

Analiza în componente principale pentru compresia/restaurarea datelor

Conf.dr. Luminița STATE
Facultatea de Matematică, Universitatea București

Extragerea caracteristicilor lineare reprezinta unul dintre scopurile principale ale instruirii nesupervizate. În cadrul sistemelor biologice, extragerea caracteristicilor constituie prima etapa a mecanismului cognitiv în dobândirea capacitații suficiente procesării unor funcții cognitive superioare. În esență, extragerea caracteristicilor revine la un proces de eliminare a redundanței existente în date, caracteristicile rezultând ca funcții de elementele de structură ale spațiului formelor de intrare.

Cuvinte cheie: compresia datelor, restaurarea datelor, caracteristică lineară, KLT, TH, schemă de compresie/reconstrucție.

Procesul de extragere a caracteristicilor poate fi interpretat ca un proces de analiză a structurii spațiului de reprezentare în scopul detectării elementelor esențiale și este modelat din punct de vedere matematic printr-o transformare definită pe spațiul de reprezentare, $\varphi: R^n \rightarrow R^m$, cu $m \leq n$. Componentele $\varphi_1, \dots, \varphi_m$ sunt caracteristicile selectate, pentru $\forall x \in R^n, \varphi_i(x)$ fiind valoarea caracteristicii φ_i pentru forma reprezentată prin vectorul x , $1 \leq i \leq m$.

Dacă informația reprezentată de forma x este transmisă printr-un canal de comunicație prin intermediul vectorului m -dimensional $\varphi(x)$ atunci în cazul în care transformarea φ nu este inversabilă, restaurarea unei reprezentări n -dimensionale este în general afectată de erori. Cerința $m \leq n$ justifică termenii de reducere a dimensionalității respectiv compresie a datelor prin care este frecvent referit procesul de extragere a caracteristicilor.

Determinarea aplicației φ este realizată astfel încât pierderea de informație implicată de reducerea dimensiunii să fie minimizată, unde pierderea de informație este evaluată prin intermediul unei anumite funcții criteriu. Extragerea caracteristicilor lineare revi-

ne la inferența unei transformări lineare, deci, din punct de vedere matematic, la rezolvarea problemei variaționale

$$\inf_{A \in M_{n,m}(R)} E[f(X, \hat{X})]$$

unde $f(X, \hat{X})$ este valoarea funcției criteriu dacă forma inițială n -dimensională X a fost reprezentată prin intermediul vectorului m -dimensional $Y = \varphi(X) = A^T X$ al valorilor caracteristicilor selectate, pe baza căreia, în continuare, fiind efectuată reconstrucția n -dimensională \hat{X} .

Cu alte cuvinte, $f(X, \hat{X})$ exprimă o măsură a erorii corespunzătoare schemei de compresie/reconstrucție astfel rezultată $X \Rightarrow \varphi(X) \Rightarrow \hat{X}$.

Punctul clasic de vedere în abordarea problemei extragerii caracteristicilor lineare este unul dintre instrumentele importante ale analizei în componente principale. Analiza în componente principale este frecvent referită și sub numele de Transformare Karhunen-Loeve (KLT) respectiv Transformare Hotelling (TH). Extragerea caracteristicilor lineare pe baza analizei în componente principale revine la descompunerea după valorile singulare (SVD) ale matricei de covarianta corespunzătoare repartiției vectorilor formă și

determinarea reconstrucției \hat{X} ca proiecție a formei X pe spațiul linear generat de caracteristicile selectate conform valorilor singulare asociate din mulțimea vectorilor proprii.

Selectarea caracteristicilor lineare de varianță maximă

Fie $X = (X_1, \dots, X_n)^T$ vector aleator n -dimensional, presupunem $E(X) = 0$, $\text{cov}(X, X^T) = \Sigma$.

