ذ [section] 0.0

دانشگاه تربیت معلم

دانشکده علوم ریاضی و کامپیوتر

پایان نامه کارشناسی‌ارشد

آمار(گرایش آمار ریاضی)

عنوان

هسته همبستگی کانونی و کاربرد آن

تدوین

وحید شیری

استاد راهنما

دکتر عین اله پاشا

آذر ۱۳۹۰

چکیده

اندازه وابستگی بین دو مجموعه از متغیر های تصادفی مورد علاقه ی بسیاری از کارشناسان آمار است . تحلیل همبستگی کانونی کلاسیک می تواند راه گشا باشد، اما این همبستگی محدود به رابطه خطی است. دراین تحقیق یک روش هسته ناپارامتری و غیرخطی برای مطالعه وابستگی و استقلال دو مجموعه از متغیر ها بیان می شود. این هسته غیر خطی برای تحلیل همبستگی کانونی و برای تحلیل مشخصه غیر خطی نیز بکارمی رود. ما اجرای روش هسته برای همبستگی کانونی را در بدنه ی همبستگی کانونی کلاسیک بوسیله دنباله‌ای از مجمو عه های مستقل در هسته فضای هیلبرت وابسته جای می دهیم. چنین جایگذاری یک راه ساده برای انجام روش هسته برای تحلیل همبستگی کانونی فراهم می کند. آزمایش های عددی و مقایسه با روش های ناپارامتری دیگر نیز نشان داده خواهد شد. هسته همبستگی کانونی برای تعمیم تحلیل همبستگی کانونی کلاسیک برای زمینه غیر خطی آن فراهم می‌کند.چنین تعمیمی ناپارامتریک است. در اینجا به بررسی پیوند هسته همبستگی کانونی به تحلیل متعارف غیر خطی می پردازیم و به کاربرد های هسته همبستگی کانونی در زمینه اندازه وابستگی و کاهش بعد و آزمون استقلال، بدون مفروضات گاوسی، پرداخته می شود.

واژه‌های کلیدی: اندازه وابستگی، تحلیل همبستگی کانونی، روش هسته، هسته تکثیر کننده، آزمون استقلال

رده بندی موضوعی ریاضی (۲۰۱۰): 62H20، 62M99

فهرست مندرجات

پیشگفتار

توصیف و اندازه وابستگی بین دو مجموعه ازمتغیرها مورد علاقه ی بسیاری از محققان بوده است. هتلینگ در سال (۱۹۳۶) برای توصیف ارتباط خطی بین دو مجموعه نظریه ی تحلیل همبستگی کانونی خطی (LCCA) را معرفی کرد. این همبستگی با رابطه خطی سروکار دارد و می خواهد یک دستگاه مختصات متعامد جدید برای هر یک از دو مجموعه تعریف کند. بنابراین دستگاه مختصات جدید برای ماکزیمم کردن همبستگی بهینه است. این دستگاه یک دستگاه خطی از نمونه اصلی آن است، پس همبستگی کانونی فقط برای توصیف رابطه خطی بکار می رود. تحت مفروضات گاوسی این همبستگی همچنین می تواند برای آزمون استقلال تصادفی بین دو مجموعه از متغیر ها بکار رود. اگر داده ها گاوسی نباشد و به صورت متقارن توزیع نشود تمام این کارها نامعتبر خواهد بود. در سال های اخیر از دستگاه هسته استفاده خوبی به عمل آمده که در $LCCA$ نیز از روش هسته استفاده شده است که به $KCCA$ معروف است.

$KCCA$ از دستگاه های یادگیری سرچشمه می گیرد که می تواند حالت خاصی از تحلیل همبستگی کانونی غیر خطی باشد. در فصل اول به تعریف همبستگی کانونی به روش هسته و سپس به تعریف همبستگی کانونی غیر خطی می پردازیم. در اینجا ما ابتدا به مبحث $LCCA$ و سپس به مبحث $KCCA$ و در بخش های بعدی به موضوع پیاده سازی و منظم سازی و انتخاب پارامتر می‌پردازیم. کاربردهای $KCCA$ عبارت است ازاندازه وابستگی غیر خطی، کاهش بعد برای تحلیل مشخصه غیر خطی و آزمون استقلال و در آخر به معرفی مختصر $NLCCA$ می‌پردازیم. همبستگی کانونی به روش هسته حالت خاصی از همبستگی کانونی غیر خطی است.

