Determinarea caracteristicilor lineare optimale din punct de vedere informaţional în compresia şi decompresia datelor

Conf. dr. Luminţa STATE
Catedra de Informatică, Universitatea Bucureşti
Prep. Cătălina COCIANU
Catedra de Informatică Economică, A.S.E., Bucureşti

Extragerea caracteristicilor liniare revine, în esenţa, la un proces de eliminare a redundanţei existente în date, caracteristicile rezultând ca funcţii de elementele de structură ale spaţiului formelor de intrare. Modelele informaţionale dezvoltate furnizează algoritmi foarte eficienţi de compresie/decompresie, reconstrucţie şi recunoaştere de forme. Este exemplificat algoritmul RLS, cu aplicaţii în compresia/decompresia imaginilor (binare şi cu un număr de 256 nivele de gri de la alb la negru), calităţile sale fiind relevate şi comparativ cu un model standard de extragere de caracteristici, şi anume modelul Karhunen-Loeve.
Cuvinte cheie: *extractor linear de caracteristici, entropie, layer, vector sinaptic, eroare Hamming, funcţie criteriu.*

Fie X vector aleator n-dimensional $E(X) = \mu$, $cov(X,X^T) = \Sigma$ matrice simetrică şi pozitiv definită. Dacă $T\in M_{m\times n}(R)$ este un extractor de caracteristici lineare unde $m\le n$ şi $Y=T^TX$ este vectorul aleator reprezentând valorile caracteristicilor selectate corespunzătoare formei X, atunci $E(Y)=T^T\mu$ şi $cov(Y,Y^T) = T^T\Sigma\, T$. Informaţia reciprocă $I(X;T^TX)$ reprezintă o masură a cantităţii de informaţie relativ la X reţinută prin utilizarea filtrului linear T:

$$I(X;T^TX) = H(T^TX) - H\left(T^TX\middle|X\right) = H(X) - H\left(X\middle|T^TX\right)\ge 0,$$ unde H este entropia Shannon.

Deoarece $H\left(T^TX\middle|X\right)=0$, rezultă că $H(T^TX)\le H(X)$, $H\left(X\middle|T^TX\right)=H(X) - H(T^TX)$, deci $H\left(X\middle|T^TX\right)$ reprezintă o măsură a pierderii de informaţie implicate de utilizarea filtrului linear T. Extragerea de caracteristici lineare optimale din punct de vedere informaţional revine astfel la determinarea unui filtru T care să minimizeze entropia condiţionată $H\left(X\middle|T^TX\right)$, ceea ce este echivalent cu maximizarea entropiei $H(T^TX)$.
În continuare vom considera numai extractori de caracteristici lineare care conduc la repartiţii nedegenerate, matricea $T^T\Sigma\, T$ fiind pozitiv definită dacă şi numai dacă rang(T)=m. Evident pentru orice $T\in M_{m\times n}(R)$ astfel încât $T^TT=I_m$ rezultă rang(T)=m. Fie, prin notaţie, mulţimea $M(m,n)=\left\{T\middle|T\in M_{n\times m}(R), \left|T^T\Sigma T\right|\neq 0, T^TT=I_m\right\}$.
Fie $\Phi=(\Phi_1, ..., \Phi_n)$ matricea vectorilor proprii unitari şi $\Lambda=diag(\lambda_1, ..., \lambda_n)$ matricea diagonală a valorilor proprii asociate corespunzători matricei Σ; $\Phi^T\Sigma\Phi=\Lambda$, Φ matrice ortogonală. Determinarea caracteristicilor lineare optimale din punct de vedere informaţional pentru o repartiţie dată, corespunde problemei variaţionale:

$$(1)\quad \sup_{T\in M(m,n)} H\left(T^TX\right)$$

Lema 1 Fie $A\in M_{m\times n}(R)$, $A=\|a_{ij}\|$, $S\in M_m(R)$, $S=\|\sigma_{ij}\|$, matrice simetrică astfel încât $\left|A^TSA\right|\neq 0$, matrice simetrică. Atunci,

$$(2)\ \frac{\partial}{\partial A}\ln\left|A^TSA\right| = 2SA\left(A^TSA\right)^{-1},\text{ unde } \frac{\partial}{\partial A}\ln\left|A^TSA\right| = \left\|\frac{\partial}{\partial a_{ij}}\ln\left|A^TSA\right|\right\|.$$

Demonstrație Notând cu $R(A^TSA)_{kl}$ cofactorul corespunzator componentei a_{kl} din matricea A^TSA, $1\le k, l\le m$ rezultă,

$$\frac{\partial}{\partial A}\ln\left|A^TSA\right| = \frac{1}{\left|A^TSA\right|}\left\|\frac{\partial}{\partial a_{ij}}\left|A^TSA\right|\right\| = \frac{1}{\left|A^TSA\right|}\left\|\frac{\partial}{\partial a_{ij}}\sum_{k=1}^{m}\sum_{l=1}^{m}\left(A^TSA\right)_{kl}\left|R\left(A^TSA\right)_{kl}\right|\right\|$$

Componentele matricei $\frac{\partial}{\partial A}\ln\left|A^TSA\right|$ sunt,

$$\frac{\partial}{\partial a_{ij}}\left|A^TSA\right| = \sum_{k=1}^{m}\sum_{l=1}^{m}\frac{\partial}{\partial\left(A^TSA\right)_{kl}}\left|A^TSA\right|\frac{\partial}{\partial a_{ij}}\left(A^TSA\right)_{kl} = \sum_{k=1}^{m}\sum_{l=1}^{m}\left|R\left(A^TSA\right)_{kl}\right|\frac{\partial}{\partial a_{ij}}\left(A^TSA\right)_{kl}$$

