

برازش، تخمین، نایقینی I

mamwad@mailaps.org

محمد خرمی

برازش یک دسته داده بررسی میشود. مشخصتر، بهینه-سازی پارامترها ی برازش و نایقینی ی این پارامترها، بر اساس نتایج سنجش، بررسی میشود.

0 درآمد

بین یک دسته متغیر X روابط ی برقرار است. این روابط با یک دسته پارامتر A مشخص میشوند:

$$Q(A, X) = 0. \quad (1)$$

اگر X به دقت معلوم باشد، و تعدادِ مئلفها ی Q با تعدادِ مئلفها ی A باشد، به شرطی که معادلهای (1) مستقل باشند (که از این پس فرض میکنم چنین است)، میشود A را حساب کرد. عملن X با سنجش به دست میتاید، پس خطا دارد، و تعدادِ مئلفها ی Q معمولن بیش از تعدادِ مئلفها ی A است. پس اگر مقادارها ی سنجیده برای X به جا ی خُد X به کار رود، ممکن است معادلات (1) برای A جواب نداشته باشند (معمولن چنین است). جواب اگر وجود داشته باشد هم با مقدار واقعی فرق دارد، چون به جا ی X تخمین ی از آن به کار رفته. هدف این است که با داشتن مقادارها ی سنجیده،

تخمین ی برای X و A به دست آید، و انحراف این تخمینها از مقادیر واقعی تخمین زده شود. مقدار ی که با سنجش برای X به دست میآید را با Y نشان میدهم. چگالی-ی-احتمال را هم با pro نشان میدهم. از

$$\begin{aligned} & \text{pro}(Y = y|A = a, X = x) \text{pro}(A = a, X = x) \\ &= \text{pro}(A = a, X = x|Y = y) \text{pro}(Y = y) \end{aligned} \quad (2)$$

یک شکل قضیه ی بیز [1] نتیجه میشود (مثلن [2]):

$$\begin{aligned} \text{pro}(A = a, X = x|Y = y) &= \frac{\text{pro}(A = a, X = x)}{\text{pro}(Y = y)} \\ & \text{pro}(Y = y|A = a, X = x). \end{aligned} \quad (3)$$

مقدار واقعی ی Ω را با $\hat{\Omega}$ نشان میدهم.

1 توزیع احتمال

برای $\text{pro}(A = a, X = x)$ این معلوم است که a و x باید قیدی شبیه (1) را برآورند. پس

$$\text{pro}(A = a, X = x) = g(a, x) \delta[Q(a, x)], \quad (4)$$

که g هموار فرض میشود. متغیرها ی (A, X) را به متغیرها ی (B, B') تغییر میدهم، چنان که $B' = 0$ با $Q = 0$ هم‌مرز است. از جمله میشود B' را خُذ Q گرفت. (A, X) ی که (1) را برآورد چنین میشود

$$A = f_0(B). \quad (5)$$

$$X = f_1(B). \quad (6)$$

عنصر حجم برای (A, X) را هم میشود بر حسب B و B' نوشت. به این ترتیب، با انتگرالگیری از (3) بر B' چگالی ی احتمال برای B به شرط $Y = y$ به دست میآید:

$$\begin{aligned} \text{pro}(B = b|Y = y) &= \frac{g[f_0(b), f_1(b)]}{\text{pro}(Y = y)} h(b) \\ & \text{pro}[Y = y|A = f_0(b), X = f_1(b)], \end{aligned} \quad (7)$$

که h ناشی از عنصر انتگرالگیری و رابطه B' با Q است. تعداد متغیرها (تعداد سنجشها) را با N ، تعداد پارامترها را با m ، و تعداد رابطها بین اینها (تعداد قیدها) را با η نشان میدهیم. تعداد متغیرهای مستقل (تعداد مثلثها B) را هم با ϕ نشان میدهیم. دیده میشود

$$\phi = N + m - \eta. \quad (8)$$

اگر خطای سنجش کم باشد، چگالی احتمال Y به شرط X و A حل X جایگزیده و تیز است. با فرض این که بقیه عبارتها طرف راست (7) نسبت به b کندتغییرند، میشود در آنها به جای b مقدار y را گذاشت که $pro[Y = y | A = f_0(b), X = f_1(b)]$ را بیشینه میکند. به این ترتیب،

$$pro(B = b | Y = y) = \mathcal{N}_1 pro[Y = y | A = f_0(b), X = f_1(b)], \quad (9)$$

که \mathcal{N}_1 یک ضریب بهنجارش است.