در فصل ۲ به تعریف هسته و موارد کاربرد آن و همبستگی کانونی به روش هسته می پردازیم و در فصل ۳ الگوریتم منظم سازی و انواع آن را مورد بحث قرار می دهیم و در فصل آخر به همبستگی کانونی غیر خطی و چند مثال از همبستگی کانونی به روش هسته می‌پردازیم و سپس مثال‌هایی با استفاده از نرم افزار R را معرفی می کنیم.

فصل ۱

همبستگی کانونی و متغیر های کانونی

۱.۱ مقدمه

دراین بخش دو مجموعه از متغیر ها با توزیع توام را بررسی می کنیم و همبستگی بین متغیر های یک مجموعه را با متغیر های مجموعه‌دیگر مورد بررسی قرار می دهیم و در فضای هر مجموعه از متغیر ها یک دستگاه مختصات جدید پیدا می کنیم. به طور دقیق در هر مجموعه از متغیر ها یک ترکیب خطی از متغیر ها را در نظر می گیریم که دارای بیشترین همبستگی باشند. چنین ترکیبات خطی اولین مختصات در دستگاه جدید هستند. تحلیل همبستگی کانونی روی همبستگی بین یک ترکیب خطی از متغیر های یک مجموعه و یک ترکیب خطی از متغیر های مجموعه دیگر متمرکز می شود. ابتدا هدف این است که دو ترکیب خطی با بیشترین همبستگی را تعیین کنیم. سپس دو ترکیب خطی را تعیین می کنیم که در میان تمام زوج های نا همبسته با زوج انتخاب شده اول دارای بیشترین همبستگی باشد. این روش تا جایی که دستگاه مختصات جدید به طور کامل مشخص شود ادامه می یابد. شاید علاقمند به پیدا کردن وابستگی بین دو مجموعه از متغیرها با بعد وسیع باشیم در این حالت تعداد کمی از ترکیبات خطی را در نظر می گیریم. در این گونه مجموعه ها ترکیبات خطی را در نظر می گیریم که دارای بیشترین همبستگی باشد. تحلیل

همبستگی کانونی شناخت و کمی کردن رابطه بین دو مجموعه از متغیر ها را برسی می کند. این روش برای اولین بار توسط هتلینگ(۱۹۳۶)بسط داده شد.او مثالی از رابطه سرعت محاسبه وتوان محاسبه برای سرعت خواندن و توان خواندن را بیان می کند. در این بخش همبستگی کانونی و متغیر های کانونی در جامعه تعریف می شود و آزمون استقلال و رتبه ماتریس همبستگی مورد برسی قرار می گیرد

۲.۱ همبستگی کانونی و متغیر های کانونی در جامعه

فرض کنید که بردار تصادفی X با p مولفه دارای ماتریس کوواریانس Σ باشد. در اینجا فرض می‌کنیم E(X)=0. X را به دو زیر بردار با $p_۱$ و $p_۲$ مولفه تقسیم می کنیم

$$X = \begin{pmatrix} X^{(۱)} \\ X^{(۲)} \end{pmatrix}$$

برای راحتی فرض کنیم که $p_۱ \leq p_۲$. ماتریس کوواریانس نیز به صورت زیر باشد

$$\sum = \begin{pmatrix} \Sigma_{۱۱} & \Sigma_{۱۲} \\ \Sigma_{۲۱} & \Sigma_{۲۲} \end{pmatrix}$$