Deoarece $\frac{\partial}{\partial a_{ij}}\left(A^TSA\right)_{kl} = \frac{\partial}{\partial a_{ij}}\sum_{p=1}^{n}\sum_{q=1}^{n}a_{pk}\sigma_{pq}a_{ql} = \sum_{p=1}^{n}\sum_{q=1}^{n}\frac{\partial a_{pk}}{\partial a_{ij}}\sigma_{pq}a_{ql} + \sum_{p=1}^{n}\sum_{q=1}^{n}\frac{\partial a_{pk}}{\partial a_{ij}}a_{pk}\sigma_{pq}$

rezultă $\frac{\partial}{\partial a_{ij}}\left(A^TSA\right)_{kl} = \begin{cases} 0, k\ne j, l\ne j \\ (SA)_{il}, k=j, l\ne j \\ (SA)_{ik}, k\ne j, l=j \\ 2(SA)_{ij}, k=l=j\end{cases}$

Obținem astfel,

$$\frac{\partial}{\partial a_{ij}}\left|A^TSA\right| = \sum_{\substack{l=1\\ l\ne j}}^{m}\left|R\left(A^TSA\right)_{jl}\right|(SA)_{il} + \sum_{\substack{k=1\\ l\ne j}}^{m}\left|R\left(A^TSA\right)_{kj}\right|(SA)_{ik} + 2(SA)_{ij}\left|R\left(A^TSA\right)_{jj}\right| =$$

$$2\sum_{k=1}^{m}\left|R\left(A^TSA\right)_{kj}\right|(SA)_{ik} = 2\left((SA)\left(A^TSA\right)^*\right)_{ij}$$

În concluzie, $\left\|\frac{\partial}{\partial a_{ij}}\left|A^TSA\right|\right\| = 2\left|A^TSA\right|SA\left(A^TSA\right)^{-1}$, deci $\frac{\partial}{\partial A}\ln\left|A^TSA\right| = 2SA(A^TSA)^{-1}$

Lema 2 Fie $A\in M_{m\times n}(R)$, $A=\left\|a_{ij}\right\|$, $S\in M_m(R)$, $S=\left\|\sigma_{ij}\right\|$, matrice simetrică. Atunci,

$$(3)\ \frac{\partial}{\partial A}\mathrm{tr}\left\{ASA^T\right\} = 2AS,\text{ unde } \frac{\partial}{\partial A}\mathrm{tr}\left\{ASA^T\right\} = \left\|\frac{\partial}{\partial a_{ij}}\mathrm{tr}\left\{ASA^T\right\}\right\|.$$

Demonstrație Deoarece $\mathrm{tr}\{ASA^T\}=\mathrm{tr}\{SA^TA\}$, obținem

$$\frac{\partial}{\partial a_{ij}}\mathrm{tr}\left\{ASA^T\right\} = \frac{\partial}{\partial a_{ij}}\sum_{p=1}^{m}\sum_{q=1}^{m}\sum_{r=1}^{n}\sigma_{pq}a_{rq}a_{rp} = \sum_{p=1}^{m}\sum_{q=1}^{m}\sum_{r=1}^{n}\sigma_{pq}\frac{\partial a_{rq}}{\partial a_{ij}}a_{rp} + \sum_{p=1}^{m}\sum_{q=1}^{m}\sum_{r=1}^{n}\sigma_{pq}a_{rq}\frac{\partial a_{rp}}{\partial a_{ij}} =$$

$2\sum_{p=1}^{m}a_{ip}\sigma_{pj} = 2(AS)_{ij}$ deci $\frac{\partial}{\partial A}\mathrm{tr}\left\{ASA^T\right\} = 2AS.$

Teorema 1 Dacă $X\sim N(\mu,\Sigma)$ atunci $\sup_{T\in M(m,n)} H\left(T^TX\right) = \frac{1}{2}\sum_{i=1}^{m}\ln\lambda_i + \frac{m}{2}\ln 2\pi e$, unde $\lambda_1\ge\lambda_2\ge\ldots\ge\lambda_m\ge\ldots\lambda_n$ sunt valorile proprii ale matricei de covarianță Σ. Mulțimea soluțiilor problemei variaționale (1) este $\left\{(\Phi_1,\ldots,\Phi_m)\Psi \mid \Psi\in M_m(R), \Psi^T\Psi = I_m\right\}$.

Demonstrație Deoarece $X \sim N(\mu,\Sigma)$, pentru orice $T\in M(m,n)$, rezultă că $T^TX \sim N(T^T\mu, T^T\Sigma T)$, deci $H(T^TX) = \frac{m}{2}\ln 2\pi e + \frac{1}{2}\ln\left|T^T\Sigma T\right|$. Notăm cu $S(\varpi)$spațiul punctelor critice ale funcției obiectiv.

$J_m(T) = \ln\left|T^T\Sigma T\right| - tr\left\{\varpi\left(T^TT - I_m\right)\right\}$ unde ϖ este o matrice simetrică,

$$S(\varpi)=\left\{T \middle| T\in M_{m\times n}(R), \left|T^T\Sigma T\right|\neq 0, \frac{\partial J_m(T)}{\partial T}=0\right\}$$

Deoarece $\frac{\partial \ln\left|T^T\Sigma T\right|}{\partial T} = 2\Sigma T(T^T\Sigma T)^{-1}$, $\frac{\partial tr\left(T\varpi T^T\right)}{\partial T} = 2T\varpi$ rezultă că $T\in S(\varpi)$ dacă și numai dacă $\left|T^T\Sigma T\right|\neq 0$ și $2\Sigma T(T^T\Sigma T)^{-1} = 2T\varpi$ adică $\Sigma T = T\varpi T^T\Sigma T$.

Obținem în continuare, $T^T\Sigma T = T^TT\varpi\ TT\Sigma T$. ceea ce evident implică $I_m = T^TT\varpi$. În particular, pentru $\varpi = I_m$ obținem $S(I_m)\subseteq M(m,n)$.

Deoarece $S(I_m)$ conține punctele de extrem ale funcției obiectiv și pentru orice $T\in S(I_m)$, $\Sigma T = TT^T\Sigma T$, $J_m(T) = \ln\left|T^T\Sigma T\right|$, rezultă în particular că $S(I_m)$ conține toate soluțiile problemei variaționale (1).