Definiție. Spunem că vectorul $\Psi_1 \in R^n$ este prima componentă principală în sensul varianței dacă $\|\Psi_1\| = 1$ și $\text{var}(\Psi_1^T X) = \sup_{\substack{\Phi \in R^n \\ \|\Phi\|=1}} \text{var}(\Phi^T X)$. Pentru k

număr natural $2 \leq k \leq n$, spunem că $\Psi_k \in R^n$ este cea de a k -a componentă principală, în sensul varianței, dacă avem relațiile: $\|\Psi_k\| = 1$ și $\text{var}(\Psi_k^T X) = \sup_{\substack{\Phi \in L^\perp(\Psi_1, \dots, \Psi_{k-1}) \\ \|\Phi\|=1}} \text{var}(\Phi^T X)$, unde $L^\perp(\Psi_1, \dots, \Psi_{k-1})$

este spațiul ortogonal spațiului linear generat de primele $k-1$ componente principale $\Psi_1, \dots, \Psi_{k-1}$.

Evident componentele principale în sensul varianței corespunzătoare unei repartiții sunt vectori ortonormali.

Definiție. Pentru m număr natural fixat, $1 \leq m \leq n$, caracteristicile optimale în sensul varianței pentru reprezentarea formei X sunt $\Psi_1, \Psi_2, \dots, \Psi_m$. Reprezentarea formei X în termenii setului de caracteristici Ψ_1, \dots, Ψ_m este vectorul aleator $Y = (Y_1, \dots, Y_m)^T$ unde $Y_k = \Psi_k^T X$, $1 \leq k \leq m$.

Dacă Ψ_1, \dots, Ψ_n este o bază ortonormală a spațiului R^n , notând $Y_i = \Psi_i^T X$, $1 \leq i \leq n$, rezultă $E(Y_i) = 0$, $\text{var}(Y_i) = E(Y_i^T X X^T \Psi_i) = \Psi_i^T \Sigma \Psi_i$, $1 \leq i \leq n$,

$$X = \sum_{i=1}^n Y_i \Psi_i. \text{ Notând } \Psi = [\Psi_1, \dots, \Psi_n]$$

rezultă $\Psi \Psi^T = \Psi^T \Psi = I_n$.

Deoarece $X = \Psi Y$ rezultă $Y = \Psi^T X$, deci $E(Y) = 0$ și $\text{cov}(Y, Y^T) = \Psi^T \Sigma \Psi$.

Observație. În particular dacă Ψ_1, \dots, Ψ_n sunt vectorii proprii ai matricei Σ atunci $\text{cov}(Y, Y^T) = \Psi^T \Sigma \Psi = \Lambda$ unde $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ este matricea diagonală a valorilor proprii asociate. Obținem astfel că dacă Ψ_1, \dots, Ψ_n sunt vectorii proprii ai matricei Σ transformarea $Y = \Psi^T X$ realizează decorrelarea variabilelor aleatoare X_1, \dots, X_n . În cazul particular al repartiției normale $X \sim N(0, \Sigma)$, rezultă că $Y_i = \Psi_i^T X$, $1 \leq i \leq n$ sunt variabile aleatoare independente, $Y_i \sim N(0, \lambda_i)$, $1 \leq i \leq n$.

Teorema 1. Fie $X = (X_1, \dots, X_n)^T$ vector aleator n -dimensional, presupunem $E(X) = 0$, $\text{cov}(X, X^T) = \Sigma$. Pentru orice k , $1 \leq k \leq n$, cea de a k -a componentă principală în sensul varianței este vectorul propriu Φ_k asociat valorii proprii λ_k a matricei Σ , unde $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$.

Demonstrație. Deoarece pentru orice $\Psi \in R^n$, $\text{var}(\Psi^T X) = \Psi^T \Sigma \Psi$, cum $\|\Psi\| = 1$, $\lambda_1 = \Phi_1^T \Sigma \Phi_1$ și

$$\lambda_1 = \sup_{\Psi \in R^n \setminus \{0\}} \frac{\Psi^T \Sigma \Psi}{\|\Psi\|^2} = \sup_{\substack{\Psi \in R^n \\ \|\Psi\|=1}} \Psi^T \Sigma \Psi$$

obținem că prima componentă principală în sensul varianței este Φ_1 . Presupunem că primele $k-1$ componente principale sunt $\Phi_1, \dots, \Phi_{k-1}$.