یک ترکیب خطی از مولفه های $x^{(۱)}$ ویک ترکیب خطی از مولفه های $X^{(۲)}$ مانند زیر را در نظر می گیریم

$$U = \alpha' X^{(۱)}, V = \gamma' X^{(۲)}$$

ابتدا ترکیبات خطی از $X^{(۱)}$ و $X^{(۲)}$ را در نظر می گیریم که دارای بیشترین همبستگی باشد. نیاز داریم که α و γ را طوری بسازیم که U وV دارای واریانس واحد باشند

$$۱ = E(U^۲) = E(\alpha' x^{(۱)} x^{(۱)\prime} \alpha) = \alpha' \sum\nolimits_{۱۱} \alpha \qquad (۱)$$

$$۱ = E(V^۲) = E(\gamma' x^{(۲)} x^{(۲)\prime} \gamma) = \gamma' \sum\nolimits_{۲۲} \gamma \qquad (۲)$$

می دانیم که E(U)=0 و به طور مشابه E(V)=0، پس کوواریانس بین U و V به شکل زیر است

$$E(UV) = E(\alpha' x^{(1)} x^{(2)\prime} \gamma) = \alpha' \sum_{12} \gamma \quad (3)$$

می خواهیم α و γ را طوری پیدا کنیم که رابطه (۳) با توجه به رابطه (۱)و(۲) ماکزیمم شود، پس معادله لاگرانژ زیر را تشکیل می دهیم

$$\psi = \alpha' \sum_{12} \gamma - \frac{1}{2}\lambda(\alpha' \sum_{11} \alpha - 1) - \frac{1}{2}\mu(\gamma' \sum_{22} \gamma - 1) \quad (4)$$

λ و μ ضرایب لاگرانژ هستند. حال از رابطه (۴) نسبت به α و γ مشتق می گیریم و برابر صفر قرار می دهیم

$$\frac{\partial \psi}{\partial \alpha} = \sum_{12} \gamma - \lambda \sum_{11} \alpha = 0 \quad (5)$$

$$\frac{\partial \psi}{\partial \gamma} = \sum_{12}' \alpha - \mu \sum_{22} \gamma = 0 \quad (6)$$

با ضرب رابطه (۵) در α' و رابطه (٦) در γ' روابط زیر بدست می آید:

$$\alpha' \sum_{12} \gamma - \lambda \alpha' \sum_{11} \alpha = 0 \quad (7)$$

$$\gamma' \sum_{12}' \alpha - \mu \gamma' \sum_{22} \gamma = 0 \quad (8)$$

از آ نجایی که $\alpha' \sum_{11} \alpha$ و $\gamma' \sum_{22} \gamma$ برابر یک است، این نشان می دهد که

$$\lambda = \mu = \alpha' \sum_{12} \gamma$$

با جایگذاری در(۵) و (٦) می توان بدست آورد:

$$-\lambda \sum_{11} \alpha + \sum_{12} \gamma = 0 \quad (9)$$

$$\sum_{21} \alpha - \lambda \sum_{22} \gamma = 0 \quad (10)$$

می دانیم $\Sigma_{12}' = \Sigma_{21}$ پس معادله بالا را می توان به شکل ماتریس نوشت:

$$\begin{pmatrix} -\lambda \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda \Sigma_{22} \end{pmatrix} \begin{pmatrix} \alpha \\ \gamma \end{pmatrix} = 0 \quad (11)$$

ماتریس سمت چپ باید یکتا باشد که آن

$$\begin{vmatrix} -\lambda\sum_{۱۱} & \sum_{۱۲} \\ \sum_{۲۱} & -\lambda\sum_{۲۲} \end{vmatrix} = ۰ \qquad (۱۲)$$

این دترمینان چند جمله ای از درجه $p_۱ + p_۲ = p$ است. برای نشان دادن این موضوع بسط لاپلاس را بوسیله کهاد $p_۱$ ستون اولیه در نظر بگیرید.