وقت ی چگالی احتمال Y به شرط X و A حل X جایگزیده و تیز باشد، میشود آن را با یک گاوسی تقریب کرد:

$$\begin{aligned} pro(Y = y | A = a, X = x) &= \mathcal{N}_2 \exp \left[-\frac{1}{2} C_{ij}^{-1} (y^i - x^i) (y^j - x^j) \right], \\ &= \mathcal{N}_2 \exp \left[-\frac{1}{2} C^{-1} (y - x, y - x) \right] \end{aligned} \quad (10)$$

که \mathcal{N}_2 یک ضریب بهنجارش است و C همبستگی Y است:

$$E\{[Y^i - E(Y^i)][Y^j - E(Y^j)]\} = C^{ij}. \quad (11)$$

E مقدار چشمداشتی ست. به این ترتیب،

$$\begin{aligned} pro(B = b | Y = y) &= \mathcal{N}_3 \exp \left\{ -\frac{1}{2} C_{ij}^{-1} [y^i - f_1^i(b)] [y^j - f_1^j(b)] \right\}, \\ &= \mathcal{N}_3 \exp \left\{ -\frac{1}{2} C^{-1} [y - f_1(b), y - f_1(b)] \right\}, \end{aligned} \quad (12)$$

که \mathcal{N}_3 یک ضریب بهنجارش است.

$pro(B = b | Y = y)$ نسبت به b تیز است. پس میشود آن را با یک گاوسی (نسبت به b) تقریب

کرد:

$$\begin{aligned} \text{pro}(B = \mathbf{b} | Y = \mathbf{y}) &= \mathcal{N}_4 \exp \left\{ -\frac{1}{2} L_{\alpha\beta} [b^\alpha - \bar{B}^\alpha(\mathbf{y})] [b^\beta - \bar{B}^\beta(\mathbf{y})] \right\}, \\ &= \mathcal{N}_4 \exp \left\{ -\frac{1}{2} \mathbf{L} [\mathbf{b} - \bar{\mathbf{B}}(\mathbf{y}), \mathbf{b} - \bar{\mathbf{B}}(\mathbf{y})] \right\} \end{aligned} \quad (13)$$

که \mathcal{N}_4 یک ثابت بهنجارش است، $\bar{\mathbf{B}}(\mathbf{y})$ و \mathbf{L} مستقل از \mathbf{b} اند. دیده میشود $\bar{\mathbf{B}}(\mathbf{y})$ مقداری برای \mathbf{b} است که $\text{pro}(B = \mathbf{b} | Y = \mathbf{y})$ را بیشینه میکند. پس میشود آن را با بیشینه-کردن (12) نسبت به \mathbf{b} حساب کرد:

$$C_{ij}^{-1} \{f_1^i[\bar{\mathbf{B}}(\mathbf{y})] - y^i\} \{(D_\beta f_1^j)[\bar{\mathbf{B}}(\mathbf{y})]\} = 0, \quad (14)$$

که \mathbf{D} مشتقگیری است. بیشینه-کردن نمای گاوسی در (12) نسبت به \mathbf{b} ، هم‌مرز با بیشینه-کردن نمای گاوسی در (10) نسبت به \mathbf{x} (و \mathbf{a}) با قید (1) است. این بیشینه-کردن را میشود بدون حل قید هم انجام داد. به روش ضریبهای لگرانژ [3]، تعریف میکنم

$$F(\mathbf{a}, \mathbf{x}, \boldsymbol{\lambda}) = -\frac{1}{2} C_{ij}^{-1} (y^i - x^i) (y^j - x^j) + \lambda_\nu Q^\nu(\mathbf{a}, \mathbf{x}). \quad (15)$$

شرط فرینه-شدن با قید میشود (مثلن [4])

$$(D_{0,p} F)[\bar{\mathbf{A}}(\mathbf{y}), \bar{\mathbf{X}}(\mathbf{y}), \bar{\boldsymbol{\Lambda}}(\mathbf{y})] = 0, \quad (16)$$

$$(D_{1,k} F)[\bar{\mathbf{A}}(\mathbf{y}), \bar{\mathbf{X}}(\mathbf{y}), \bar{\boldsymbol{\Lambda}}(\mathbf{y})] = 0, \quad (17)$$

$$(D_{2,\rho} F)[\bar{\mathbf{A}}(\mathbf{y}), \bar{\mathbf{X}}(\mathbf{y}), \bar{\boldsymbol{\Lambda}}(\mathbf{y})] = 0, \quad (18)$$

