Předmět B2M31DSP/Př. 10

PS

Přednáška 10: Rozklad na hlavní složky - PCA

Obsah

Rozklad na hlavní komponenty – PCA

Používané názvy:

- PCA – Principal Component Analysis nebo EVD – Eigenvalue decomposition - rozklad kovarianční matice na vlastní vektory a vlastní čísla
- KLT – Karhunenova-Loevevova transformace používána pro ztrátovou kompresi

PCA je lineární metoda analýzy dat (signálů)

- nepředpokládá nic o rozdělení pravděpodobnosti[1] dat/signálů $\rightarrow$ jednoduchá neparametrická metoda (poskytuje relevantní informace z často matoucího souboru dat)
- hledá směry největšího rozptylu dat/signálů - jsou určeny vlastními vektory
- lze ji použít pro redukci dimenze příznakového prostoru a tedy pro ztrátovou kompresi dat
- pozor na nelineární případy: PCA neposkytuje správný výsledek

[1] I když původní práce vycházely z normálního rozdělení dat.

Smysl PCA:

nalézt vhodnou bazi[2] a v ní znovu vyjádřit data $\rightarrow$ dekoreluje původní data (diagonalizuje kovarianční matici dat $\mathbf{C}_x$) a poskytuje vlastní čísla a vlastní vektory

Kovarianční matici dat

$$\mathbf{C}_x = \mathbf{X}\mathbf{X}^\mathbf{T},$$

kde $\mathbf{X}$ je matice dat (sloupce tvoří jedno měření, řádky obsahují měřené hodnoty v čase)

lze rozložit:

$$\mathbf{C}_x = \mathbf{V}\mathbf{D}\mathbf{V}^\mathbf{T}$$

matice $\mathbf{D}$ je diagonální a obsahuje vlastní čísla = rozptyly dat ve směrech určených vlastními vektory = sloupce matice $\mathbf{V}$

Úpravou předchozí rovnice lze získat

$$\mathbf{D} = \mathbf{V}^\mathbf{T}\mathbf{C}_x\mathbf{V}$$

Pozn.: bázové vektory jsou u PCA totožné s vlastními vektory

[2] Před provedením PCA je třeba data centrovat a normovat

Smysl PCA:
Dekorelaci dat provádí transformace (KLT)

$$\mathbf{Y} = \mathbf{V}^T\mathbf{X},$$

kde $\mathbf{Y}$ je matice nekorelovaných dat

Ověření:
Platí

$$\mathbf{C}_y = \mathbf{Y}\mathbf{Y}^{\mathbf{T}}$$

po dosazení za $\mathbf{Y}$ a úpravě získáme

$$\mathbf{C}_y = \mathbf{V}^{\mathbf{T}}\mathbf{C}_x\mathbf{V} = \mathbf{D}$$

Matice $\mathbf{C}_y$ je diagonální a tedy $\mathbf{Y}$ obsahuje dekorelovaná data

Pozn.:
Matice $\mathbf{V}$ je ortogonální, platí tedy[3] $\mathbf{V}^{\mathbf{T}}\mathbf{V} = \mathbf{E}$, z čehož plyne[4] $\mathbf{V}^{\mathbf{T}} = \mathbf{V}^{-1}$

[3] $\mathbf{E}$ je jednotková matice

[4] Po násobení obou stran rovnice rovnice inverzní maticí zprava

Ztrátová komprese 1-D signálů

Princip ztrátové komprese pro 1-D data

- Provedeme transformaci 1-D signálu $\mathbf{x}$ pomocí ortogonální transformace dané maticí[5] $\mathbf{T}$, čímž získáme vektor nekorelovaných vzorků $\mathbf{y}$

$$\mathbf{y} = \mathbf{T}\mathbf{x}$$

- Ponecháme nejsilnější složky[6] $\mathbf{y} \rightarrow \tilde{\mathbf{y}}$
- Provedeme zpětnou transformaci

$$\tilde{\mathbf{x}} = \mathbf{T}^{\mathbf{T}}\tilde{\mathbf{y}}$$

Používanými ortogonálními transformacemi jsou

- DFT, DCT, DWT - signálově nezávislé
- PCA - signálově závislá -na rozdíl od DCT je baze PCA závislá na datech, a proto má PCA vyšší výpočetní nároky než DCT; za to nabízí větší kompresní poměr