اولین جمله $\left| -\lambda\sum_{۱۱} \right| \left| -\lambda\sum_{۲۲} \right| = (-\lambda)^{p_۱+p_۲} \left| \sum_{۲۲} \right|$

$$\left| \sum_{۱۱} \right|$$

جملات دیگر بسط درجات پایین تری در λ هستند، برای اینکه یک یا بیشتر سطر های هر کهاد در $p_۱$ ستون اولیه شامل λ نیستند. از آنجا که Σ معین مثبت است، $\left| \sum_{۱۱} \right| \left| \sum_{۲۲} \right| \neq ۰$ این نشان می دهد که رابطه (۱۲) یک چند جمله ای از درجه p است و دارای p ریشه است. $\lambda_۱ \geq \lambda_۲ \geq ... \geq \lambda_p$ از رابطه (۷) دیدیم که همبستگی بین $U = \alpha' x^{(۱)}$ و $V = \gamma' x^{(۲)}$ برای بعضی مقادیر α و γ که به ازای مقادیری از از λ که در رابطه (۱۲) صدق می کند برابر است با $\lambda = \alpha' \sum_{۱۲} \gamma$، چون به دنبال بیشترین همبستگی هستیم پس $\lambda = \lambda_۱$ یک جواب برای رابطه (۱۱) به ازای $\lambda = \lambda_۱$ برابر $\gamma^{(۱)}, \alpha^{(۱)}$ است. پس داریم $V_۱ = \gamma^{(۱)\prime} x^{(۲)}, U_۱ = \alpha^{(۱)\prime} x^{(۱)}$ $V_۱, U_۱$ دو ترکیب خطی استاندارد شده (میانگین صفر و واریانس یک) از $X^{(۱)}$ و $X^{(۲)}$ هستند که دارای بیشترین همبستگی هستند. حال دومین ترکیبات خطی از U و V در نظر می گیریم که در میان ترکیبات خطی ناهمبسته با $U_۱$ و $V_۱$ دارای بیشترین همبستگی هستند. این روش را ادامه می دهیم به طوری که در r امین مرحله داریم $U_۱ = \alpha^{(۱)\prime} x^{(۱)}, V_۱ = \gamma^{(۱)\prime} x^{(۲)} \cdots : U_r = \alpha^{(r)\prime} x^{(۱)}, V_r = \gamma^{(r)\prime} x^{(۲)}$ با همبستگی های $\lambda^{(۱)} = \lambda_۱, \lambda^{(۲)} ... \lambda^{(r)}$

۱.۲.۱ قضیه.

$U_{r+۱} = \alpha^{(r+۱)\prime} x^{(۱)}, V_{r+۱} = \gamma^{(r+۱)\prime} x^{(۲)}$ در میان ترکیبات خطی نا همبسته با $U_۱, V_۱, ... U_r, V_r$ دارای بیشترین همبستگی است.

برهان : شرط اینکه U با U_i ناهمبسته باشد این است که

$$\circ = E(UU_i) = \alpha' \sum_{۱۱} \alpha^{(i)} \tag{۱۳}$$

$$\circ = E(UV_i) = \alpha' \sum_{۱۲} \gamma^{(i)} \tag{۱۴}$$

وشرط اینکه V با V_i ناهمبسته باشد این است که

$$\circ = E(VV_i) = \gamma' \sum_{۲۲} \gamma^{(i)} \tag{۱۵}$$

$$\circ = E(VU_i) = \gamma' \sum_{۲۱} \alpha^{(i)} \tag{۱٦}$$

حال می خواهیم $E(U_{r+۱}V_{r+۱})$ را با انتخاب α و γ که در شرط (۱) و (۲) و (۱۳) و (۱۵) به ازای i=1,2,...,r صدق کند را ماکزیمم کنیم.

$$\psi_{r+۱} = \alpha' \sum_{۱۲} \gamma - \frac{۱}{۲}\lambda(\alpha' \sum_{۱۱} \alpha - ۱) - \frac{۱}{۲}\mu(\gamma' \sum_{۲۲} \gamma - ۱) \tag{۱۷}$$
$$+ \sum_{i=۱}^{r} \theta_i \gamma' \sum_{۲۲} + \sum_{i=۱}^{r} \nu_i \alpha' \sum_{۱۱} \alpha^{(i)} \gamma^{(i)}$$

