Lite Linjär Algebra

Avsnittet om kovariansmatriser och flerdimensionella normalfördelningar har starka kopplingar till grundkurserna i linjär algebra och detta är ett litet försök att belysa detta och dessutom kanske friska upp kunskaperna i linjär algebra.

Kovariansmatrisen C för den n-dimensionella stokastiska variabeln $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)^T$ är ju en nxn-matris med elementen $c_{ij} = C(X_i, X_j),\ i, j = 1, 2, \cdots, n$. Eftersom $C(X_i, X_j) = C(X_j, X_i)$ så inser man direkt att C är en symmetrisk matris. Från den linjära algebran kommer man kanske ihåg att symmetriska matriser kan diagonaliseras, dvs man kan finna ett rätvinkligt koordinatsystem (ON-system) där den linjära avbildningen $C : R^n \to R^n$ definierad av $\boldsymbol{x} \to C\boldsymbol{x}$ beskrivs av en diagonalmatris D där

$$D = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & 0 & \cdots & \lambda_n \end{pmatrix} = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$$

Ortogonala matriser U har egenskapen $UU^T = U^T U = I$ dvs att transponatet utgör invers. Dessa matriser utför byten mellan olika ortonormerade (ON) koordinatsystem och svarar i princip mot rotationer och speglingar. Diagonaliseringen innebär att man kan finna U och D (diagonalmatris) så att $C = UDU^T$ (eller $D = U^T CU$) där D har ovanstående utseende. Värdena $\lambda_1, \lambda_2, \cdots, \lambda_n$ kallas egenvärden och mot dessa svarar egenvektorer $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n$ sådana att $C\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i$. För dessa vektorer har alltså C (sedd som linjär avbildning) ett mycket enkelt utseende - en ren förlängning/förkortning och eventuell omkastning. Egenvektorer som hör till olika egenvärden är ortogonala. De kan väljas så att de utgör ett ON-system, dvs att de utgör basvektorer. Om flera egenvärden sammanfaller finns en godtycklighet i valen av motsvarande egenvektorer, men de kan väljas ortogonala.

Att C kan skrivas som $C = UDU^T$ kan tolkas på följande sätt sett från höger till vänster. U^T genomför koordinatbytet till ON-systemet av egenvektorer där avbildningen har utseendet D och U transformerar tillbaka till det ursprungliga koordinatsystemet.

Påstående: Om C är en kovariansmatris är alla egenvärden $\lambda_i \geq 0$.

Kommentar: Man kallar då C en positivt semi-definit matris. Är egenvärdena > 0 kallas den positivt definit och den är då också inverterbar.

Bevis: Tag $\boldsymbol{b} = (b_1, b_2, \cdots, b_n)^T$ godtycklig och bilda $Y = \boldsymbol{b}^T\boldsymbol{X} = \sum_1^n b_i X_i$. Vi har då eftersom varianser är icke-negativa och enligt momentsatsen

$$0 \leq V(\boldsymbol{b}^T\boldsymbol{X}) = \boldsymbol{b}^T C\boldsymbol{b} = \boldsymbol{b}^T UDU^T\boldsymbol{b} = (U^T\boldsymbol{b})^T D(U^T\boldsymbol{b})$$

där vi också utnyttjat att $(AB)^T = B^T A^T$ - observera den omkastade ordningen! Vi kan nu eftersom $\boldsymbol{b}$ var godtycklig skriva detta som

$$0 \leq \boldsymbol{a}^T D\boldsymbol{a} = \sum_1^n \lambda_i a_i^2$$

för godtyckliga $\boldsymbol{a}$-vektorer (tag $\boldsymbol{a} = U^T\boldsymbol{b}$ dvs $\boldsymbol{b} = U\boldsymbol{a}$). Eftersom $\sum_1^n \lambda_i a_i^2 \geq 0$ för *alla* val av $\boldsymbol{a}$ måste $\lambda_1, \lambda_2, \cdots, \lambda_n \geq 0$. Vi kan ju t ex ta $\boldsymbol{a} = \boldsymbol{e}_i = (0, \cdots, 0, 1, 0, \cdots, 0)^T$ dvs som den

i:te enhetsvektorn. Alltså har vi visat att C är positivt semi-definit. Alla positivt semi-definita symmetriska matriser duger som kovariansmatriser.

Konstruktion av A-matris

Vi vill nu hitta någon matris A sådan att $AA^T = C$ där C är en kovariansmatris och alltså positivt semi-definit. Detta kan uppfattas som att hitta ett slags kvadratrot till C.

Liksom för vanliga tal är denna kvadratrot ej entydigt bestämd. Ett (av många val) är följande.