Fie Ψ_m matricea având coloanele vectorii proprii unitari ai matricei $T^T\Sigma T$ și Λ_m matricea diagonală a valorilor proprii asociate; $T^T\Sigma T = \Psi_m\Lambda_m\Psi_m^T$.

Pentru orice $T\in S(I_m)$ rezultă $\Sigma T = T\Psi_m\Lambda_m\Psi_m^T$ adică $\Sigma T\ \Psi_m = T\Lambda_m\Psi_m^T$, deci componentele matricei Λ_m sunt valori proprii ale matricei Σ iar coloanele matricei $T\Psi_m$ sunt vectorii proprii unitari asociați și $\ln\left|T^T\Sigma T\right| = \sum_{j=1}^{m}\lambda_{i_j}$ unde $\Lambda_m = diag\left(\lambda_{i_1},\ldots,\lambda_{i_m}\right)$.

Evident, pentru orice m valori proprii $\lambda_{i_1},\ldots,\lambda_{i_m}$ ale matricei Σ, există $T\in S(I_m)$ astfel încât $\ln\left|T^T\Sigma T\right| = \sum_{j=1}^{m}\ln\lambda_{i_j}$. Într-adevăr, dacă $T = \left(\Phi_{i_1},\ldots,\Phi_{i_m}\right)$ atunci $\left|T^T\Sigma T\right| = \prod_{j=1}^{m}\lambda_{i_j}\neq 0$ și

$\Sigma T = T\,diag\left(\lambda_{i_1},\ldots,\lambda_{i_m}\right) = T\ T^T\Sigma T$ deci $T\in S(I_m)$ și $\ln\left|T^T\Sigma T\right| = \sum_{j=1}^{m}\ln\lambda_{i_j}$.

Rezultă $\sup\limits_{T\in M(m,n)}\ln\left|T^T\Sigma T\right| = \sum_{i=1}^{m}\ln\lambda_i$ deci $\sup\limits_{T\in M(m,n)} H\left(T^TX\right) = \frac{1}{2}\sum_{i=1}^{m}\ln\lambda_i + \frac{m}{2}\ln 2\pi e$ unde $\lambda_1,\ldots,\lambda_m$ sunt cele mai mari m valori proprii ale matricei de covarianță Σ.

Fie $T\in S=\left\{(\Phi_1,\ldots,\Phi_m)\Psi \middle| \Psi\in M_m(R), \Psi^T\Psi = I_m\right\}$arbitrar.

Evident $\left|(\Phi_1,\ldots,\Phi_m)^T\Sigma(\Phi_1,\ldots,\Phi_m)\right| = \prod_{i=1}^{m}\lambda_i \neq 0$,

$T\ T^T\Sigma T = (\Phi_1, \ldots, \Phi_m)\ \Psi\Psi^T(\Phi_1, \ldots, \Phi_m)^T\Sigma T = \Sigma T$ deci $T\in S(I_m)$.

Deoarece entropia este invariantă față de transformările ortogonale obținem:

$$H(T^TX) = H(\Psi^T(\Phi_1,\ldots,\Phi_m)X) = H((\Phi_1,\ldots,\Phi_m)X) = \frac{1}{2}\ln\left|(\Phi_1,\ldots,\Phi_m)^T\Sigma(\Phi_1,\ldots,\Phi_m)\right| + \frac{m}{2}\ln 2\pi e$$

$$= \frac{1}{2}\sum_{i=1}^{m}\ln\lambda_i + \frac{m}{2}\ln 2\pi e.$$

Reciproc, fie $T\in S(I_m)$ astfel încât $\ln\left|T^T\Sigma T\right| = \sum_{i=1}^{m}\ln\lambda_i$; conform celor demonstrate anterior,

$T\Psi_m = (\Phi_1,...,\Phi_m)$, unde Ψ_m este matricea având coloanele vectorii proprii unitari ai matricei $T^T\Sigma T$, deci $T=(\Phi_1,...,\Phi_m)\,\Psi_m^{\ T}$ adică $T\in S$.

În concluzie, pentru repartiția normală, direcțiile principale din punct de vedere informațional sunt vectorii proprii ai matricei de covarianță, reducerea dimensionalității revenind la selectarea caracteristicilor reprezentate de vectorii proprii asociați celor mai mari valori proprii. Determinarea caracteristicilor optimale din punct de vedere informațional pentru repartiții care nu mai sunt de tip normal comportă un studiu elaborat, frecvent problema devenind de nerezolvat.

În cele ce urmează vor fi prezentate argumente de ordin informațional pentru justificarea simplificării problemei prin acceptarea unui anumit tip de suboptimalitate, caracteristicile astfel determinate fiind referite prin termenul de caracteristici sigure.

Fie $T\in M_{n\times m}(R)$, $U\in M_{n\times(n-m)}(R)$ astfel încât $T^TT=I_m$, $U^TU=I_{n-m}$, $|T^T\Sigma T|\neq 0$ $|U^T\Sigma U|\neq 0$, $T^T\Sigma U=0$. Evident, există cel puțin o pereche de matrice care să verifice aceste cerințe și anume $T=(\Phi_1,...,\Phi_m)$, $U=(\Phi_{m+1},...,\Phi_n)$, unde $\Phi_1,...,\Phi_n$ sunt vectorii proprii ai matricei de covarianță Σ.