$\lambda, \mu, \nu_۱ ... \nu_r, \theta_۱, ..., \theta_r$ ضرایب لاگرانژهستند. از رابطه بالا نسبت به α و γ مشتق می گیریم

$$\frac{\partial \psi_{r+۱}}{\partial \alpha} = \sum_{۱۲} \gamma - \lambda \sum_{۱۱} \alpha + \sum_{i=۱}^{r} \nu_i \sum_{۱۱} \alpha^{(i)} \tag{۱۸}$$

$$\frac{\partial \psi_{r+۱}}{\partial \gamma} = \sum_{۲۱} \alpha - \mu \sum_{۲۲} \gamma + \sum_{i=۱}^{r} \theta_i \sum_{۲۲} \gamma^{(i)} \tag{۱۹}$$

با ضرب رابطه (۱۸) در $\alpha^{(j)\prime}$ و رابطه (۱۹) در$\gamma^{(j)\prime}$ داریم

$$\circ = \nu_j \; \alpha^{(j)\prime} \sum_{۱۱} \alpha^{(j)} = \nu_j \tag{۲۰}$$

$$\circ = \theta_j \; \gamma^{(j)\prime} \sum_{۲۲} \gamma^{(j)} = \theta_j \tag{۲۱}$$

روابط (۱۸) و (۱۹) به سادگی (۹)، (۱۰) و به طور متناوب (۱۱) است، بنابر این بزرگترین ریشه λ_i را در نظر می گیریم و آن را $\lambda^{(r+1)}$ را می نامیم، یک جواب وجود دارد که به ازای

$i = ۱, ..., r$ در روابط (۱۱)، (۱)، (۲)، (۱۳)، (۱۵) صدق می کند و این جواب $\gamma^{(r+۱)}$ و $\alpha^{(r+۱)}$ است پس داریم $U_{r+۱} = \alpha^{(r+۱)\prime} X^{(۱)}, V_{r+۱} = \gamma^{(r+۱)\prime} X^{(۲)}$

اگر $A = [\alpha^{(۱)}, ..., \alpha^{(m)}], \Gamma'_۱ = [\gamma^{(۱)}, ...\gamma(m)]$، پس داریم $m = \min(p_۱ \le p_۲)$.

$$\Lambda = \begin{pmatrix} \lambda^{(۱)} & ۰ & . & . & . & ۰ \\ ۰ & \lambda^{(۲)} & . & . & . & ۰ \\ . & . & . & & & \\ . & . & . & & & \\ . & . & . & & & \\ ۰ & ۰ & . & . & . & \lambda^{(m)} \end{pmatrix} \qquad (۲۲)$$

مسله بهینه سازی (۱) و (۱۳) به شکل زیر است

$$A' \sum_{۱۱} A = I \qquad (۲۳)$$

از آنجا که $\sum_{۱۱}$ دارای رتبه $p_۱$ و I دارای رتبه m است پس داریم $m \le p_۱$. ما باید نشان دهیم که $m < p_۱$ با استفاده از نقیض توسط نشان دادن اینکه بردار های دیگری در این مورد صدق می کند.

از آنجا که $A' \sum_{۱۱}$، $p_۱ \times m$ است، وجود دارد یک ماتریس $p_۱ \times (p_۱ - m)$ مانند E که $A' \sum_{۱۱} E = ۰$ و به طور مشابه یک ماتریس $p_۲ \times (p_۲ - m)$ مانند F وجود دارد که $\Gamma'_۱ \sum_{۲۲} F = ۰$ ما همچنین داریم $\Gamma'_۱ \sum_{۲۱} E = \Lambda A' \sum_{۱۱} E = ۰$ و $A' \sum_{۱۲} F = \Lambda \Gamma'_۱ \sum_{۲۲} F = ۰$.چون E دارای رتبه $p_۱ - m$ است پس $E' \sum_{۱۱} E$ یکتا نیست(اگر $m < p_۱$) و به طور مشابه $F' \sum_{۲۲} F$ نیز یکتا نیست. بنابراین حداقل یک ریشه از

$$\begin{vmatrix} -\nu E' \sum_{۱۱} E & E' \sum_{۱۲} F \\ F' \sum_{۲۱} E & -\nu F' \sum_{۲۲} F \end{vmatrix} = ۰ \qquad (۲۴)$$

وجود دارد زیرا $\left| F' \sum_{۲۲} F \right| . \left| E' \sum_{۱۱} E \right| \neq ۰$ با استفاده از محاسبات جبری پیشین