Skriv $C = UDU^T$ där $D = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$ dvs diagonalisera C. Vi kan nu ta $A = U\text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \cdots, \sqrt{\lambda_n})U^T = UD^{1/2}U^T$. Notera att $\lambda_i \geq 0$ eftersom C är positivt semi-definit. Notera vidare att $D^{1/2}$ är symmetrisk. Vi får då

$$AA^T = UD^{1/2}U^T(UD^{1/2}U^T)^T = UD^{1/2}U^TUD^{1/2^T}U^T = UD^{1/2}D^{1/2}U^T = UDU^T = C$$

så detta val av A duger alltså.

Ett annat tänkbart val av A skulle vara med hjälp av Gram-Schmidts ortogonaliseringsförfarande i enlighet med vad som beskrivits i ett tidigare sammanhang. Vi får då A triangulär.

Ytterligare ett val skulle vara som (faktiskt den unika) symmetriska positivt semi-definita matris $C^{1/2}$ som uppfyller $C^{1/2}C^{1/2} = C$ - Matlab beräknar denna genom funktionen *sqrtm*.

Det finns alltså stor frihet att välja A-matris, men det centrala är att om vi tar $\boldsymbol{X} = \boldsymbol{m} + A\boldsymbol{Z}$ där $\boldsymbol{Z}$ består av oberoende $N(0,1)$-fördelade variabler, så blir fördelningen för $\boldsymbol{X}$ densamma oavsett val av A så länge som $AA^T = C$. $\boldsymbol{X}$ säges då vara $N_n(\boldsymbol{m}, C)$.

A behöver inte ens vara nxn, utan kan vara nxm med varierande m - vilket svarar mot att vi introducerar m st Z-variabler. Dock kan inte m vara hur litet som helst utan måste vara åtminstone r =rangen för C=dimensionen av det rum som spänns av kolonn-vektorerna i C. Är $r = n$ dvs att kolonnerna i C är linjärt oberoende så är C inverterbar. Då är faktiskt också A inverterbar om vi valt A som en nxn-matris. Detta beror på att rummet som spänns av kolonnerna i AA^T är samma som rummet som spänns av kolonnerna i A. Att inse det är lite mer avancerat men för fullständighetens skull ges här ett kort (och aningen abstrakt bevis).

Påstående: Rummen som spänns av kolonnerna i A respektive AA^T är identiska.

Kommentar: Rummet som spänns av ett antal vektorer är det linjära rum vi kan få av linjärkombinationer av dessa vektorer.

Bevis: Vi betecknar rummet som spänns av kolonnerna i en matris A med $M(A)$.

Låt A vara en nxm-matris. Vi nu tar en godtycklig n-dimensionell kolonnvektor $\boldsymbol{x}$ som är ortogonal mot alla de m (n-dimensionella) kolonnerna i A, dvs vi har $\boldsymbol{x}^TA = \boldsymbol{0}^T$. Då gäller att $\boldsymbol{x}^TAA^T = \boldsymbol{0}^T$, dvs $\boldsymbol{x}$ är ortogonal mot alla de n kolonnerna i AA^T. Detta betyder att $M(AA^T)$ ligger i $M(A)$.

Omvänt kan vi ta en godtycklig n-dimensionell kolonnvektor $\boldsymbol{x}$ som är ortogonal mot alla de n kolonnerna i $M(AA^T)$, dvs att $\boldsymbol{x}^TAA^T = \boldsymbol{0}^T$. Då gäller att $\boldsymbol{x}^TAA^T\boldsymbol{x} = 0$ dvs att $(A^T\boldsymbol{x})^T(A^T\boldsymbol{x}) = 0$ som ger $A^T\boldsymbol{x} = \boldsymbol{0}$ som innebär att $\boldsymbol{x}^TA = \boldsymbol{0}^T$. Men detta betyder att $\boldsymbol{x}$ är ortogonal mot alla kolonnerna i A. Detta betyder att $M(A)$ ligger i $M(AA^T)$.

Eftersom $M(A)$ innehåller $M(AA^T)$ och $M(AA^T)$ innehåller $M(A)$ så gäller $M(A) = M(AA^T)$, vsb.

En konsekvens av detta resultat är att om $AA^T = BB^T$ så gäller att $M(A) = M(AA^T)$ =(eftersom $AA^T = BB^T$) $= M(BB^T) = M(B)$. De underrum av R^n vi kan få genom att bilda $\boldsymbol{Y} = A\boldsymbol{Z}$

respektive $\boldsymbol{Y}' = B\boldsymbol{Z}'$ är alltså identiska, dvs massan ligger på samma underrum vare sig vi tittar på $\boldsymbol{Y}$ eller $\boldsymbol{Y}'$. Detta underrum har dimensionen r =rangen för A=rangen för $C = AA^T$.

Att välja $m > r$ =rangen för C innebär att man egentligen har med "onödiga" Z-variabler. Eftersom rangen för $C = AA^T$ är r så skulle vi kunna finna en ortogonal transformation U som gör att UAA^TU^T får följande utseende

$$\begin{pmatrix} D_r & 0 \\ 0 & 0 \end{pmatrix}$$

där $D_r = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_r)$ där alla dessa är > 0, dvs att D_r är inverterbar.