

Asosiy komponentlar tahlili (Principal component analysis)

R. Maxammadiyev, U. Djumanazarov, Sh. Mahmudov

Ushbu maqoladagi qarashlar mualliflarning shaxsiy fikr va mulohazalari bo'lib, O'zbekiston Respublikasi Markaziy bankining rasmiy pozitsiyasi bilan mos kelmasligi mumkin. O'zbekiston Respublikasi Markaziy banki maqola mazmuniga javobgarlik olmaydi. Taqdim qilingan materiallarni har qanday uslubda qayta ishlatish faqatgina mualliflar ruxsati bilan amalga oshiriladi.

Annotatsiya

Ushbu maqolada indikatorlarning o'lchamini kamaytirish bo'yicha asosiy komponentlar tahlili (principal component analysis) modeli yoritilgan. Asosiy komponentlar tahlilida o'zaro korrelyatsiyaga ega bo'lgan bir nechta indikatorlarning mavjud dispersiyasini maksimal darajada saqlab qolgan holda o'lchamini chiziqli almashtirish orqali kamaytiriladi. Mazkur maqola moliyaviy o'zgarishlar bo'yicha umumiy xulosalar chiqarishda indikatorlar xususiyatlarini maksimal saqlab qolgan holda yagona umumiy ko'rsatkichga birlashtirish amaliyotini ishlab chiqish maqsadida tayyorlangan.

Tayanch so'zlar: determinant, komponent, kovariatsiya matritsasi, korrelyatsiya, vektor.

Asosiy komponentlar tahlili (Principal component analysis)

Asosiy komponentlar tahlili o‘zaro korrelyatsiyaga ega bo‘lgan bir nechta indikatorlarning mavjud dispersiyasini maksimal darajada saqlab qolgan holda ularning o‘lchamini chiziqli almashtirish orqali kamaytirish hisoblanadi¹.

Ko‘p sonli indikatorlarning o‘zgarishiga bir nechta umumiy komponentlar sabab bo‘ladi. Shuningdek, komponentlar va indikatorlar o‘rtasidagi bog‘liqlik quyidagicha ifodalanadi²:

$$\begin{aligned}x_{it} &= \lambda_i F_t + e_{it} \\ x_{it} &= C_{it} + e_{it}\end{aligned}$$

Bu yerda: x_{it} – indikator, λ_i – komponent va indikator o‘rtasidagi munosabat parametri (eigenvalue), F – asosiy komponent (principal component), e_{it} – indikatorga tegishli bo‘lgan dispersiya, C_{it} – modelning umumiy komponenti, i – indikatorlar soni, t – ma’lum davr.

Masalan, asosiy komponentlar tahlili 3 ta shartli indikatorlar (x_1, x_2, x_3) uchun quyidagicha amalga oshiriladi. Ushbu indikatorlar ma’lumotlari dastlab standartlashtirilib³ so‘ngra matritsa ko‘rinishiga keltirib olinadi:

$$A = \begin{vmatrix} X_{11} & X_{21} & X_{31} \\ X_{12} & X_{22} & X_{32} \\ X_{13} & X_{23} & X_{33} \\ X_{14} & X_{24} & X_{34} \\ X_{15} & X_{25} & X_{35} \end{vmatrix}$$

Bu yerda, X_{it} – t davrdagi i indikatorning standartlashtirilgan qiymati.

Ushbu standartlashtirilgan indikatorlar uchun kvadratik⁴ va simmetrik matritsa⁵ xossalriga ega kovariatsiya matritsasi (covariance matrix) quyidagicha hisoblanadi:

¹ James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning.

² Petrovska, M., & Mihajlovska, E. M. (2013). Measures of Financial Stability in Macedonia. Journal of Central Banking Theory and Practice.

³ Indikator qiymatidan jami kuzatuvlar davridagi o‘rtacha qiymati ayirilib standart chetlanishiga bo‘linadi.

⁴ Ustun va qatorlar soni teng bo‘lgan matritsa.

⁵ Matritsa elementlari bosh diagonalga nisbatan simmetrik joylashuvga ega matritsa.

$$\text{Kovariatsiya matritsasi} = \begin{vmatrix} \text{cov}(X_1, X_1) & \text{cov}(X_1, X_2) & \text{cov}(X_1, X_3) \\ \text{cov}(X_2, X_1) & \text{cov}(X_2, X_2) & \text{cov}(X_2, X_3) \\ \text{cov}(X_3, X_1) & \text{cov}(X_3, X_2) & \text{cov}(X_3, X_3) \end{vmatrix}$$

$$\text{cov}(X_i, X_j) = \frac{\sum_{t=1}^T ((X_{it} - \overline{X_{it}}) \times (X_{jt} - \overline{X_{jt}}))}{T-1};$$

Bu yerda, X_{it} – t davrdagi i standartlashtirilgan indikator qiymati, $X_{jt} - t$ davrdagi j standartlashtirilgan indikator qiymati, $\overline{X_{it}}$ va $\overline{X_{jt}}$ – X_i hamda X_j standartlashtirilgan indikatorning jami kuzatuvlar davridagi mos ravishda o'rtacha qiymatlari, T – kuzatuvlar soni.

Keyingi bosqichda kovariatsiya matritsa uchun xos qiymat (eigenvalue) va xos vektorlar (eigenvector) hisoblanadi. Kovariatsiya matritsasining xos qiymat va xos vektorlar bilan quyidagi bog'liqlikka ega:

$$B \cdot v = \lambda \cdot v$$

Bu yerda, B – kovariatsiya matritsasi, v – xos vektor va λ – xos qiymat.