Fie $A=(T,U)\in M_n(R)$; rezultă că $|A^T\Sigma A|=|T^T\Sigma T||U^T\Sigma U|\neq 0$. Dacă $Y=A^TX=$ $=\begin{pmatrix}Y^{(1)}\\ Y^{(2)}\end{pmatrix}$, unde $Y^{(1)}=T^TX$, $Y^{(2)}=U^TX$; atunci $cov(Y^{(1)},T^{(2)T})=0$, deci $Y^{(1)}$, $Y^{(2)}$ sunt necorelați. În particular, dacă X este repartizat normal, atunci și Y este repartizat normal, deci $Y^{(1)}$, $Y^{(2)}$ sunt independenți ceea ce implică $H(X)=H(Y)-\ln|A|=$ $=H(T^TX)+H(U^TX)-\ln|A|$. Obținem deci,

$$H(X|T^TX)=H(X)-H(T^TX)=H(U^TX)-\ln|A|,$$

ceea ce înseamnă că tentativa de identificare a unui extractor de caracteristici $T\in M_{n\times m}(R)$ suboptimal din punct de vedere informațional poate fi reformulată prin problema variațională:

(2) $\inf\limits_{U\in M^*(m,n)} H(U^TX)$, unde

$$M^*(m,n)=\left\{U\,|\,U\in M_{n\times(n-m)}(R), |U^T\Sigma U|\neq 0, U^TU=I_{n-m}\right\}$$

Fie X vector aleator de medie μ și matrice de covarianță Σ. Dacă U este o soluție a problemei (2), atunci matricea $T\in M_{n\times m}(R)$ având coloanele vectori unitari și ortogonali pe vectorii coloană ai matricei U definește un extractor de caracteristici lineare "aproape" optimal din punct de vedere informațional.

Conform teoremei Gibbs, pentru orice $U\in M^*(m,n)$, are loc inegalitatea: $H(U^TX)\leq\frac{n-m}{2}\ln 2\pi e+\frac{1}{2}\ln|U^T\Sigma U|$, deci minimizarea entropiei $H(U^TX)$ pe spațiul $M^*(m,n)$ poate fi realizată selectând $U\in M^*(m,n)$ care minimizează termenul $\ln|U^T\Sigma U|$, deci soluție a problemei variaționale min-max:

(3) $\inf\limits_{U\in M^*(m,n)}\ \sup\limits_{f\in G(U^T\mu,U^T\Sigma U)} H(f)$,

unde $G(U^T\mu,U^T\Sigma U)$ este clasa funcțiilor de densitate (n-m)-dimensionale, de medie $U^T\mu$ și matrice de covarianță $U^T\Sigma U$.

Caracteristicile lineare definite de matricea $T\in M_{n\times m}(R)$, având coloanele vectori unitari și ortogonali pe vectorii coloană ai matricei U, soluție a problemei (3), sunt evident suboptimale din punct de vedere informațional și, rezultând pe baza unui criteriu min-max, au semnificația de caracteristici sigure.

Utilizând argumentele prezentate în demonstrația teoremei 1 rezultă că mulțimea soluțiilor problemei variaționale (3) este $\left\{(\Phi_{m+1},...,\Phi_n)\Psi\,|\,\Psi\in M_{n-m}(R), \Psi^T\Psi=I_{n-m}\right\}$, unde $\Phi_{m+1},...,\Phi_n$ sunt vectorii proprii unitari asociați celor mai mici (n-m) valori proprii ale matricei Σ. În consecință, caracteristicile sigure corespunzatoare unei repartiții de matrice de covarianță Σ sunt caracteristicile lineare optimale din punct de vedere informațional pentru repartiția normală de matrice de covarianță Σ, adică

$$\left\{(\Phi_1,\dots,\Phi_m)\Psi \middle| \Psi \in M_m(R),\ \Psi^T\Psi = I_m\right\}$$

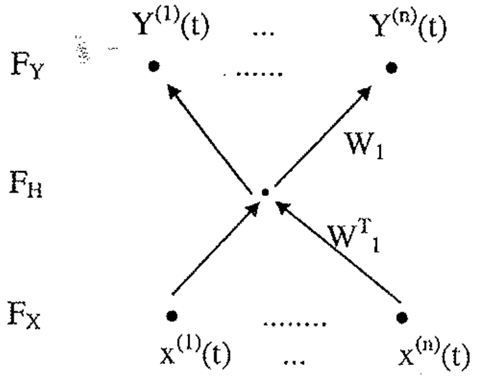
unde $\Phi_1,\dots,\Phi_m$ sunt vectorii proprii unitari asociaţi celor mai mari m valori proprii ale matricei Σ.

Algoritmul RLS

Algoritmul RLS furnizează o metodă eficientă de compresie/decompresie prin extragerea optimală a primelor *m* caracteristici principale ale unei intrări corespunzând unui proces staţionar *n*-dimensional (*n>m*) (X(t)).

Extragerea primei caracteristici principale

Arhitectura neurală utilizată pentru extragerea primei caracteristici principale conţine 3 layere F_X, F_H, F_Y, unde F_X este layer-ul de intrare, F_Y este layer-ul de ieşire, $|F_X| = |F_Y| = n$ şi F_H este layer-ul ascuns, $|F_H| = 1$. Reprezentarea arhitecturii:



$$F_X \xrightarrow[W_1^T]{} F_H \xrightarrow[W_1]{} F_Y$$

în care:

- $X(t) = [\,X^{(1)}(t),\ \dots,\ X^{(n)}(t)]$ reprezintă intrarea (proces staţionar);
- activarea neuronului ascuns se face după regula: $h_1(t) = W^T{}_1(t-1)\,X(t)$;
- $Y(t) = [\,Y^{(1)}(t),\ \dots,\ Y^{(n)}(t)]$ reprezintă ieşirea, detrminată conform regulii $Y(t) = W_1(t-1)h_1(t) = W_1(t-1)W^T{}_1(t-1)\,X(t)$.

Funcţia criteriu pentru determinarea estimaţiei filtrului (ponderii) W_1, optimală din punct de vedere al erorii medii pătratice, este: $J_1(t) = \sum_{k=1}^{t}\varepsilon^2(k)$, unde $\varepsilon^2(k) = \|X(k) - Y(k)\|^2$.

