

PROCESAREA SEMNALELOR - CURS 11

SERII DE TIMP - MODELE NEPARAMETRICE

Cristian Rusu

CUPRINS

- criterii pentru alegerea modelelor AR
- staționaritatea seriilor de timp

MODELUL AR

- folosim modelul AR și faptul că ne dorim $y[i] \approx \hat{y}[i]$ pentru fiecare i

$$\begin{bmatrix} y[i] \\ y[i-1] \\ y[i-2] \\ \dots \end{bmatrix} = \begin{bmatrix} y[i-1] & y[i-2] \\ y[i-2] & y[i-3] \\ y[i-3] & y[i-4] \\ \dots & \dots \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

- scris echivalent $\mathbf{y} = \mathbf{Y}\mathbf{x}$
- ce dimensiune are fiecare variabilă?
 - \mathbf{y} este $m \times 1$
 - \mathbf{Y} este $m \times p$
 - \mathbf{x} este $p \times 1$
- m se numește orizontul de timp
- p este dimensiunea modelului AR
- acesta este un model parametric

REVIEW METODA PCA

- la tablă
- rezumat:
 - se dă un set de date, o matrice de dimensiune $d \times N$, d reprezintă numărul de caracteristici (features) iar N numărul de elemente din setul de date
 - se aplică algoritmul Singular Value Decomposition (SVD) și se obține
$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$
 - (există o legătură între această factorizare și factorizarea cu valori/vectori proprii a matricei de covarianță $\mathbf{X}\mathbf{X}^T = \mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^T$)
 - în această descompunere, matricele \mathbf{U} și \mathbf{V} sunt ortogonale (de dimensiune $d \times d$ și a doua $N \times N$) iar matricea $\mathbf{\Sigma}$ este diagonală, iar pe diagonala sa se află valorile singulare ale matricei \mathbf{X} care sunt non-negative și sunt aranjate în ordine descrescătoare
 - primele p coloane din matricea \mathbf{U} (le notăm cu \mathbf{U}_p) reprezintă proiecția pe sub-spațiul în care este concentrată cea mai mare cantitate de energie
 - proiecția pe acest sub-spațiu se calculează cu $\tilde{\mathbf{X}} = \mathbf{U}_p^T \mathbf{X}$
 - $\tilde{\mathbf{X}}$ are dimensiunea $p \times N$
 - reconstrucția este $\mathbf{X}_{\text{rec}} = \mathbf{U}_p \tilde{\mathbf{X}}$ și are dimensiunea $d \times N$

SINGULAR SPECTRUM ANALYSIS

- primim o serie de timp de dimensiune N

$$\{x_1, x_2, x_3, \dots, x_N\}$$

- fixăm dimensiunea unei ferestre L
- construim matricea de întârziere (este o matrice de tip Hankel)

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 & x_3 & \dots & x_K \\ x_2 & x_3 & x_4 & \dots & x_{K+1} \\ x_3 & x_4 & x_5 & \dots & x_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & x_{L+2} & \dots & x_N \end{bmatrix}$$

- calculăm PCA pentru $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_{i=1}^L \sigma_i \mathbf{u}_i \mathbf{v}_i^T$

SINGULAR SPECTRUM ANALYSIS

- calculăm PCA pentru $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_{i=1}^L \sigma_i \mathbf{u}_i \mathbf{v}_i^T$
- calculăm matricele $\mathbf{X}_i = \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ (fiecare e de dimensiune $d \times N$)
- luăm fiecare matrice \mathbf{X}_i și o Hankel-izăm
- din matricea \mathbf{X}_i extragem astfel seria de tip de lungime L numită $\hat{\mathbf{x}}_i$
- în felul acesta am realizat descompunerea seriei de tip și avem

$$\mathbf{x} = \sum_{k=1}^L \hat{\mathbf{x}}_k$$

DATA VIITOARE

- procese Gaussiene

