Асосий компонентлар таҳлили (Principal component analysis)

Р. Махаммадиев, У. Джуманазаров, Ш. Маҳмудов

Аннотация

Ушбу мақолада индикаторларнинг ўлчамини камайтириш бўйича асосий компонентлар таҳлили (principal component analysis) модели ёритилган. Асосий компонентлар таҳлилида ўзаро корреляцияга эга бўлган бир нечта индикаторларнинг мавжуд дисперсиясини максимал даражада сақлаб қолган ҳолда ўлчамини чизиқли алмаштириш орқали камайтирилади. Мазкур мақола молиявий ўзгаришлар бўйича умумий хулосалар чиқаришда индикаторлар хусусиятларини максимал сақлаб қолган ҳолда ягона умумий кўрсаткичга бирлаштириш амалиётини ишлаб чиқиш мақсадида тайёрланган.

Таянч сўзлар: детерминант, компонент, ковариация матрицаси, корреляция, вектор.

Асосий компонентлар таҳлили
(Principal component analysis)

Асосий компонентлар таҳлили ўзаро корреляцияга эга бўлган бир нечта индикаторларниг мавжуд дисперсиясини максимал даражада сақлаб қолган ҳолда уларнинг ўлчамини чизиқли алмаштириш орқали камайтириш ҳисобланади[1].

Кўп сонли индикаторларнинг ўзгаришига бир нечта умумий компонентлар сабаб бўлади. Шунингдек, компонентлар ва индикаторлар ўртасидаги боғлиқлик қуйидагича ифодаланади[2]:

$$x_{it} = \lambda_i F_t + e_{it}$$
$$x_{it} = C_{it} + e_{it}$$

Бу ерда: x_{it} – индикатор, λ_i – компонент ва индикатор ўртасидаги муносабат параметри (eigenvalue), F – асосий компонент (principal component), e_{it} – индикаторга тегишли бўлган дисперсия, C_{it} – моделнинг умумий компоненти, i – индикаторлар сони, t – маълум давр.

Масалан, асосий компонентлар таҳлили 3 та шартли индикаторлар (x_1, x_2, x_3) учун қуйидагича амалга оширилади. Ушбу индикаторлар маълумотлари дастлаб стандартлаштирилиб[3] сўнгра матрица кўринишига келтириб олинади:

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

Бу ерда, X_{it} – t даврдаги i индикаторнинг стандартлаштирилган қиймати.

[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] Индикатор қийматидан жами кузатувлар давридаги ўртача қиймати айирилиб стандарт четланишига бўлинади.

Ушбу стандартлаштирилган индикаторлар учун квадратик[4] ва симметрик матрица[5] хоссаларига эга ковариация матрицаси (covariance matrix) қуйидагича ҳисобланади:

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

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

Бу ерда, $X_{it} - t$ даврдаги i стандартлаштирилган индикатор қиймати, $X_{jt} - t$ даврдаги j стандартлаштирилган индикатор қиймати, $\overline{X_{\iota t}}$ ва $\overline{X_{\jmath t}} - X_i$ ҳамда X_j стандартлаштирилган индикаторнинг жами кузатувлар давридаги мос равишда ўртача қийматлари, T – кузатувлар сони.

Кейинги босқичда ковариация матрица учун хос қиймат (eigenvalue) ва хос векторлар (eigenvector) ҳисобланади. Ковариация матрицасининг хос қиймат ва хос векторлар билан қуйидаги боғлиқликка эга:

$$B.v = \lambda.v$$

Бу ерда, B – ковариация матрицаси, v – хос вектор ва λ – хос қиймат.

Агар шундай бир λ сон танлаш мумкин бўлсаки, бунда $B.v = \lambda.v$ тенгликни қаноатлантирувчи ҳар қандай нолмас v векторга чизиқли алмаштиришнинг хос вектори дейилади. λ соннинг ўзига эса чизиқли алмаштиришнинг хос векторга мос келувчи хос қиймати дейилади[6].

Тенгламанинг ўнг тарафидаги индикаторлар чап тарафга ўтказилинади $(B - \lambda).v = 0$ ҳамда v хос векторнинг нолга тенг эмаслигини инобатга олган ҳолда, қуйидаги детерминант тенгламасини ҳисоблаш орқали λ қийматлари аниқланади. Шунингдек, λ бирлик матрицага кўпайтирилади:

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

[4] Устун ва қаторлар сони тенг бўлган матрица.

[5] Матрица элементлари бош диагоналга нисбатан симметрик жойлашувга эга матрица.

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

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

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

Юқоридаги детерминантни ҳисоблаш орқали λ аниқланади[7]. Сўнгра аниқланган ҳар бир хос қийматга тегишли хос векторлар топилади. Бунда юқоридаги ковариация матрицасининг хос қиймат ва хос векторлар билан боғлиқлигини ифодаловчи тенгликдан фойдаланилади:

$$B.v = \lambda.v$$

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

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

Юқорида келтирилган тенгламалар системаси орқали ҳар бир λ_iга тегишли хос векторлар қийматлари аниқланади. Аниқланган хос векторлар λ_i қийматлари камайиб бориш тартиби бўйича матрицага жойлаштирилади, яъни биринчи устунда энг катта λ_i тегишли бўлган хос вектор бўлиб дастлабки индикаторлар маълумотларидаги энг катта дисперсияни ифодалайди.

Дастлабки стандартлаштирилган индикаторлар аниқланган хос векторларга кўпайтириш орқали чизиқли алмаштирилади (linear transformation):

[7] Детерминантнинг қийматини нолга тенглаб, λ нинг қиймати аниқланади. Бунда детерминантнинг 3X3 эканлигига кўра λ га нисбатан кубли тенглама ҳосил бўлади ва λ нинг қийматлари 3 та бўлиши мумкин ($\lambda = \lambda_i;\ i = 1,2,3$).

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

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

Бу ерда, X – стандартлаштирилган индикаторлар, v – хос векторлар, F – асосий компонентлар.

Юқоридаги чизиқли алмаштиришдан ҳосил бўлган янги қийматлар тўплами асосий компонентлар деб аталади. Бунда биринчи устундаги асосий компонентлар тўплами индикаторлар маълумотларига тегишли бўлган энг катта дисперсияни ифодалайди.