Presupunând că la toate momentele de timp k, $1 \le k \le t$, a fost utilizat filtrul $W_1(t)$, atunci:

$$J_1(t) = \frac{1}{t}\sum_{k=1}^{t}\left(X(k) - W_1(t)h_1(k)\right)^T\left(X(k) - W_1(t)h_1(k)\right),$$

$$\hat{W}_1(t) = \frac{\overline{X}^T(t)h^T{}_1(t)}{\left\|\overline{h}_1^T(t)\overline{h}_1(t)\right\|^2}$$

Algoritmul RSL pentru determinarea estimaţiei $\hat{W}_1$ (componentei principale) este:

$$h_1(t) = W^T{}_1(t-1)\,X(t)$$

$$k_1(t) = \frac{P_1(t-1)h_1(t)}{1 + h_1^2(t)P_1(t-1)}$$

$$P_1(t) = [\,1 - k_1(t)h_1(t)\,]\,P_1(t-1)$$

$$\hat{W}_1(t) = \hat{W}_1(t-1) + k_1(t)\left[X(t) - h_1(t)\hat{W}_1(t-1)\right]$$

Teoremă Algoritmul RSL permite învăţarea primei componente principale ϕ_1 la nivelul vectorului sinaptic corespunzător neuronului ascuns: $\lim_{t\to\infty}\hat{W}_1(t) = \phi_1$, unde ϕ_1 reprezintă vectorul propriu corespunzător valorii proprii λ_1 a matricei de covarianţă Σ ($\lambda_1 \ge \lambda_2 \ge \dots \ge \lambda_n$).

Demonstrație

$$J_1(t)=\frac{1}{t}\sum_{k=1}^{t}\left[X(k)-W_1(t)W_1^T(t)X(k)\right]^T\left[X(k)-W_1(t)W_1^T(t)X(k)\right]=$$

$$=\frac{1}{t}\sum_{k=1}^{t}X^T(k)X(k)-\frac{2}{t}\sum_{k=1}^{t}X^T(k)W_1(t)W_1^T(t)X(k)+\frac{1}{t}\sum_{k=1}^{t}X^T(k)W_1(t)W_1^T(t)W_1(t)W_1^T(t)X(k)$$

Cum $\lim_{t\to\infty}\frac{1}{t}\sum_{k=1}^{t}X(k)X^T(k)=\Sigma$, rezultăcă $\frac{1}{t}\sum_{k=1}^{t}X^T(k)X(k)\cong \mathrm{tr}\,\Sigma$

$$\frac{1}{t}\sum_{k=1}^{t}X^T(k)W_1(t)W_1^T(t)X(k)=W_1^T(t)\left(\frac{1}{t}\sum_{k=1}^{t}X^T(k)X(k)\right)W_1(t)\cong W_1^T(t)\,\Sigma\,W_1(t)$$

$$\frac{1}{t}\sum_{k=1}^{t}X^T(k)W_1(t)W_1^T(t)W_1(t)W_1^T(t)X(k)\cong\left\|W_1(t)\right\|^2W_1^T(t)\,\Sigma\,W_1(t)$$

Deci $J_1(t)\cong \mathrm{tr}\,\Sigma-2W_1^T(t)\,\Sigma\,W_1(t)+\left\|W_1(t)\right\|^2W_1^T(t)\,\Sigma\,W_1(t)$

$\mathrm{grad}\,J_1(t)=-4\,\Sigma\,W_1(t)+2W_1(t)W_1^T(t)\,\Sigma\,W_1(t)+2\left\|W_1(t)\right\|^2\Sigma\,W_1(t)$

Cum $\left\|W_1(t)\right\|^2\Sigma\,W_1(t)=W_1(t)W_1^T(t)\,\Sigma\,W_1(t)$, rezultăcă:

$\mathrm{grad}\,J_1(t)=-4\,\Sigma\,W_1(t)+4W_1(t)W_1^T(t)\,\Sigma\,W_1(t)$

Fie $S=\left\{\hat{W}\in R^n/\mathrm{grad}\,J_1(t)_{W_1=\hat{W}}=0\right\}$

Observații:

1. Toți vectorii proprii ai matricii Σ aparțin spațiului S.
Într-adevăr, dacă $\hat{W}_1=\phi_i$, atunci $\mathrm{grad}\,J_1(t)_{\hat{W}=\phi_i}=-4\,\Sigma\,\phi_i+4\phi_i\phi_i^T\,\Sigma\,\phi_i=-4\lambda_i\phi_i+4\lambda_i\phi_i=0$

2. Orice vector nenul din S este unitar.
Într-adevăr, dacă $\hat{W}\in S$, atunci:

$\Sigma\hat{W}=\hat{W}\hat{W}^T\Sigma\hat{W}$, $\hat{W}^T\Sigma\hat{W}=\hat{W}^T\hat{W}\,\hat{W}^T\Sigma\hat{W}$. Deci $\hat{W}^T\Sigma\hat{W}\left(1-\left\|\hat{W}\right\|^2\right)=0$. Cum Σ este pozitiv definită si $\hat{W}$ este nenul, rezultă $\hat{W}^T\Sigma\hat{W}\neq 0$, deci $\left\|\hat{W}\right\|$.

3. Punctele de minim ale functiei J_1 sunt exact vectorii proprii corespunzători matricii Σ.
Într-adevăr, fie functiile criteriu $\bar{J}_i$ ($i=\overline{1,n}$) definite pe spatiul S astfel:

$\bar{J}_i(t)=\mathrm{tr}\,\Sigma-2\hat{W}^T\,\Sigma\,\hat{W}+\left\|\hat{W}\right\|^2\hat{W}^T\,\Sigma\,\hat{W}+\lambda_i\left(\hat{W}^T\hat{W}-1\right),\quad i=\overline{1,n}$

Cum $\left\|\hat{W}\right\|=1$ pe spatiul S, rezultă că $\bar{J}_i(t)=\mathrm{tr}\,\Sigma-\hat{W}^T\,\Sigma\,\hat{W}+\lambda_i\left(\hat{W}^T\hat{W}-1\right),\quad i=\overline{1,n}$

Functiile J_1 si $\bar{J}_i$ au acelesi puncte critice si aceleasi puncte de minim în raport cu spatiul S.