Deoarece $\Phi_k \in L^\perp(\Phi_1, \dots, \Phi_{k-1})$ și $\sup_{\Phi \in L^\perp(\Phi_1, \dots, \Phi_{k-1})} \text{var}(\Phi^T X) = \sup_{\Phi \in L^\perp(\Phi_1, \dots, \Phi_{k-1})} \Phi^T \Sigma \Phi$

$$\text{cum } \lambda_k = \Phi_k^T \Sigma \Phi_k = \sup_{\Phi \in L^\perp(\Phi_1, \dots, \Phi_{k-1})} \Phi^T \Sigma \Phi$$

rezultă Φ_k este cea de a k -a componentă principală în sensul varianței.

Selectarea caracteristicilor lineare optimale din punctul de vedere al erorii medii pătratice

Fie $X = (X_1, \dots, X_n)^T$ vector aleator n -dimensional $E(X) = 0$, $\text{cov}(X, X^T) = \Sigma$ și m număr natural fixat, $1 \leq m \leq n$. Presupunem că sunt selectate caracteristicile reprezentate de vectorii coloană ai matricei $W \in M_{n,m}(R)$ astfel încât matricea $W^T \Sigma W$ este nesingulară. Reprezentarea formei X în termenii caracteristicilor lineare selectate este $Y = W^T X$ deci $E(Y) = 0$, $\text{cov}(Y, Y^T) = W^T \Sigma W$.

Presupunem de asemenea că reconstrucția unei reprezentări n -dimensionale este realizată pe baza unei transformări lineare aplicate vectorului Y , $\hat{X} = UY$, $U \in M_{n,n}(R)$.

Teorema 2. Pentru $W \in M_{n,m}(R)$ fixat astfel încât matricea $W^T \Sigma W$ este nesingulară, reconstrucția optimală din punctul de vedere al erorii medii pătratice este realizată pentru $U = \Sigma W(W^T \Sigma W)^{-1}$.

Demonstrație Eroarea medie patratice implicată de schema de compresie/reconstrucție este: $X \xrightarrow{w^T} \varphi(X) = Y = W^T X \xrightarrow{U} \hat{X} = UY$ este $\varepsilon^2(\Sigma, W, U) = E(\|X - \hat{X}\|^2) = E(\|(I_n - UW^T)X\|^2) = E(X^T X) - 2E(X^T (UW^T)^T X) + E(X^T (UW^T)^T (UW^T)X)$ deci, $\varepsilon^2(\Sigma, W, U) = \text{tr}(XX^T) - 2E(X^T (UW^T)^T X) + E(X^T (UW^T)^T (UW^T)X)$.

Aplicând pentru $a \in R^n$, $B \in M_{n,n}(R)$ regulile de calcul, $\frac{\partial}{\partial B}(a^T B^T a) = aa^T$,

$$\frac{\partial}{\partial B}(a^T B^T Ba) = 2Baa^T, \quad \text{obținem} \\ \frac{\partial}{\partial U} \varepsilon^2(\Sigma, W, U) = -2E(XX^T W) +$$

$$+ 2E(UW^T XX^T W) = -2(\Sigma W - UW^T \Sigma W).$$

Rezultă că matricea U care asigură minimizarea erorii medii patratice este soluție a ecuației $0 = \frac{\partial}{\partial U} \varepsilon^2(\Sigma, W, U) = -2(\Sigma W - UW^T \Sigma W) \Rightarrow U = \Sigma W(W^T \Sigma W)^{-1}$. În consecință, reconstrucția lineară a formei X din reprezentarea $Y = W^T X$ este $\hat{X} = UY = \Sigma W(W^T \Sigma W)^{-1} W^T X$.

Observație. Pentru $W = [W_1, \dots, W_m]$ dat,

schema de compresie $X \xrightarrow{w^T} Y = W^T X$ conduce la reprezentarea m -dimensională $Y = (W_1^T X, \dots, W_m^T X)^T$ coloanele matricei W corespunzând caracteristicilor selectate. Deoarece $\cos(W_i, X) = \frac{W_i^T X}{\|W_i^T\| \|X\|}$, dacă $\|W_i\| = 1$, atunci valoarea fiecărei caracteristici W_i relativ la forma X este dată de proiecția vectorului X pe W_i , $1 \leq i \leq m$.

Observație. Fie $A = \Sigma W(W^T \Sigma W)^{-1} W^T$.