---

[5] Matice $\mathbf{T}$ je v případě DFT rovna matici $\mathbf{W}$ a v případě PCA je to transponovaná matice vlastních vektorů $\mathbf{V}^T$

[6] Odpovídající největším vlastním číslům

Redukce dimenzionality pomocí PCA

Princip redukce dimenzionality vícedimenzionálních signálů pomocí PCA

- Sestavíme matici dat $\mathbf{X}$ rozměru [M, N], jejíž sloupce jsou jednotlivá měření[7]; počet sloupců N určuje počet těchto měření
- Určíme čtvercovou korelační matici [M, M]

$$\mathbf{C_x} = \mathbf{XX^T}$$

- Určíme vlastní vektory $\mathbf{v_i}$ a vlastní čísla $\lambda_i,\ i = 1, 2, ..., M$ matice $\mathbf{C_x}$; víme, že platí[8]

$$\mathbf{C}_x = \mathbf{VDV^T}$$

[7] viz příklad pro M=6 souřadnic získaných 3 kamerami snímajících přímočarý pohyb

[8] Nebo pro jednotlivé vlastní vektory a čísla $\mathbf{C_x v_i} = \lambda_i \mathbf{v_i}$

Redukce dimenzionality pomocí PCA

- Počet významých vlastních čísel určuje dimenzionalitu dat

Pokud je třeba provést rekonstrukci dat, postupujeme podobně jako u ztrátové komprese 1-D dat

- Provedeme transformaci

$$\mathbf{Y} = \mathbf{V}^T\mathbf{X},$$

  protože platí $\mathbf{C_x} = \mathbf{VDV^T}$, pak $\mathbf{C_y} = \mathbf{D}$ je diagonální matice vlastních čísel $\rightarrow$ matice $\mathbf{Y}$ tedy obsahuje nekorelovaná data

- Provedeme redukci

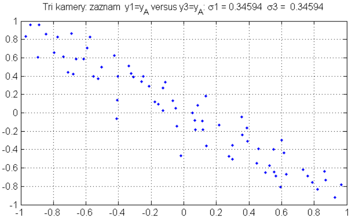
$$\mathbf{Y} \rightarrow \tilde{\mathbf{Y}}$$

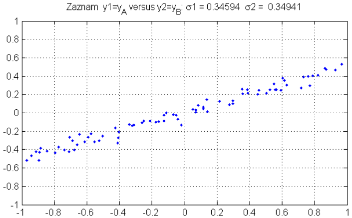
- Rekonstrukci dat s nižší dimenzinalitou získáme zpětnou transformací, při které používáme pouze složky, které odpovídají nejvýznamnějším vlastním číslům

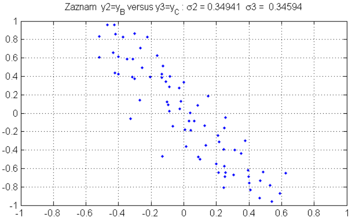
$$\tilde{\mathbf{X}} = \mathbf{V}\tilde{\mathbf{Y}}$$

Příklad PCA pro redukci dimenzionality

Snímání 1-D scény 3 kamerami → 6-D úloha?







