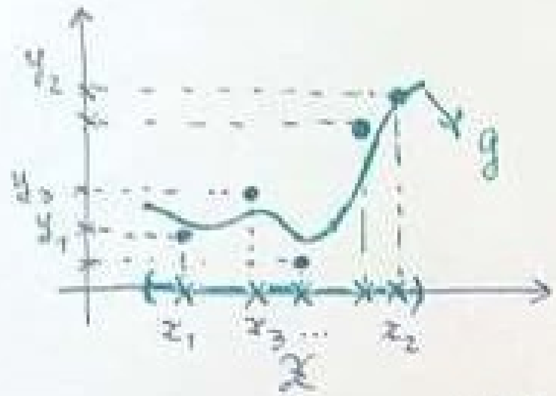
որտեղ յուրաքանչյուր x_i իրենից ներկայացնում է p-չափանի վեկտոր՝ $x_i \in \mathbb{R}^p$ իսկ յուրաքանչյուր y_i (միաչափ) իրական թիվ է՝ $y_i \in \mathbb{R}$: Այս երկու արժեքները՝ x_i-ն և y_i-ն չափումներ են արված նույն առարկայի, օբյեկտի կամ անձի վրա:

Նպատակն է կառուցել ֆունկցիա՝

$$g: \mathcal{X} \to \mathcal{Y} \quad \text{որտեղ} \quad \underbrace{\mathcal{X}}_{\text{բոլոր } x_i\text{-երը պարունակող բազմություն}} \subset \mathbb{R}^p \quad \underbrace{\mathcal{Y}}_{\text{բոլոր } y_i\text{-երը պարունակող բազմություն}} \subset \mathbb{R}$$

այնպիսին, որ երբ մեզ տրամադրեն մի նոր օբյեկտի վրա արված x_{n+1} չափումները, մենք հաշվելով $\hat{y}_{n+1} = g(x_{n+1})$ արժեքը կկանխատեսենք կամ կգուշակենք այդ օբյեկտի վրա արված y_{n+1} չափումը:

Այսուհետ կասենք, որ x_i-երը հատկանիշների (label)
վեկտորներ են (feature vector), y_i-երը պիտակներ են,
$g: \mathcal{X} \to \mathcal{Y}$ չափելի ֆունկցիաները՝ կանխատեսիչներ
(prediction function):



* Եթե $\mathcal{Y} = \{0;1\}$, կասենք բինար դասակարգման խնդիր: (binary classification)
* Եթե $\mathcal{Y} = \mathbb{R}$ կամ $\mathcal{Y} = [a,b]$, կասենք ռեգրեսիայի խնդիր:

Որպեսզի գնահատենք կանխատեսման որակը, պետք է ունենանք «կորստի» ֆունկցիա (loss function)

* $l : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$.
* Եթե $l(y, \hat{y})$-ը փոքր է, ապա $\hat{y}$-ը y-ի ճշգրիտ կանխատեսում է:

Հիմնական ենթադրությունը (Ենթադրաբանական կամ Ենթադրական)

Գոյություն ունի $(\Omega, \mathcal{F}, \mathbb{P})$ հավանականային տարածություն, ~~այնպիսին որ~~ և նրա վրա որոշված $(X_1, Y_1), \dots, (X_n, Y_n), (X_{n+1}, Y_{n+1})$ անկախ և միատեսակ բաշխված (independent and identically distributed, iid) պատահական վեկտորներ, այնպիսին, որ

$\exists \omega \in \Omega$ s.t. $X_i(\omega) = x_i \quad Y_i(\omega) = y_i \quad \forall i \in \{1, \dots, n, n+1\}$

Կանխատեսման խնդիրը՝ գտնել $g: \mathcal{X} \to \mathcal{Y}$ կանխատեսիչ, այնպիսին, որ $\mathbb{E}[l(Y_{n+1}, g(X_{n+1}))]$-ը լինի փոքր:

2) Ռիսկ և Բայեսի կանխորոշիչը

Նշանակենք $R(g) = \mathbb{E}\left[\ell(Y_{n+1}, g(X_{n+1}))\right]$ -ով $g: \mathcal{X} \to \mathcal{Y}$ կանխորոշիչի ռիսկը:

Օրինակներ՝

1) $\underbrace{\ell(y, y') = \mathbb{1}(y \neq y')}_{\text{0-1 կորուստ}}$ $\mathcal{Y} = \{0; 1\}$ բինար դասակարգում

$$R(g) = \mathbb{E}\left[\mathbb{1}(Y_{n+1} \neq g(X_{n+1}))\right] = \mathbb{P}(Y_{n+1} \neq g(X_{n+1})) = \mathbb{P}(\{\omega : Y_{n+1}(\omega) \neq g(X_{n+1}(\omega))\})$$

2) $\underbrace{\ell(y, y') = (y - y')^2}_{\text{քառակուսային կորուստ}}$ $\mathcal{Y} = \mathbb{R}$ Նպատակային փոփոխականների դեպքերում:

$$R(g) = \mathbb{E}\left[(Y_{n+1} - g(X_{n+1}))^2\right]$$

Կարևոր դիտարկում (Նրանց համար ովքեր անցել են չափի տեսություն)

$(X_1, Y_1) : \Omega \longrightarrow \mathcal{X} \times \mathcal{Y} \subset \mathbb{R}^p \times \mathbb{R}$ պատահական մեծությունն ունի բաշխում $(\mathbb{R}^p \times \mathbb{R}, \mathcal{B}(\mathbb{R}^p \times \mathbb{R}))$ չափելի տարածության վրա որոշված: Այդ բաշխումը կնշանակենք $\mathbb{P}_{(X,Y)}$ կամ կարճեցված $\mathbb{P}^*$: Սա նշանակում է, որ

$$\forall A \in \mathcal{B}(\mathbb{R}^p \times \mathbb{R}) \quad \mathbb{P}^*(A) = \mathbb{P}(\{\omega : (X_1(\omega), Y_1(\omega)) \in A\}).$$

Ռիսկը կարող ենք գրել այսպես

$$R(g) = \mathbb{E}\left[\ell(Y_{n+1}, g(X_{n+1}))\right] =$$

$$= \underbrace{\int_{\mathcal{X}\times\mathcal{Y}} \ell(y, g(x))\, d\mathbb{P}^*(x,y)}$$

Լեբեգի ինտեգրալը $\mathbb{P}^*$ հավանականային չափի նկատմամբ:

Սահմ 1 Երբ տրված են ℓ կորստի ֆունկցիան և $\mathbb{P}^*$ բաշխումը (data generating distribution), կանվանենք Բայեսի կանխատեսիչ կամայական $g^*: \mathcal{X} \to \mathcal{Y}$ այնպիսին, որ

$$g^* \in \arg\min_g R(g)$$

$$\iff R(g^*) = \min_g R(g):$$

Դիտարկում

1) g^*-ը կարող է գոյություն չունենալ (երբ մինիմումը հասանելի չէ, այլ ինֆիմում է):

2) կարող է գոյություն ունենալ մի քանի Բայեսի կանխատեսիչ: Բայց նրանք բոլորը մեզ համար կլինեն համարժեք, քանի որ ունեն միևնույն ռիսկը:

Սահմ 2 Այսուհետև հետևյալ ֆունկցիան

$$\eta^*(x) = \mathbb{E}\left[Y_1 \mid X_1 = x\right]$$

կանվանենք ռեգրեսիա կամ ռեգրեսիայի ֆունկցիա:

regression function

Օրինակ Դիտարկենք $X_i = (W_i, S_i) \in [40,70]\times\{0,1\}$ և $Y_i = 100 + W_i + 10\times S_i + \xi_i$ որտեղ ξ_i-ն 0-միջինով Գաուսյան պատահական մեծություն է՝ անկախ (W_i,S_i)-ից:

Այս դեպքում $\forall\, x=(\omega,s)$-ի համար

$$
\begin{aligned}
\eta^*(x) = \eta^*(\omega,s) &= \mathbb{E}\left[Y_1 \mid X_1=(\omega,s)\right] \\
&= \mathbb{E}\left[100+W_1+10S_1+\xi_1 \mid W_1=\omega, S_1=s\right] \\
&= 100 + \mathbb{E}\big[\underbrace{\omega+10s}_{\text{պատահական չէ}} + \xi_1 \mid W_1=\omega, S_1=s\big] \\
&= 100+10s+\omega + \mathbb{E}\left[\xi_1 \mid W_1=\omega, S_1=s\right] \\
&= 100+10s+\omega+\mathbb{E}\left[\xi_1\right] \quad \longleftarrow \text{քանի որ } \xi_1\text{-ը անկախ է } (W_1,S_1)\text{-ից} \\
&= 100+\omega+10s
\end{aligned}
$$

Հետևաբար $\eta^*(x)=\eta^*(\omega,s)=100+\omega+10s$:

(Այս օրինակը համապատասխանում է հետևյալ իրավիճակին՝ ω-քաշով և s-սեռով ունեցող անձի հասակի գուշակում-կանխորոշում: Այստեղ $s=1$-ը համապատասխանում է արական, իսկ $s=0$-ն՝ իգական սեռին:)

Թեորեմ 1. Եթե $\mathcal{Y}=\{0;1\}$ և $\ell(y,y')=\mathbb{1}(y\neq y')$, ապա $g^*(x)=\mathbb{1}(\eta^*(x)\geq 1/2)$-ը Բայեսյան կանխորոշիչն է:

Թեորեմ 2. Եթե $\mathcal{Y} = \mathbb{R}$ և $\ell(y, y') = (y - y')^2$, ապա $g^*(x) = \eta^*(x)$ -ը Բայեսյան կանխորդիչն է:

Դիտողություն

Երկու թեորեմների Բայեսյան կանխորդիչները կիրառելի չեն, քանի որ հիմնված են $\mathbb{P}^*$ բաշխման իմացության վրա: Իրոք, որպեսզի հաշվենք η^*-ը, պետք է իմանանք $\mathbb{P}^*$ բաշխումը: Մեքենայական ուսուցման ինչպես նաև վիճակագրության նպատակն է լավագույն եղանակով փոխարինել $\mathbb{P}^*$-ի իմացությունը $\{(X_i, Y_i); i = 1, \dots, n\}$ նմուշների իմացությամբ:

Թեորեմ 1-ի ապացույց

Կարևոր է նկատել, որ $\forall X \in \mathcal{X}\ \forall Y \in \{0;1\}$

a) $(Y - g(X))^2 = \mathbb{1}(Y \neq g(X)) \quad \forall g: \mathcal{X} \to \{0;1\}$

b) $(\eta^*(X) - g(X))^2 \geqslant (\eta^*(X) - g^*(X))^2 \quad \forall g: \mathcal{X} \to \{0;1\}$

c) $\mathbb{E}[Y \cdot g(X)] = \mathbb{E}[\eta^*(X) \cdot g(X)]$

Վերը նշված a) և c) կետերից հետևում է, որ

$$R(g) \overset{a)}{=} \mathbb{E}\left[(Y - g(X))^2\right] \overset{c)}{=} \mathbb{E}\left[(Y^2 - \eta^*(X)^2)\right] + \mathbb{E}\left[(\eta^*(X) - g(X))^2\right]$$

$$\overset{b)}{\geqslant} \mathbb{E}\left[(Y^2 - \eta^*(X)^2)\right] + \mathbb{E}\left[(\eta^*(X) - g^*(X))^2\right]$$

$$\overset{a)+c)}{=} R(g^*) \quad \blacksquare$$