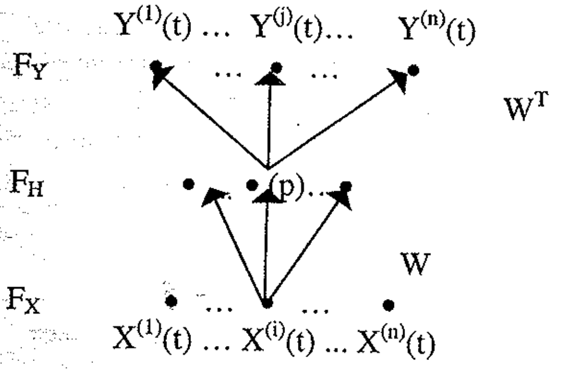
$\mathrm{grad}\,\bar{J}_i(t)=-\Sigma\,\hat{W}+\lambda_i\hat{W}$. Pe S $\mathrm{grad}\,\bar{J}_i(t)=0\Rightarrow\Sigma\,\hat{W}=\lambda_i\hat{W}$, deci $\hat{W}$ reprezintă vectorul propriu corespunzător lui λ_i. Deci punctele de minim ale functiei J_1 sunt exact vectorii proprii corespunzători matricii Σ.

$J_1(t)_{\hat{W}=\phi_i}=\mathrm{tr}\,\Sigma-\phi_i^T\,\Sigma\,\phi_i=\sum_{j=1}^{n}\lambda_j-\lambda_i=\sum_{\substack{j=1\\ j\neq i}}^{n}\lambda_j$, deci minimul absolut pentru J_1 corespunzător lui $\hat{W}$ apartinând lui S este dat de $\hat{W}=\phi_1$.

Algoritmul RLS garantează minimizarea lui J_1 ; suprafata de eroare are unul sau mai multe puncte de minim local, dar valoarea criteriului J_1 este aceeasi; punctul de minim absolut pentru J_1 este ϕ_1, deci, asimtotic, vectorul optimal este ϕ_1 : $\lim_{t\to\infty}\hat{W}_1(t)=\phi_1$.

Extragerea primelor m componente principale

Arhitectura neurală utilizată în scopul extragerii primelor *m* componente principale este, cu aceleaşi notaţii, următoarea:



$$F_X \xrightarrow[W^T]{} F_H \xrightarrow[W]{} F_Y$$

Reprezentarea internă calculată de arhitectura neurală e dată de vectorii sinaptici între F_X şi F_H. Antrenarea arhitecturii se realizează secvenţial astfel încât procesul de antrenare pentru neuronul *p* începe când s-a decis terminarea procesului de antrenare pentru neuronul *p-1*.

$W_p = [\, W_{1p}, \ldots, W_{np}]^T$

$h_p(t) = W^T(t-1)\, X(t)$

$Y(t) = W(t-1)h_p(t) = W(t-1)\; W^T(t-1)\; X(t)$

$p \in 1..m$

$D(t) = X(t)$ - răspunsul dorit

Răspunsul dorit la momentul de timp *t* în procesul de calcul al neuronului *p* de pe layer-ul F_H este dat prin relaţia:

$$d_p(t) = X(t) - \sum_{i=1}^{p-1} \hat{W}_i(t) h_i(t)$$

Funcţia criteriu pentru neuronul *p* la momentul *t* este dată de relaţia:

$$J_p(t) = \frac{1}{t}\sum_{k=1}^{t}\left(d_p(k) - W_p(k)h_p(k)\right)^T\left(d_p(k) - W_p(k)h_p(k)\right)$$

$$h_p(t) = \frac{1}{t}\sum_{k=1}^{t}\left[X(k) - \sum_{i=1}^{p-1}\hat{h}_i(t)\hat{W}_i(t) - W_p(k)h_p(k)\right]^T\left[X(k) - \sum_{i=1}^{p-1}\hat{h}_i(t)\hat{W}_i(t) - W_p(k)h_p(k)\right]$$

Presupunând că la toate momentele de timp $k \in 1..t\text{-}1$ ar fi utilizat filtrul $W_p(t)$ în locul filtrelor $W_p(1), \ldots, W_{p-1}(t)$, atunci $\operatorname{grad} J_p(t) = -\sum_{k=1}^{t} h_p(k)\left(d_p(k) - W_p(t)h_p(k)\right)$, iar minimul lui J_p se atinge pentru $\sum_{k=1}^{t} h_p(k) d_p(k) = W_p(t)\sum_{k=1}^{t}\left(h_p(k)\right)^2$. Fie $\overline{h}_p(t) = [\, h_p(1), \ldots, h_p(t)]^T$, $\overline{D}_p(t) = [d_p(1), \ldots, d_p(t)]^T$; obţinem: $\overline{D}_p(t)\,\overline{h}_p(t) = W_p(t)\left\|\overline{h}_p(t)\right\|^2$, deci $W_p(t) = \frac{1}{\left\|\overline{h}_p(t)\right\|^2}\overline{D}_p^T(t)\overline{h}_p(t)$.