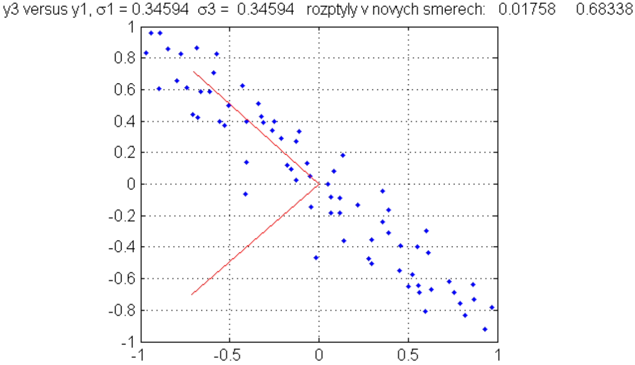
Příklad PCA pro redukci dimenzionality

Pro jednu kameru získáme dvě nenulová různá vlastní čísla - jedno číslo je dominantní: $\lambda_1 = 0.6834$, $\lambda_2 = 0.0176$, a tedy úloha pro 1 kameru není 2-D ale pouze 1-D $\rightarrow$ redukce dimenzionality je možná

Podobně pro všechny kamery získáme 6 různých vlastních čísel
$\lambda_1 = 1.79\, \lambda_2 = 0.026$, $\lambda_3 = 0.002$, $\lambda_4 = 0.0015$, $\lambda_5 = 0.001$, $\lambda_6 = 0.0008$
pouze jedno je dominantní, a tedy úloha je 1-D a nikoliv 6-D

Příklad PCA pro redukci dimenzionality

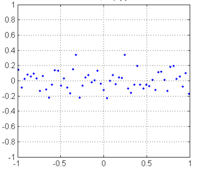
Rozptyly dat v původních a transformovaných=vlastních směrech (červené úsečky) pro 1 kameru

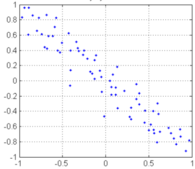


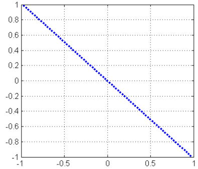
Příklad PCA pro redukci dimenzionality

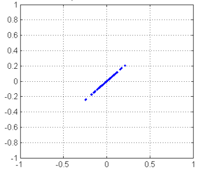
A. data a rozptyly v transformovaných a v původních směrech pro 1 kameru - první dva obrázky;

B. rekonstrukce po redukci dimenzionality ze 2-D na 1-D = rekonstruované průměty dat do vlastních směrů[9] jsou různě dlouhé :-) - 3. a 4. obrázek





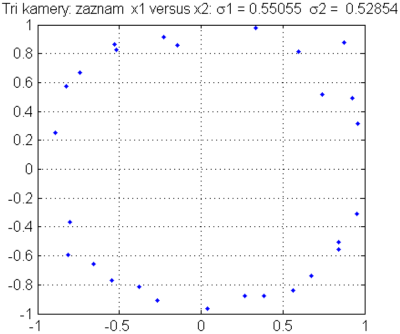




[9] Hlavní komponenty (Principal Components) jsou definovány jako průměty dat do podprostoru, i když některé prameny ztotožňují hlavní komponenty s vlastními vektory.