Deoarece $A^2 = A$ transformarea $\hat{X} = AX$ este un operator de proiecție deci reconstrucția optimală din punctul de vedere al erorii medii pătratice corespunde unei proiecții a spațiului R^n pe subspațiul $L = \{Ax | x \in R^n\}$. Evident $\text{rang}(A) = \text{rang}(W) = m$ deci $\dim(L) = m$.

Teorema 3. Aplicația lineară definită de matricea $A = \Sigma W(W^T \Sigma W)^{-1} W^T$ este o proiecție pe $L(W)$ spațiul linear generat de vectorii coloană ai matricei W .

Demonstrație. Se observă că spațiul linear generat de vectorii coloană ai matricei ΣW coincide cu $L(W)$.

Într-adevăr, fie $x \in L(W)$ deci $\exists \theta \in R^m$, $x = W\theta = \Sigma^{-1}(\Sigma W\theta)$. Deoarece $\Sigma W\theta$ aparține spațiului linear $L(\Sigma W)$ obținem $x \in L(\Sigma W)$ deci $L(W) \subset L(\Sigma W)$. Evident $\dim(L(\Sigma W)) \leq m$. Fie $\lambda \in R^m$ arbitrar. Deoarece Σ este inversabilă obținem $\Sigma W\lambda = 0 \Leftrightarrow W\lambda = 0 \Leftrightarrow \lambda = 0 \Rightarrow \dim(L(\Sigma W)) = m$. Din $L(W) \subset L(\Sigma W)$ și $\dim(L(\Sigma W)) = m = \dim(L(W))$ rezultă $L(W) = L(\Sigma W)$. Demonstrăm ca $L = L(\Sigma W)$. Fie $x \in L(\Sigma W)$ și $\theta \in R^m$ astfel încât $x = \Sigma W\theta$. Obținem $Ax = \Sigma W(W^T \Sigma W)^{-1} W^T x = \Sigma W(W^T \Sigma W)^{-1} W^T \Sigma W\theta = \Sigma W\theta = x$ deci $L(\Sigma W) \subset L$. Dacă $x \in L$ atunci $x = Ay = \Sigma W(W^T \Sigma W)^{-1} W^T y = \Sigma Wz$ pentru $y \in R^n$, $z = (W^T \Sigma W)^{-1} W^T y$ deci $x \in L(\Sigma W)$.

În concluzie $L = L(W)$, deci aplicația lineară definită de matricea $A = \Sigma W(W^T \Sigma W)^{-1} W^T$ este o proiecție a spațiului R^n pe spațiul linear generat de vectorii coloană ai matricei W .

Observație. Deoarece în general $A^T \neq A$, pentru operația de compresie lineară definită de matricea W , reconstrucția optimală din punctul de vedere al erorii medii pătratice corespunde unei proiecții neortogonale pe $L(W)$. În continuare vom determina matricea W astfel încât eroarea medie patratice $\varepsilon^2(\Sigma, W) = \inf_{U \in M_{n,m}(R)} \varepsilon^2(\Sigma, W, U) =$

$$= \varepsilon^2(\Sigma, W, \Sigma W(W^T \Sigma W)^{-1})$$

corespunzătoare schemei de compresie/reconstrucție lineară:

$$\begin{aligned} X &\xrightarrow{W^T} Y = W^T X \xrightarrow{U = \Sigma W(W^T \Sigma W)^{-1}} \\ &\Rightarrow \hat{X} = \Sigma W(W^T \Sigma W)^{-1} W^T X \end{aligned}$$

să fie minimizată.

Definiție. Eroarea medie patratice minimă pentru schema de compresie/recon-

strucție este $\varepsilon^2(\Sigma) = \inf_{W \in M_{n,m}} \varepsilon^2(\Sigma, W)$.

Schema de compresie/reconstrucție lineară $X \xrightarrow{W^T} Y = W^T X \xrightarrow{U} \hat{X} = UY$ este optimală din punctul de vedere al erorii medii pătratice, dacă $\varepsilon^2(\Sigma) = \varepsilon^2(\Sigma, W, U)$.