Agar shunday bir λ son tanlash mumkin bo'lsaki, bunda $B \cdot v = \lambda \cdot v$ tenglikni qanoatlantiruvchi har qanday nolmas v vektorga chiziqli almashtirishning xos vektori deyiladi. λ sonning o'ziga esa chiziqli almashtirishning xos vektorga mos keluvchi xos qiymati deyiladi⁶.

Tenglamaning o'ng tarafidagi indikatorlar chap tarafga o'tkazilinadi ($B - \lambda$). $v = 0$ hamda v xos vektorning nolga teng emasligini inobatga olgan holda, quyidagi determinant tenglamasini hisoblash orqali λ qiymatlari aniqlanadi. Shuningdek, λ birlik matritsaga ko'paytiriladi:

$$\det(B - \lambda \cdot I) = 0$$

$$\det \left(\begin{pmatrix} \text{cov}(X_1, X_1) & \text{cov}(X_1, X_2) & \text{cov}(X_1, X_3) \\ \text{cov}(X_2, X_1) & \text{cov}(X_2, X_2) & \text{cov}(X_2, X_3) \\ \text{cov}(X_3, X_1) & \text{cov}(X_3, X_2) & \text{cov}(X_3, X_3) \end{pmatrix} - \lambda \times \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right) = 0$$

$$\det \begin{pmatrix} \text{cov}(X_1, X_1) - \lambda & \text{cov}(X_1, X_2) & \text{cov}(X_1, X_3) \\ \text{cov}(X_2, X_1) & \text{cov}(X_2, X_2) - \lambda & \text{cov}(X_2, X_3) \\ \text{cov}(X_3, X_1) & \text{cov}(X_3, X_2) & \text{cov}(X_3, X_3) - \lambda \end{pmatrix} = 0$$

⁶ Kuttler, K. (2022). Linear Algebra, Theory and Applications.

Yuqoridagi determinantni hisoblash orqali λ aniqlanadi⁷. So‘ngra aniqlangan har bir xos qiymatga tegishli xos vektorlar topiladi. Bunda yuqoridagi kovariatsiya matritsasining xos qiymat va xos vektorlar bilan bog‘liqligini ifodalovchi tenglikdan foydalaniladi:

$$B \cdot v = \lambda \cdot v$$

$$\begin{pmatrix} cov(X_1, X_1) & cov(X_1, X_2) & cov(X_1, X_3) \\ cov(X_2, X_1) & cov(X_2, X_2) & cov(X_2, X_3) \\ cov(X_3, X_1) & cov(X_3, X_2) & cov(X_3, X_3) \end{pmatrix} \times \begin{pmatrix} v_{1i} \\ v_{2i} \\ v_{3i} \end{pmatrix} = \lambda_i \times \begin{pmatrix} v_{1i} \\ v_{2i} \\ v_{3i} \end{pmatrix}$$

$$\begin{cases} cov(X_1, X_1) \times v_{1i} + cov(X_1, X_2) \times v_{2i} + cov(X_1, X_3) \times v_{3i} = \lambda_i \times v_{1i} \\ cov(X_2, X_1) \times v_{1i} + cov(X_2, X_2) \times v_{2i} + cov(X_2, X_3) \times v_{3i} = \lambda_i \times v_{2i} \\ cov(X_3, X_1) \times v_{1i} + cov(X_3, X_2) \times v_{2i} + cov(X_3, X_3) \times v_{3i} = \lambda_i \times v_{3i} \end{cases}$$

Yuqorida keltirilgan tenglamalar sistemasi orqali har bir λ_i ga tegishli xos vektorlar qiymatlari aniqlanadi. Aniqlangan xos vektorlar λ_i qiymatlari kamayib borish tartibi bo‘yicha matritsaga joylashtiriladi, ya’ni birinchi ustunda eng katta λ_i tegishli bo‘lgan xos vektor bo‘lib dastlabki indikatorlar ma’lumotlaridagi eng katta dispersiyani ifodalaydi.

Dastlabki standartlashtirilgan indikatorlar aniqlangan xos vektorlarga ko‘paytirish orqali chiziqli almashtiriladi (linear transformation):

$$\begin{pmatrix} X_{11} & X_{21} & X_{31} \\ X_{12} & X_{22} & X_{32} \\ X_{13} & X_{23} & X_{33} \\ X_{14} & X_{24} & X_{34} \\ X_{15} & X_{25} & X_{35} \end{pmatrix} \times \begin{pmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{pmatrix} = \begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \\ F_{41} & F_{42} & F_{43} \\ F_{51} & F_{52} & F_{53} \end{pmatrix}$$

$$\begin{cases} F_{11} = X_{11} \times v_{11} + X_{21} \times v_{21} + X_{31} \times v_{31} \\ F_{12} = X_{11} \times v_{12} + X_{21} \times v_{22} + X_{31} \times v_{32} \\ F_{13} = X_{11} \times v_{13} + X_{21} \times v_{23} + X_{31} \times v_{33} \\ \dots \\ F_{53} = X_{51} \times v_{13} + X_{52} \times v_{23} + X_{53} \times v_{33} \end{cases}$$

Bu yerda, X – standartlashtirilgan indikatorlar, v – xos vektorlar, F – asosiy komponentlar.

⁷ Determinantning qiymatini nolga tenglab, λ ning qiymati aniqlanadi. Bunda determinantning 3X3 ekanligiga ko‘ra λ ga nisbatan kubli tenglama hosil bo‘ladi va λ ning qiymatlari 3 ta bo‘lishi mumkin ($\lambda = \lambda_i; i = 1,2,3$).

Yuqoridagi chiziqli almashtirishdan hosil bo'lgan yangi qiymatlar to'plami asosiy komponentlar deb ataladi. Bunda birinchi ustundagi asosiy komponentlar to'plami indikatorlar ma'lumotlariga tegishli bo'lgan eng katta dispersiyani ifodalaydi.