که \mathbf{D}_0 و \mathbf{D}_1 و \mathbf{D}_2 مشتقگیری نسبت به، به ترتیب، \mathbf{a} و \mathbf{b} و $\boldsymbol{\lambda}$ اند. این رابطها میشوند

$$[\bar{\boldsymbol{\Lambda}}_\nu(\mathbf{y})] \{(D_{0,p} Q^\nu)[\bar{\mathbf{A}}(\mathbf{y}), \bar{\mathbf{X}}(\mathbf{y})]\} = 0, \quad (19)$$

$$C_{kj}^{-1} [y^j - \bar{X}^j(\mathbf{y})] + [\bar{\boldsymbol{\Lambda}}_\nu(\mathbf{y})] \{(D_{1,k} Q^\nu)[\bar{\mathbf{A}}(\mathbf{y}), \bar{\mathbf{X}}(\mathbf{y})]\} = 0, \quad (20)$$

$$Q^\rho[\bar{\mathbf{A}}(\mathbf{y}), \bar{\mathbf{X}}(\mathbf{y})] = 0. \quad (21)$$

دیده میشود (16) و (17) و (18)، یا هم‌مرز با آنها (19) و (20) و (21)، به ترتیب m و N و η معادله

اند. رُشن است که

$$\bar{\Omega}(\hat{Y}) = \hat{\Omega}. \quad (22)$$

$$f_1(\hat{B}) = \hat{Y}. \quad (23)$$

از اینجا دیده میشود

$$f_1(\mathbf{b}) - \mathbf{y} = [f_1(\mathbf{b}) - f_1(\hat{B})] - (\mathbf{y} - \hat{Y}). \quad (24)$$

اگر تزیعها تیز باشند، فقط برای انحرافها ی کوچک \mathbf{y} از \hat{Y} و \mathbf{b} از \hat{B} است که چگالی-ی-احتمالها چشمگیر ند. پس،

$$f_1^i(\mathbf{b}) - f_1^i(\hat{B}) = \Gamma_{\alpha}^i (b^{\alpha} - \hat{B}^{\alpha}), \quad (25)$$

که Γ_{α}^i مشتق f_1^i نسبت به b^{α} است، که جا بی نزدیک \hat{B} ، از جمله $\bar{B}(\mathbf{y})$ حساب شده.

با جاگذاری ی (25) و (24) در (12)، نتیجه میشود

$$L_{\alpha\beta} = C_{ij}^{-1} \Gamma_{\alpha}^i \Gamma_{\beta}^j. \quad (26)$$

شرط این که (13) یک چگالی-ی-احتمال (بهنجارش-پذیر) برای B باشد این است که L مثبت-معین باشد. و شرط این-یکی هم آن است که ϕ (تعداد متغیرها ی مستقل) از N (تعداد متغیرها) بیشتر نباشد. اگر ϕ بزرگتر از N باشد، معادلات

$$f_1(\mathbf{b}) = \mathbf{y} \quad (27)$$

برای \mathbf{b} جواب دارند و بُعد مجموعه ی این جوابها $(\phi - N)$ است. در این صورت طرف راست (12) بر این مجموعه ثابت است، که نشان میدهد طرف راست (13) بر این مجموعه ثابت است. پس یک زیرفضا ی $(\phi - N)$ بُعدی هست که اثر L بر آن صفر میشود. از این پس فرض میکنم ϕ کوچکتر از N است (همرئز با آن، m کوچکتر از η است) و L مثبت-معین (و در نتیجه وارون-پذیر) است. وارون L را با K نشان میدهم.

برای تزیعها ی تیز، (14) به این شکل در میآید.

$$C_{ij}^{-1} \{ \Gamma_{\alpha}^i [\bar{B}^{\alpha}(\mathbf{y}) - \hat{B}^{\alpha}] - (y^i - \hat{Y}^i) \} \Gamma_{\beta}^j = 0, \quad (28)$$

که نتیجه میدهد

$$\bar{B}^{\alpha}(\mathbf{y}) - \hat{B}^{\alpha} = K^{\alpha\beta} [C_{ij}^{-1} \Gamma_{\beta}^j (y^i - \hat{Y}^i)]. \quad (29)$$