În consecinţă, algoritmul RLS extins pentru neuronul *p* este următorul:

intrare: $X(t)$ proces staţionar de medie 0 şi matrice de covarianţă Σ

$h_p(t) = W^T(t-1)\, X(t)$

$$k_p(t) = \frac{P_p(t-1)h_p(t)}{1 + h_p^2(t)P_p(t-1)}$$

$P_p(t) = [\, 1 - k_p(t)h_p(t)\,]\, P_p(t-1)$

$$\hat{W}_p(t) = \hat{W}_p(t-1) + k_p(t)\left[X(t) - h_p(t)\hat{W}_p(t-1)\right]$$

Teoremă. Algoritmul RLS descris pentru antrenarea neuronului *p* determină asimtotic ca vector sinaptic vectorul propriu ϕ_p: $\lim_{t\to\infty}\hat{W}_p(t) = \phi_p$

Demonstraţia este similară celei din teorema anterioară (se poate găsi în [1])

Experimente şi concluzii

Algoritmul RLS a fost implementat şi testat de autori pentru două clase mari de probleme, şi anume reducerea dimensionalităţii unui semnal binar, exemplificat în cazul de faţă prin caractere, respectiv a unui semnal oarecare, exemplificat prin imagini cu 256 nivele de gri de la alb la negru.
Testele efectuate au demonstrat, pe de o parte, un comportament stabil la perturbaţii, iar pe de altă parte, variaţii foarte reduse de informaţie la o reducere a dimensionalităţii la o zecime şi, în cazul imaginilor, chiar la o reducere mai drastică.
Implementarea acestui algoritm în cazul caracterelor a fost realizată considerând ca semnal de intrare un proces staţionar de dimensiune 100 (rezultat în urma liniarizării matricilor caracter 10×10), cu câte 11 instanţe pentru fiecare literă, dintre care o instanţă reprezentând prototipul literei respective.
Curba de eroare Hamming corespunzătoare refacerii literelor în urma efectuării reducerii numărului de componente principale (deci a dimensionalităţii) de la 100 (numărul total de componente) la 5, cu pasul 13, demonstrează că perturbaţii puternice survin în cazul semnalului redus după aproximativ 7 paşi, adică atunci când din cele 100 componente sunt selectate mai puţin de 15 componente principale.
Rezultatele obţinute prin aplicarea algoritmului RLS în cazul semnalelor binare devin şi mai interesante în momentul comparării lor cu rezultate similare ale aplicării unei metode standard de extragere de caracteristici, şi anume metoda Karhunen-Loeve; se remarcă astfel că erorile apărute la refacere prin compactarea cu RLS rămân acceptabile după o reducere foarte drastică, în timp ce, în cazul compactării cu Karhunen-Loeve, deşi la o reducere mică erorile sunt nesemnificative, în cazul unei selecţii foarte mici de caracteristici, erorile sunt mult mai mari decât cele date de RSL.
Pentru exemplificare este prezentată în figurile 1 şi 2 refacerea unui set de caractere (selectie asupra literei *a*) în urma reducerii dimensionalităţii prin metodele RLS şi Karhunen Loeve, în paralel cu prezentarea curbei de eroare Hamming pentru prototipuri pentru fiecare caz în parte.
Testele efectuate în cazul unor imagini cu 256 nivele de gri de la alb la negru au demonstrat de asemenea posibilităţile mari de reducere pe care le oferă acest algoritm şi în special stabilitatea la erorile survenite în determinarea efectivă a componentelor principale în cazul unor intrări foarte mari. Datorită dimensiunilor foarte mari ale unei imagini liniarizate, modul de implementare diferă de cel al extragerii caracteristicilor pentru caractere; astfel, imaginea nu este tratată în întregime la o implementare, ci linie cu linie, acest lucru neafectând, după cum se va vedea, rezultatul refacerii.
În continuare (figurile 3 şi 4) sunt prezentate două exemple, pentru fiecare menţionându-se doar prototipul intrării şi rezultatul refacerii în urma extragerii primelor 3 caracteristici principale(din totalul de 320) pentru 8 instanţe în cazul primului exemplu, respectiv primelor 25 caracteristici principale (din acelaşi total) pentru 35 de instanţe în cazul celui de-al doilea exemplu.