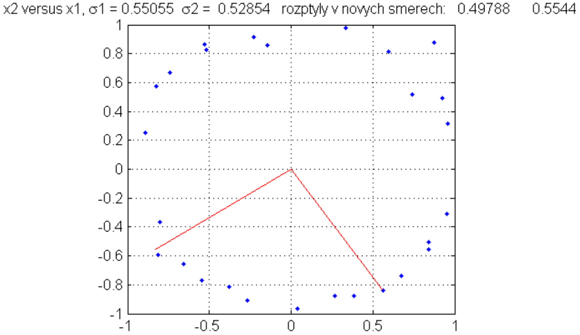
Příklad selhání PCA

Nelineární úloha - rozložení dat po kružnici



Příklad selhání PCA

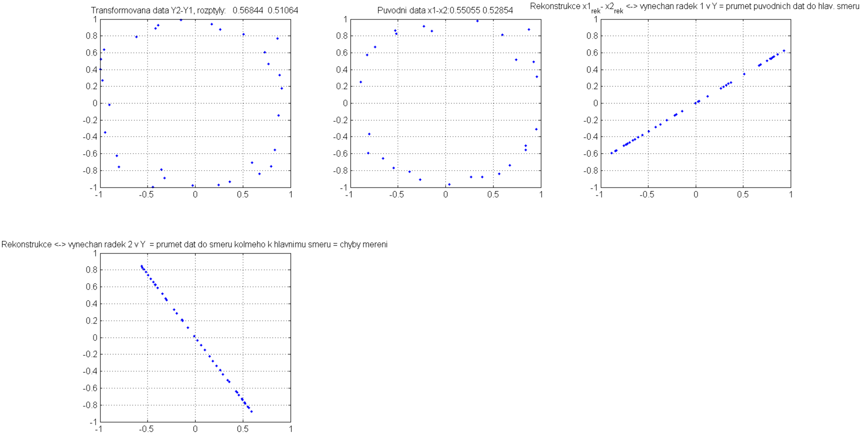
Vlastní čísla jsou stejně velká!!! $\rightarrow$ dimenzionalitu snížit nelze
Nové směry nalezeny, ale žádná redukce rozptylu !!! – PCA směry zvolila náhodně



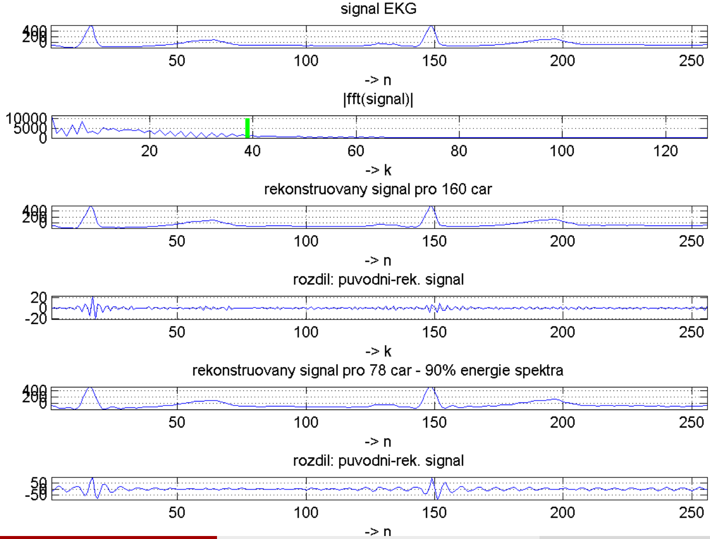
Příklad selhání PCA

Nová (rekonstruovaná) a původní data – redukce dimenz. není možná - oba průměty jsou stejně dlouhé :-(

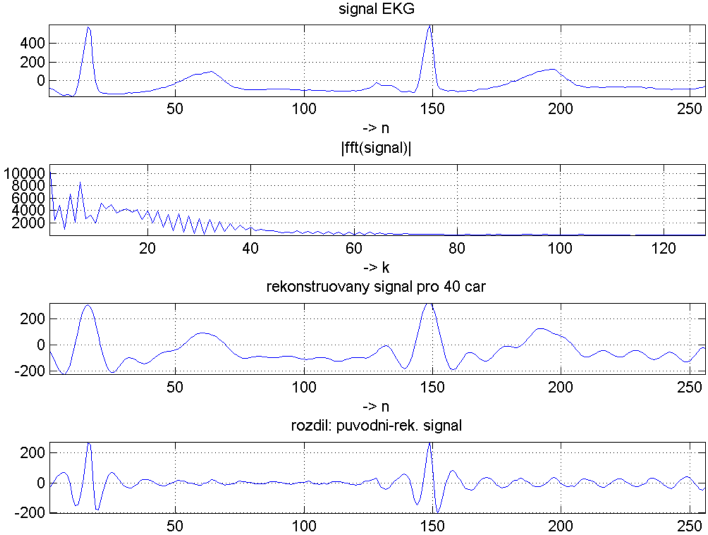
Řešení: použít nelineární transformaci dat a následně PCA nebo použít jinou metodu analýzy, která obsahuje nelinearitu