دیدیم که بردار های a ،b وجود دارند به طوریکه

$$E'\sum_{12}Fb=\nu E'\sum_{11}Ea \qquad (25)$$

$$F'\sum_{21}Ea=\nu F'\sum_{22}Fb \qquad (26)$$

فرض کنید که $Ea = g$ و $Fb = h$. می خواهیم نشان دهیم که ν,g,h شکل راه حل جدید $\lambda^{(m+1)}, \alpha^{(m+1)}, \gamma^{(m+1)}$ هستند. فرض کنید $\Sigma_{11}^{-1}\Sigma_{12}h = k$ چون $A'\Sigma_{12}Fb = 0$ $A'\Sigma_{11}k =$,k به سطر های $A'\Sigma_{11}$ متعامد است ، بنابراین یک ترکیب خطی از ستون های E است که Ec گفته می شود. برابری $\Sigma_{12}h = \Sigma_{11}k$ را می توان به این صورت نوشت

$$\sum_{12}Fb=\sum_{11}Ec \qquad (27)$$

طرفین رابطه را در E' ضرب می کنیم

$$E'\sum_{12}Fb=E'\sum_{11}Ec \qquad (28)$$

از انجا که $E'\Sigma_{11}E$ یکتا نیست با مقایسه روابط (۲۵) و(۲۸) داریم که $c = \nu a$ و بنابراین $k = \nu g$ پس داریم

$$\sum_{12}h=\nu\ \sum_{11}\ g \qquad (29)$$

، به طور مشابه می توان نشان داد که

$$\sum_{21}g=\nu\ \sum_{22}\ h \qquad (30)$$

بنابراین $\nu = \lambda^{(m+1)}$,$g = \alpha^{(m+1)}$,$h = \gamma^{(m+1)}$ یک جواب دیگر است. اما این خلاف این فرض است که $\lambda^{(m)}$,$\alpha^{(m)}$,$\gamma^{(m)}$ که اخرین جواب است پس $m = p_1$ شرایط بر روی λ,α,γ می تواند به شکل زیر خلاصه شود.

$$A'\sum_{11}A=I \qquad (31)$$

$$A'\sum_{۱۲}\Gamma_۱=\Lambda \qquad (۳۲)$$

$$\Gamma'_۱\sum_{۲۲}\Gamma_۱=I \qquad (۳۳)$$

فرض کنید $\Gamma_۲=(\gamma^{(p_۱+۱)}...\gamma^{(p_۲)})$ یک ماتریس $p_۲\times(p_۲-p_۱)$ است که در روابط زیر صدق می کند.

$$\Gamma'_۲\sum_{۲۲}\Gamma_۱=\circ \qquad (۳۴)$$

$$\Gamma'_۲\sum_{۲۲}\Gamma_۲=I \qquad (۳۵)$$

هر $\Gamma_۲$ می تواند بوسیله ماتریس متعامد $(p_۲-p_۱)\times(p_۲-p_۱)$ تکثیر شود. این ماتریس هر بار یک ستون را تشکیل می دهد. $\gamma^{(p_۱+۱)}$ به $\sum_{۲۲}\Gamma_۱$ بردار متعامد و نرمال شده است، پس $۱=\gamma^{(p_۱+۱)\prime}\sum_{۲۲}\gamma^{(p_۱+۱)}$ و $\gamma^{(p_۱+۲)}$ بردار متعامد به $\sum_{۲۲}(\Gamma_۱\gamma^{(p_۱+۱)})$ است، پس $۱=\gamma^{(p_۱+۲)\prime}\sum_{۲۲}\gamma^{(p_۱+۲)}$ و الی آخر. فرض کنید که $\Gamma=(\Gamma_۱,\Gamma_۲)$، این ماتریس از آنجایی که $\Gamma'\sum_{۲۲}\Gamma=I$ یک ماتریس یکتا نیست. دترمینان زیر را در نظر بگیرید.