Lema 1. Fie $S \in M_{n,m}(R)$ astfel încât $\text{rang}(S) = m$ unde $1 \leq m < n$ și fie $D \in M_{n,n}(R)$ o matrice diagonală. Atunci, sunt îndeplinite urmatoarele egalități:

$$1. |S^T DS(S^T S)^{-1}| = |P^T DP|$$

$$2. \text{tr}\{S^T DS(S^T S)^{-1}\} = \text{tr}\{P^T DP\}$$

unde vectorii coloane ale matricei $P \in M_{n,m}(R)$ formează o baza orthonormală a spațiului linear generat de vectorii coloane ale matricei S .

Demonstrație. Fie $k = \begin{cases} n-1, m=n \\ m, m < n \end{cases}$ și $K^T = K_{k+1,k}^T \dots K_{2,1}^T$ unde matricele $K_{i,j}^T \in M_{n,n}(R)$, $1 \leq j < i \leq n$ sunt definite prin, $(K_{i,j})_{ii} = (K_{i,j})_{jj} = \frac{s_{ij}}{\sqrt{s_{jj}^2 + s_{ij}^2}}$,

$$(K_{i,j})_{ij} = -(K_{i,j})_{ji} = \frac{s_{ij}}{\sqrt{s_{jj}^2 + s_{ij}^2}}$$

și restul componentelor egale cu zero, $S = (s_{ij})$.

Evident $K_{i,j}^T$, $1 \leq j < i \leq n$ sunt matrice de rotație deci K este de asemenea o matrice de rotație. Notând $R = K^T S$ obținem $S = KR$.

Deoarece toate componentele ultimelor $(n-m)$ linii ale matricei R sunt egale cu zero, matricea S poate fi scrisă în continuare, $S = P\tilde{R}$ unde coloanele matricei $P \in M_{n,m}(R)$ sunt primele m

$$\text{coloane ale matricei } K \text{ și } R = \begin{bmatrix} \tilde{R} \\ 0 \end{bmatrix}.$$

Evident, \tilde{R} este o matrice superior triunghiulară și $\text{rang}(\tilde{R}) = m$ deci \tilde{R}

este inversabilă. Prima relație din enunț rezultă din seria de egalități,

$$\begin{aligned} S^T DS(S^T S)^{-1} &= \tilde{R}^T P^T D P \tilde{R} (\tilde{R}^T P^T P \tilde{R})^{-1} = \\ &= \tilde{R}^T P^T D P \tilde{R} (\tilde{R}^T)^{-1} (P^T P) (\tilde{R}^T P^T P \tilde{R})^{-1} (\tilde{R}^{-1})^T \\ &= \tilde{R}^T P^T D P (\tilde{R}^{-1})^T, \end{aligned}$$

$$\text{deci } |S^T DS(S^T S)^{-1}| = |P^T D P|.$$

În continuare, utilizând proprietatea $\text{tr}\{ABC\} = \text{tr}\{BCA\}$, obținem:

$$\begin{aligned} \text{tr}\{S^T DS(S^T S)^{-1}\} &= \text{tr}\{\tilde{R}^T P^T D P (\tilde{R}^{-1})^T\} = \\ &= \text{tr}\{P^T D P (\tilde{R}^{-1})^T \tilde{R}^T\} = \text{tr}\{P^T D P\} \end{aligned}$$

Teorema 4. (Schema de compresie/reconstrucție lineară optimală LMS)

Fie $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq \dots \geq \lambda_n$ valorile proprii ale matricei de covarianță Σ și $\Phi_1, \dots, \Phi_m, \dots, \Phi_n$ vectorii proprii asociati. Schema de compresie/reconstrucție lineară:

$$\begin{aligned} X &\xrightarrow{W^T} Y = W^T X \xrightarrow{U = \Sigma W (W^T \Sigma W)^{-1}} \\ &\Rightarrow \hat{X} = \Sigma W (W^T \Sigma W)^{-1} W^T X \end{aligned}$$

este optimală din punctul de vedere al erorii medii pătratice dacă $L(W) = L(\Phi_1, \dots, \Phi_m)$ unde $L(\Phi_1, \dots, \Phi_m)$ este spațiul linear generat de vectorii Φ_1, \dots, Φ_m .