از جمله دیده میشود

$$\frac{\partial \bar{B}^\alpha}{\partial y^i} = K^{\alpha\beta} C_{ij}^{-1} \Gamma_j^\beta. \quad (30)$$

شاخصها را با C و K بالا-و-پایین میبرم. به این ترتیب،

$$\frac{\partial [\bar{B}^\alpha(\mathbf{y})]}{\partial y^i} = \Gamma_i^\alpha. \quad (31)$$

همچنین،

$$\Gamma_i^\alpha \Gamma^{i\beta} = K^{\alpha\beta}. \quad (32)$$

$$\Gamma_i^\alpha \Gamma_\beta^i = \delta_\beta^\alpha. \quad (33)$$

2 گشتاورها

رابطه ی (11)، وقت ی $E(\mathbf{Y})$ با $\hat{\mathbf{Y}}$ جایگزین شود میشود

$$E[(Y^i - \hat{Y}^i)(Y^j - \hat{Y}^j)] = C^{ij}. \quad (34)$$

به این ترتیب، با استفاده از (29) یا (31)،

$$\begin{aligned} E[(\bar{B}^\alpha - \hat{B}^\alpha)(\bar{B}^\beta - \hat{B}^\beta)] &= \Gamma_i^\alpha \Gamma_j^\beta C^{ij}, \\ &= K^{\alpha\beta}. \end{aligned} \quad (35)$$

رابطه ی (5) و (6) هم برا ی انحرافها ی کم از مقدارها ی واقعی میشوند

$$\bar{A}^p - \hat{A}^p = \Theta^p_\alpha (\bar{B}^\alpha - \hat{B}^\alpha), \quad (36)$$

$$\bar{X}^i - \hat{X}^i = \Gamma^i_\alpha (\bar{B}^\alpha - \hat{B}^\alpha), \quad (37)$$

که Θ^p_α مشتق f_0^i نسبت به b^α است، که جا بی نزدیک \hat{B} (از جمله \bar{B}) حساب شده. به این ترتیب،

$$\bar{A}^p(\mathbf{y}) - \hat{A}^p = \Theta^p_j (y^j - \hat{Y}^j), \quad (38)$$

$$\bar{X}^i(\mathbf{y}) - \hat{X}^i = \Gamma^i_j (y^j - \hat{Y}^j), \quad (39)$$

که

$$\Omega = \Omega \frac{\partial [\bar{B}(y)]}{\partial y}. \quad (40)$$

این ضامن یعنی Θ و Γ ماتریسهای مشتق، به ترتیب، $\bar{A}(y)$ و $\bar{X}(y)$ نسبت به y اند. از (38) و (39) نتیجه میشود

$$E[(\bar{A}^p - \hat{A}^p)(\bar{A}^q - \hat{A}^q)] = \Theta^p_k \Theta^q_k. \quad (41)$$

$$E[(\bar{X}^i - \hat{X}^i)(\bar{X}^j - \hat{X}^j)] = \Gamma^i_k \Gamma^j_k. \quad (42)$$

یا صریحتر،

$$E[(\bar{A}^p - \hat{A}^p)(\bar{A}^q - \hat{A}^q)] = C^{kl} \Theta^p_k \Theta^q_l. \quad (43)$$

$$E[(\bar{X}^i - \hat{X}^i)(\bar{X}^j - \hat{X}^j)] = C^{kl} \Gamma^i_k \Gamma^j_l. \quad (44)$$

و به شکل دیگر،

$$E[(\bar{A}^p - \hat{A}^p)(\bar{A}^q - \hat{A}^q)] = \Theta^p_\alpha \Theta^q_\alpha. \quad (45)$$

$$E[(\bar{X}^i - \hat{X}^i)(\bar{X}^j - \hat{X}^j)] = \Gamma^i_\alpha \Gamma^j_\alpha. \quad (46)$$

یا صریحتر،

$$E[(\bar{A}^p - \hat{A}^p)(\bar{A}^q - \hat{A}^q)] = K^{\alpha\beta} \Theta^p_\alpha \Theta^q_\beta. \quad (47)$$

$$E[(\bar{X}^i - \hat{X}^i)(\bar{X}^j - \hat{X}^j)] = K^{\alpha\beta} \Gamma^i_\alpha \Gamma^j_\beta. \quad (48)$$