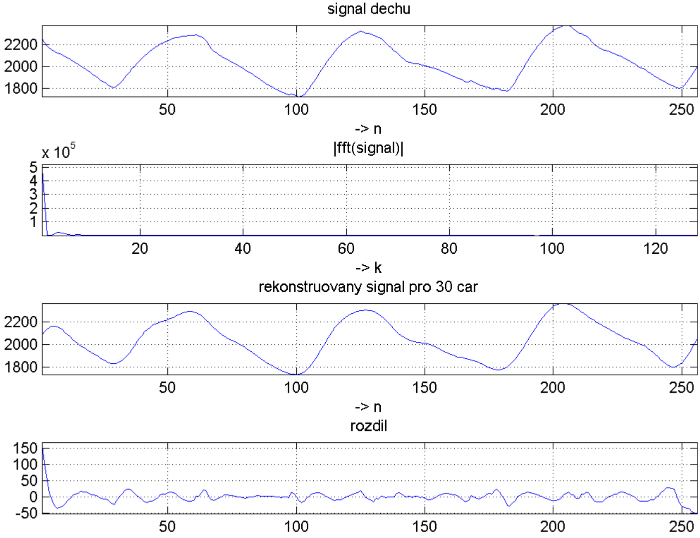
Ztrátová komprese EKG pomocí FFT



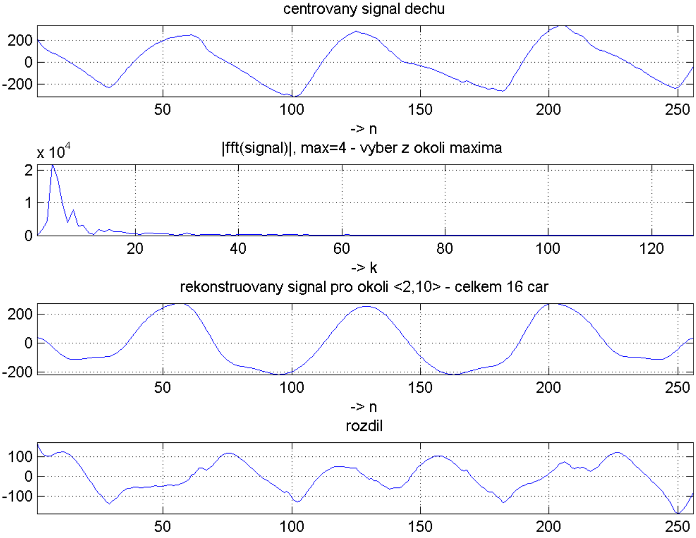
Ztrátová komprese EKG pomocí FFT



Ztrátová komprese dechu pomocí FFT

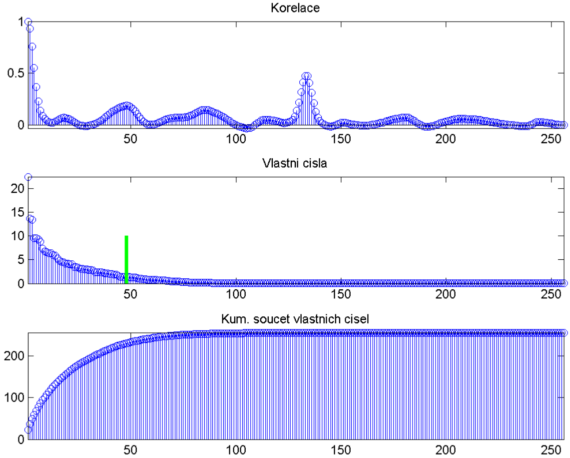


Ztrátová komprese dechu pomocí FFT



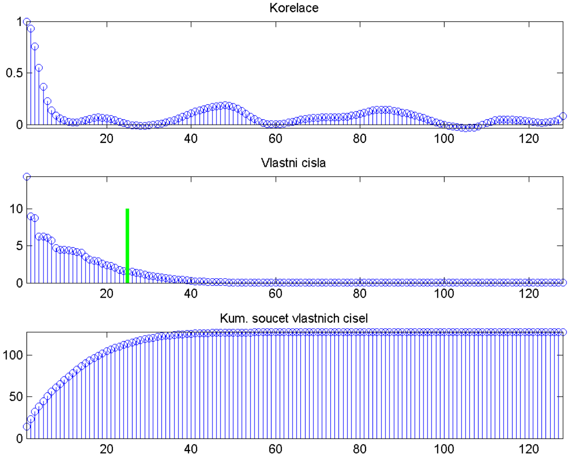
Korelace a vlastní čísla EKG

Dvě periody EKG - zelená úsečka označuje počet vlastních čísel, při kterém je dosaženo 95% původní energie signálu - to znamená, že lze provést ztrátovou rekonstrukci z 50 vlastních vektorů