Demonstrație. Conform rezultatelor precedente, pentru $W \in M_{n,m}(R)$, eroarea medie pătratică minimă a schemei de compresie/reconstrucție $\varepsilon^2(\Sigma, W) =$

$$= E(\|X - \hat{X}\|^2) = E(\|(I_n - A)X\|^2)$$

unde $A = \Sigma W (W^T \Sigma W)^{-1} W^T$. Obținem,

$$\varepsilon^2(\Sigma, W) = E\left(\left((I_n - A)X\right)^T \left((I_n - A)X\right)\right)$$

$$= \text{tr}\left\{E\left(\left((I_n - A)X\right)\left((I_n - A)X\right)^T\right)\right\} =$$

$$= \text{tr}\left\{(I_n - A)E(X X^T)(I_n - A)^T\right\} =$$

$$= \text{tr}\{\Sigma\} - 2\text{tr}\{\Sigma A^T\} + \text{tr}\{A \Sigma A^T\}.$$

Deoarece $\text{tr}\{A \Sigma A^T\} =$

$$\text{tr}\{\Sigma W (W^T \Sigma W)^{-1} W^T \Sigma W (W^T \Sigma W)^{-1} W^T \Sigma\}$$

$$= \text{tr}\{\Sigma W (W^T \Sigma W)^{-1} W^T \Sigma\} = \text{tr}\{A \Sigma\} =$$

$$= \text{tr}\{(A \Sigma)^T\} = \text{tr}\{\Sigma A^T\}$$

rezultă $\varepsilon^2(\Sigma, W) = \text{tr}\{\Sigma\} - \text{tr}\{\Sigma A^T\}$ deci eroarea medie pătratică este minimizată dacă matricea W este astfel încât:

$$\text{tr}\{\Sigma W (W^T \Sigma W)^{-1} W^T \Sigma\} =$$

$$= \sup_{B \in M_{n,m}(R)} \text{tr}\{\Sigma B (B^T \Sigma B)^{-1} B^T \Sigma\}.$$

Notăm cu T mulțimea punctelor critice ale funcției $\text{tr}\{\Sigma W (W^T \Sigma W)^{-1} W^T \Sigma\}$.

Fie $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_n\}$ matricea diagonală a valorilor proprii corespunzătoare matricei de covarianță Σ și $\Phi = [\Phi_1, \dots, \Phi_n]$ matricea având coloanele vectorii proprii ortonormali asociați, $\Sigma = \Phi \Lambda \Phi^T$. Notând $S = \Lambda^{1/2} \Phi W$, obținem $S^T S = W^T \Phi \Lambda \Phi^T W$.

Rezultă în continuare,

$$\text{tr}\{\Sigma W (W^T \Sigma W)^{-1} W^T \Sigma\} =$$

$$= \text{tr}\{\Phi \Lambda \Phi^T W (W^T \Phi \Lambda \Phi^T W)^{-1} W^T \Phi \Lambda \Phi^T\} =$$

$$= \text{tr}\{S^T \Lambda S (S^T S)^{-1}\} \text{ deci } \frac{\partial}{\partial W} \text{tr}\{S^T \Lambda S (S^T S)^{-1}\} =$$

$$= \frac{\partial}{\partial S} \text{tr}\{S^T \Lambda S (S^T S)^{-1}\} \frac{\partial S}{\partial W} =$$

$$= [-2S(S^T \Lambda S) + 2\Lambda S] \frac{\partial S}{\partial W} \text{ unde } \frac{\partial S}{\partial W}$$

este matrice inversabilă.

Obținem astfel $W \in T \Leftrightarrow$

$$\frac{\partial}{\partial W} \text{tr}\{S^T \Lambda S (S^T S)^{-1}\} = 0 \Leftrightarrow$$

$$-2S(S^T \Lambda S) + 2\Lambda S = 0 \Leftrightarrow$$

$$(SS^T)\Lambda S = \Lambda S. \text{ Evident } (SS^T)\Lambda S = \Lambda S$$

$$\Leftrightarrow (S^T \Lambda S)S^T = S^T \Lambda \Leftrightarrow \text{coloanele matricei } S^T \text{ sunt vectori proprii ai matricei } S^T \Lambda S \text{ asociați la } m \text{ dintre valorile proprii ale matricei } \Sigma. \text{ Fie } W \in T;$$

dacă $\lambda_1, \dots, \lambda_m$ sunt valorile proprii ale matricei Σ asociate punctu-

lui, atunci valoarea funcției criteriu este

$$\text{tr}\left\{\Sigma W(W^T \Sigma W)^{-1} W^T \Sigma\right\} = \text{tr}\{S^T \Lambda S\} = \sum_{j=1}^m \lambda_j.$$