3 تخمین گستاورها

از (43) و (44) دیده میشود همبستگی A و نیز همبستگی X به C مربوط است. C هم نماینده Y خطای سنجش است. این خطا را میشود از روی کمیتهای سنجیده تخمین زد. از (24) و (25) و (31) دیده میشود برای انحرافهای کم از مقادیر واقعی

$$\begin{aligned} y^i - f_1^i(\bar{B}) &= y^i - \hat{Y}^i - \Gamma^i_\alpha [\bar{B}^\alpha(y) - \hat{B}^\alpha], \\ &= y^i - \hat{Y}^i - \Gamma^i_\alpha \Gamma_k^\alpha (y^k - \hat{Y}^k). \end{aligned} \quad (49)$$

به این ترتیب،

$$\begin{aligned}
& E\{C^{-1}[Y - f_1(\bar{B}), Y - f_1(\bar{B})]\} \\
&= E\{C_{ij}^{-1} [Y^i - f_1^i(\bar{B})] [Y^j - f_1^j(\bar{B})]\} \\
&= C_{ij}^{-1} \{E[(Y^i - \hat{Y}^i)(Y^j - \hat{Y}^j)] - 2\Gamma_{\alpha}^i \Gamma_k^{\alpha} E[(Y^k - \hat{Y}^k)(Y^j - \hat{Y}^j)] \\
&\quad + \Gamma_{\alpha}^i \Gamma_k^{\alpha} \Gamma_{\beta}^j \Gamma_l^{\beta} E[(Y^k - \hat{Y}^k)(Y^l - \hat{Y}^l)]\}, \\
&= C_{ij}^{-1} [C^{ij} - 2\Gamma_{\alpha}^i \Gamma_k^{\alpha} C^{kj} + \Gamma_{\alpha}^i \Gamma_k^{\alpha} \Gamma_{\beta}^j \Gamma_l^{\beta} C^{kl}], \tag{50}
\end{aligned}$$

که نتیجه میدهد

$$\begin{aligned}
E\{C^{-1}[Y - f_1(\bar{B}), Y - f_1(\bar{B})]\} &= C_i^i - K_{\alpha}^{\alpha}, \\
&= N - \phi, \tag{51}
\end{aligned}$$

یا، با استفاده از (8)،

$$E\{C^{-1}[Y - f_1(\bar{B}), Y - f_1(\bar{B})]\} = \eta - m. \tag{52}$$

از اینجا این تخمین به دست میآید.

$$es\{C^{-1}[\mathbf{y} - \bar{X}(\mathbf{y}), \mathbf{y} - \bar{X}(\mathbf{y})]\} = \eta - m. \tag{53}$$

این در واقع یک رابطه‌ی تخمینی بین متلفها ی C است: اگر

$$C = \sigma^2 C_{\circ}, \tag{54}$$

که σ یک عدد مثبت نامعلوم است و C_{\circ} معلوم است،

$$es(\sigma^2) = \frac{C_{\circ}^{-1}[\mathbf{y} - \bar{X}(\mathbf{y}), \mathbf{y} - \bar{X}(\mathbf{y})]}{\eta - m}. \tag{55}$$

این را در (43) و (44) مینشانم:

$$es\{E[(\bar{A}^p - \hat{A}^p)(\bar{A}^q - \hat{A}^q)]\} = \frac{C_{\circ}^{-1}[\mathbf{y} - \bar{X}(\mathbf{y}), \mathbf{y} - \bar{X}(\mathbf{y})]}{\eta - m} C_{\circ}^{kl} \Theta_k^p \Theta_l^q. \tag{56}$$

$$es\{E[(\bar{X}^i - \hat{X}^i)(\bar{X}^j - \hat{X}^j)]\} = \frac{C_{\circ}^{-1}[\mathbf{y} - \bar{X}(\mathbf{y}), \mathbf{y} - \bar{X}(\mathbf{y})]}{\eta - m} C_{\circ}^{kl} \Gamma_k^i \Gamma_l^j. \tag{57}$$

دیده میشود طرفها ی راست رابطهها ی بالا با ضرب-کردن C در یک عدد تغییر نمیکنند. پس برای تخمین خطای پارامترها و متغیرها مقدار مطلق خطای سنجش لازم نیست.

4 پانوشتها

- [1] Bayes
- [2] Morris H. deGroot & Mark J. Schervish; “probability and statistics” 4th edition (Addison-Wesley, 2002)
- [3] Lagrange
- [4] Tom M. Apostol; “mathematical analysis” 2nd edition (Addison-Wesley, 1975) section 13.7