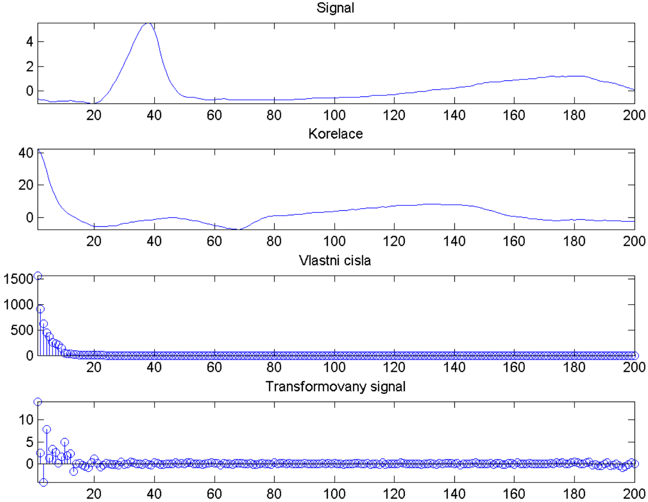


Korelace a vlastní čísla EKG

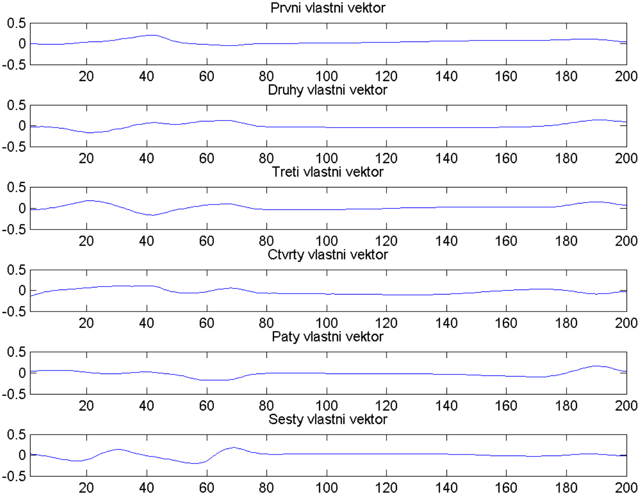
Jedna perioda EKG - zde je počet vlastních čísel a vektorů potřebných pro ztrátovou rekonstrukci nižší



Rozklad EKG pomoci PCA a jeho transformace



Vlastní vektory PCA pro 1 periodu EKG



Rekonstrukce EKG pomoci KLT

Při použití 15 vlastních vektorů (cca 65% energie signálu) je patrná značná chyba rekonstrukce především v oblasti rychlé změny (QRS komplex)