$$\begin{vmatrix} A' & \circ \\ \circ & \Gamma'_۱ \\ \circ & \Gamma'_۲ \end{vmatrix}\begin{vmatrix} -\lambda\sum_{۱۱} & \sum_{۱۲} \\ \sum_{۲۱} & -\lambda\sum_{۲۲} \end{vmatrix} \qquad (۳۶)$$

$$\begin{vmatrix} A & \circ & \circ \\ \circ & \Gamma_۱ & \Gamma_۲ \end{vmatrix}=\begin{vmatrix} -\lambda I & \Lambda & \circ \\ \Lambda & -\lambda I & \circ \\ \circ & \circ & -\lambda I \end{vmatrix}$$
$$=(-\lambda)^{p_۲-p_۱}\begin{vmatrix} -\lambda I & \Lambda \\ \Lambda & -\lambda I \end{vmatrix}$$

$$=(-\lambda)^{p_۲-p_۱}|-\lambda I|.|-\lambda I-\Lambda(-\lambda^{-۱})\Lambda|=(-\lambda)^{p_۲-p_۱}|\lambda^۲ I-\Lambda^۲|=(-\lambda)^{p_۲-p_۱}\prod(\lambda^۲-\lambda^{(i)۲})$$

به جز برای بردار ثابت چند جمله ای بالا است:

$$\begin{vmatrix} -\lambda\sum_{۱۱} & \sum_{۱۲} \\ \sum_{۲۱} & -\lambda\sum_{۲۲} \end{vmatrix} \qquad (۳۷)$$

ریشه های (۱۲) برابر ریشه های (۳۶) برابر صفر است، یعنی

$$\lambda = \pm\lambda^{(i)}, i = 1, \cdots, p_1$$

و $\lambda = 0$ پس $(\lambda_1 ... \lambda_p) = (\lambda_1, \cdots, \lambda_{p_1}, 0, 0, 0 \cdots, -\lambda_{p_1}, \cdots, -\lambda_1)$ مجموعه $\lambda^{(i)2}$ مجموعه λ_i^2 است. برای نشان دادن اینکه مجموعه $\{\lambda^{(i)}\}$ مجموعه $\{\lambda_{(i)}\}$ به ازای $i = 1, \cdots, p_1$ است کافی است فقط نشان دهیم که $\lambda^{(i)}$ نا منفی است، پس ما داریم $\lambda_{(i)} = \lambda^{(i)}$. فرض کنید

$$U = \begin{pmatrix} U_1 \\ . \\ . \\ . \\ U_{P_1} \end{pmatrix} = A'X^{(1)}$$

$$V^{(1)} = \begin{pmatrix} V_1 \\ . \\ . \\ . \\ V_{P_1} \end{pmatrix} = \Gamma_1' X^{(2)}$$

$$V^{(2)} = \begin{pmatrix} V_{p_1+1} \\ . \\ . \\ . \\ V_{P_2} \end{pmatrix} = \Gamma_2' X^{(2)}$$

مولفه های U یک مجموعه از متغیر های کانونی هستند و $V = (V^{(1)\prime}\ V^{(2)\prime})'$ مجموعه دیگر از متغیر های کانونی است

$$E \begin{pmatrix} U \\ V^{(1)} \\ V^{(2)} \end{pmatrix} (U' \quad V^{(1)\prime} \quad V^{(2)\prime}) = \begin{pmatrix} I_{p_1} & \Lambda & 0 \\ \Lambda & I_{p_1} & 0 \\ 0 & 0 & I_{p_2} - I_{p_1} \end{pmatrix} \qquad (38)$$

که در آن

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_{p_1} \end{bmatrix} \tag{۳۹}$$

۱.۱ تعریف.

فرض کنید که $X = (X^{(1)\prime}\ X^{(2)\prime})'$، که در آن $x^{(1)}$ دارای p_1 مولفه و $x^{(2)}$ دارای p_2 مولفه است. r امین متغیر کانونی زوج ترکیبات خطی $U_r = \alpha^{(r)\prime}X^{(1)}, V_r = \gamma^{(r)\prime}X^{(2)}$ هستند که هر کدام دارای واریانس واحد و با $r-1$ زوج متغیر کانونی اول نا همبسته است. همبستگی بین این جفت r امین همبستگی کانونی است.