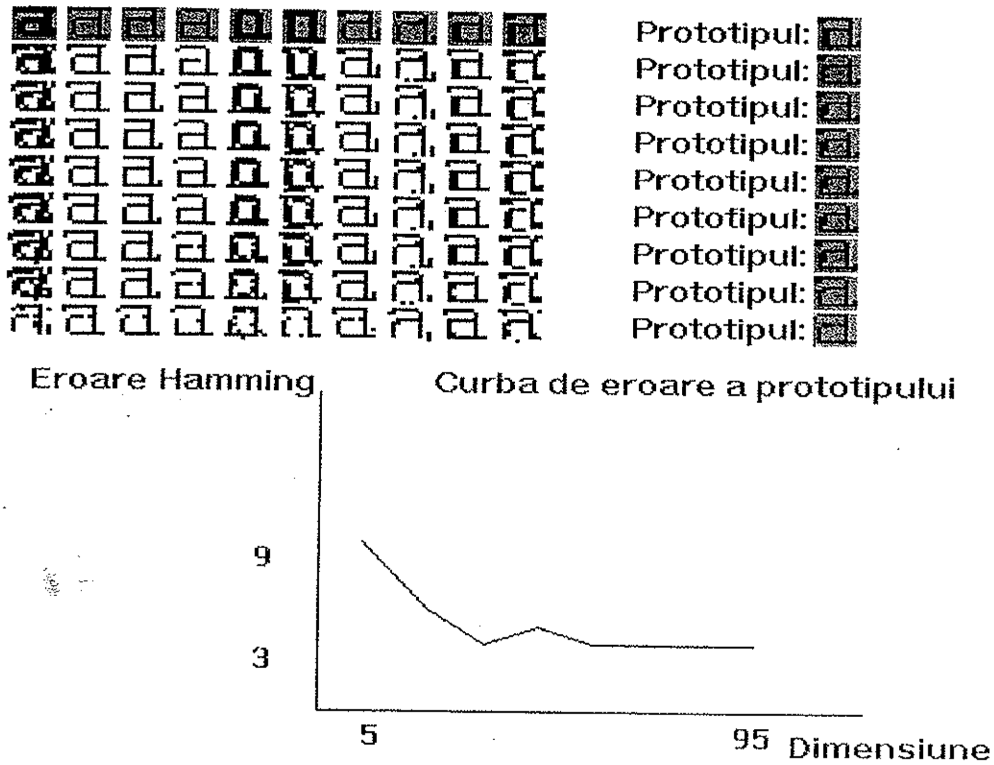


Fig. 1. Reducere Karhunen-Loeve

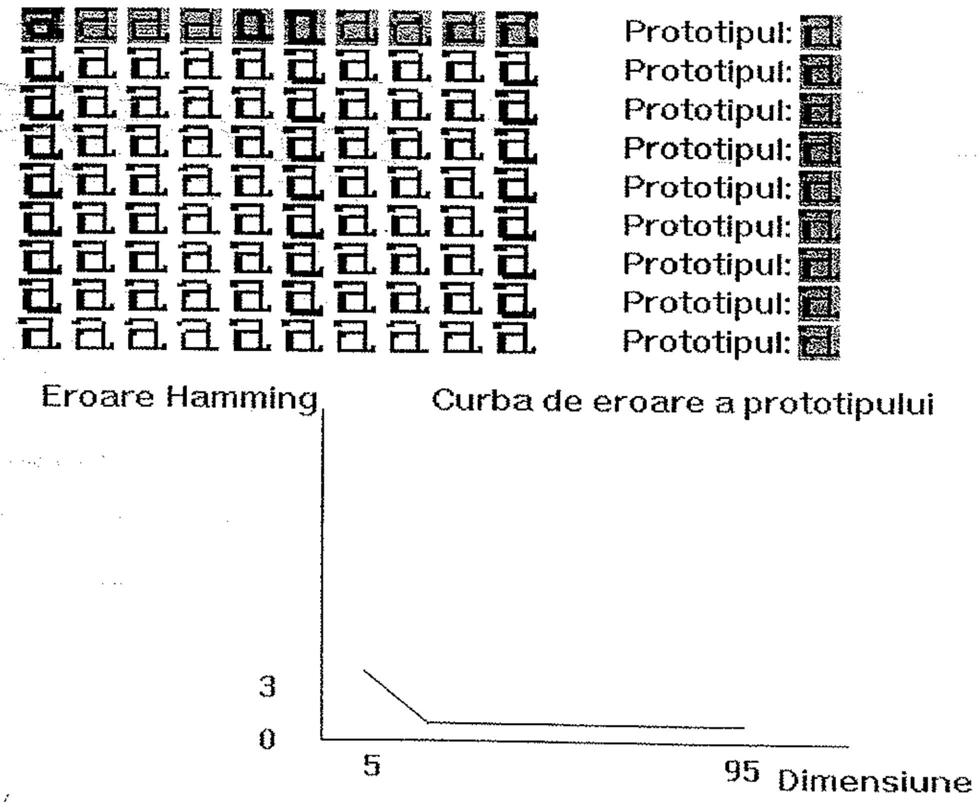


Fig. 2. Reducere RLS

Prototipul intrării Ieşirea

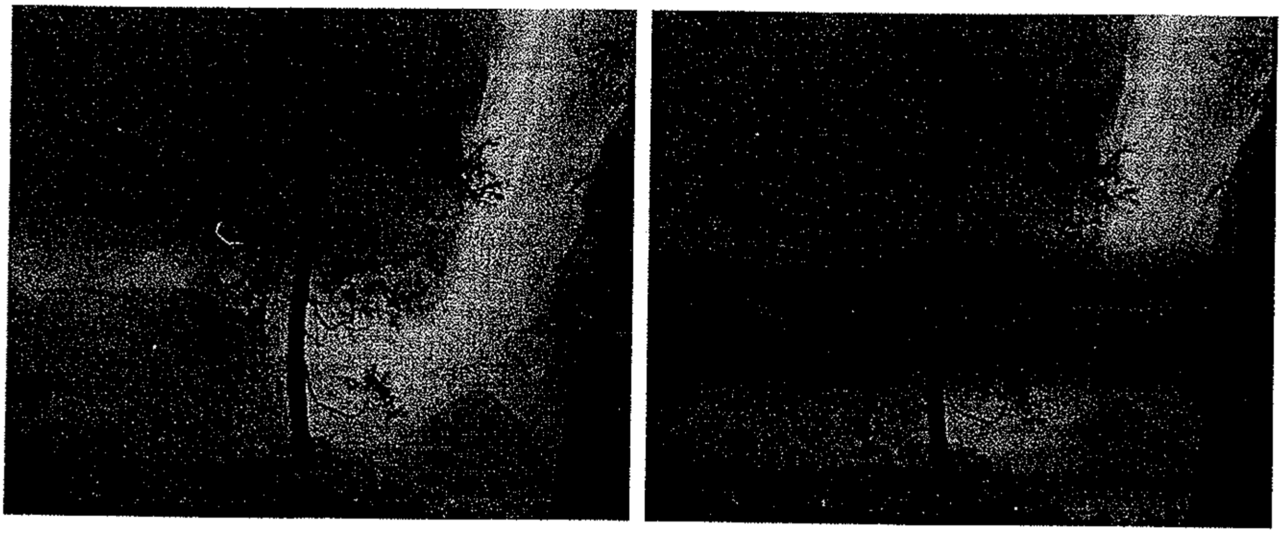


Fig. 3. Reducere de la 320 la 3 caracteristici principale (8 instanţe ale intrării)

Prototipul intrării Ieşirea

Fig. 4. Reducere de la 320 la 35 de caracteristici pǎrincipale (35 instanţe ale intrării)

Bibliografie

[1] Diamantaras K.I. *"Principal Component Neural networks: Theory and Applications"*, John Wiley&Sons, Inc. 1996

[2] Deco G., Obradovic D. *"An Information-Theoretic Approach to neural Computing"*, Springer Verlag, 1996

[3] Cover T.M., Thomas J.A. *"Elements of Information Theory"*, John Wiley&Sons, Inc, 1991

[4] State L. *"Analiza în componente principale pentru compresia/restaurarea datelor"*, Informatica Economică, nr.2/1997

[5] Gonzales Rafael, Woods Richard *"Digital Image Processing"*, Addison-Wesley, 1993

[6] Pitas Ioannis *"Digital Image Processing"*, Prentice Hall, 1993