Din relația $(SS^T)\Lambda S = \Lambda S$ obținem de asemenea $(S^T S)(S^T \Lambda S) = S^T \Lambda S$ deci cum matricea $S^T S$ este inversabilă rezultă $(S^T S)^{-1}(S^T \Lambda S) = S^T \Lambda S$ adică

$$\begin{aligned} \text{tr}\{S^T \Lambda S\} &= \text{tr}\{(S^T S)^{-1} S^T \Lambda S\} = \\ &= \text{tr}\{S^T \Lambda S(S^T S)^{-1}\}. \end{aligned}$$

În concluzie, valoarea funcției obiectiv $\text{tr}\{S^T \Lambda S(S^T S)^{-1}\}$ într-un punct critic

este egală cu $\text{tr}\{S^T \Lambda S\} = \sum_{j=1}^m \lambda_j$ pentru anume m valori proprii distincte ale matricei Σ .

Demonstrăm că orice punct $W = [\Phi_{i_1}, \dots, \Phi_{i_m}]$ aparține mulțimii T . Pentru simplificarea notației presupunem $W = [\Phi_1, \dots, \Phi_m]$. Obținem,

$$\begin{aligned} SS^T \Lambda S &= \Lambda^2 \Phi W (\Phi W)^T \Lambda^2 \Phi W = \\ &= \Lambda^2 \begin{bmatrix} I_m \\ 0 \end{bmatrix} \begin{bmatrix} I_m & 0 \end{bmatrix} \Lambda^2 \begin{bmatrix} I_m \\ 0 \end{bmatrix} = \\ &= \begin{bmatrix} \lambda_1^2 \sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & \lambda_2^2 \sqrt{\lambda_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_m^2 \sqrt{\lambda_m} \\ 0 & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

Evident, $\Lambda S = \Lambda^2 \Phi W = \Lambda^2 \begin{bmatrix} I_m \\ 0 \end{bmatrix}$ deci

$$SS^T \Lambda S = \Lambda S \text{ adică } W \in T.$$

În concluzie,

$$\sup_{B \in M_{n,m}(R)} \text{tr}\left\{\Sigma B(B^T \Sigma B)^{-1} B^T \Sigma\right\} = \sum_{j=1}^m \lambda_j$$

unde $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq \dots \geq \lambda_n$ sunt valorile proprii ale matricei de covarianță Σ , supremumul funcționalei

$$\text{tr}\left\{\Sigma B(B^T \Sigma B)^{-1} B^T \Sigma\right\} \text{ fiind atins}$$

într-un singur punct $B = [\Phi_1, \dots, \Phi_m]$.

Rezultă că schema de compresie/reconstrucție lineare de eroare medie pătratică minimă corespunde matricelor $W = [\Phi_1, \dots, \Phi_m]$, $U = \Sigma W(W^T \Sigma W)^{-1} = W$ adică este:

$$X \xrightarrow{W^T} Y = W^T X \xrightarrow{W} \hat{X} = WW^T X \text{ eroarea fiind egală cu } \varepsilon^2(\Sigma) = \inf_{W \in M_{n,m}} \varepsilon^2(\Sigma, W) =$$

$$\text{tr}\{\Sigma\} - \sum_{j=1}^m \lambda_j = \sum_{j=m+1}^n \lambda_j.$$

Referințe bibliografice

- [1] Bannour S., Azimi-Sadjani R. - *Principal Component Extraction Using Recursive Least Squares Learning*, IEEE Transactions on Neural Networks, 6, no.2, 1995;
- [2] Deco G., Obradovic D. - *An Information-Theoretic Approach to Neural Computing*, Springer-Verlag, 1996;
- [3] Diamantaras K.I. - *Principal Component Neural networks: Theory and Applications*, John Wiley & Sons, Inc. 1996;