۲.۲.۱ قضیه.

فرض کنید که $X = (X^{(1)\prime}\ X^{(2)\prime})'$ یک بردار تصادفی با ماتریس کوواریانس Σ باشد r امین همبستگی کانونی بین $X^{(2)}$، $X^{(1)}$ برابر r امین ریشه بزرگ برابری (۱۲) است.

ما می توانیم ثابت کنیم که V_1، U_1 دارای بیشترین همبستگی هستند. ترکیبات خطی $a'U = (a'A')X^{(1)}$ و $b'V = (b'\Gamma')X^{(2)}$ بوسیله ی $a'a = 1$ و $b'b = 1$ نرمال شده اند ، از آنجا که Γ،A یکتا نیستند هر بردار α را می توان به شکل Aa و هر بردار γ را می توان به شکل Γb نوشت بنابراین هر ترکیب خطی $\alpha' X^{(1)}, \gamma' X^{(2)}$ را می توان به شکل $a'U$ و $b'V$ نوشت. همبستگی بین آنها برابر

$$a'(\Lambda\ 0)b = \sum_{i=1}^{p_1} \lambda_i a_i b_i \tag{۴۰}$$

فرض کنید که $c_i = \frac{\lambda_i a_i}{\sqrt{\sum(\lambda_i a_i)^2}}$ سپس ماکزیمم $a'(\Lambda\ 0)b = \sqrt{\sum(\lambda_i a_i)^2}\sum c_i b_i$ نسبت به b برای $b_i = c_i$ است، از آنجا که $\sum b_i c_i$ کوسینوس زاویه بین b و$(c_1, \cdots, c_{p_1}, 0, \cdots, 0)$ است پس

(۴۵)برابر

$$\sqrt{\sum \lambda_i^2 a_i^2} = \sqrt{\sum_{i=2}^{p_1} (\lambda_i^2 - \lambda_1^2) a_i^2 + \lambda_1^2} \qquad (۴۱)$$

با در نظر گرفتن $a_i = 0,, \ i = 2, \cdots, p_1$ این رابطه ماکزیمم می شود. بنابراین U_1, V_1 دارای بیشترین همبستگی هستند.

۳.۲.۱ قضیه.

همبستگی کانونی نسبت به تبدیلات $X^{(i)*} = C_i X^i$ ثابت است که در آن $C_i,, i = 1, 2$ یکتا نیست و هر تابع از Σ که ثابت است یک تابع از همبستگی کانونی است.

برهان: رابطه (۱۲) به شکل زیر تغییر می کند.

$$\begin{vmatrix} -\lambda C_1 \sum_{11} C_1' & C_1 \sum_{12} C_2' \\ C_2 \sum_{21} C_1' & -\lambda C_2 \sum_{22} C_2' \end{vmatrix} = \qquad (۴۲)$$

$$\begin{vmatrix} C_1 & 0 \\ 0 & C_2 \end{vmatrix} . \begin{vmatrix} -\lambda \sum_{11} & \sum_{12} \\ \sum_{21} & -\lambda \sum_{22} \end{vmatrix}$$

$$. \begin{vmatrix} C_1' & 0 \\ 0 & C_2' \end{vmatrix}$$

بنابراین ریشه ها تغییر نمی کنند. برعکس، فرض کنید که $f(\Sigma_{11}, \Sigma_{12}, \Sigma_{22})$ بردار تابع مقادیر Σ باشد بنابراین برای تمام C_1 ،C_2 غیر منفرد داریم

$$f(C_1 \sum_{11} C_1', C_1 \sum_{12} C_2', C_2 \sum_{22} C_2') = f(\sum_{11}, \sum_{12}, \sum_{22})$$

اگر $C_1' = A$ و $C_2 = \Gamma'$ پس رابطه (۴۲) برابر رابطه (۳۶)است که فقط در همبستگی کانونی

وابسته است پس داریم $f = f(I, (\Lambda \quad 0), I)$

فرض کنید که V ،U دو متغیر تصادفی بامیانگین صفر و واریانس های σ_u^2, σ_v^2 و همبستگی ρ