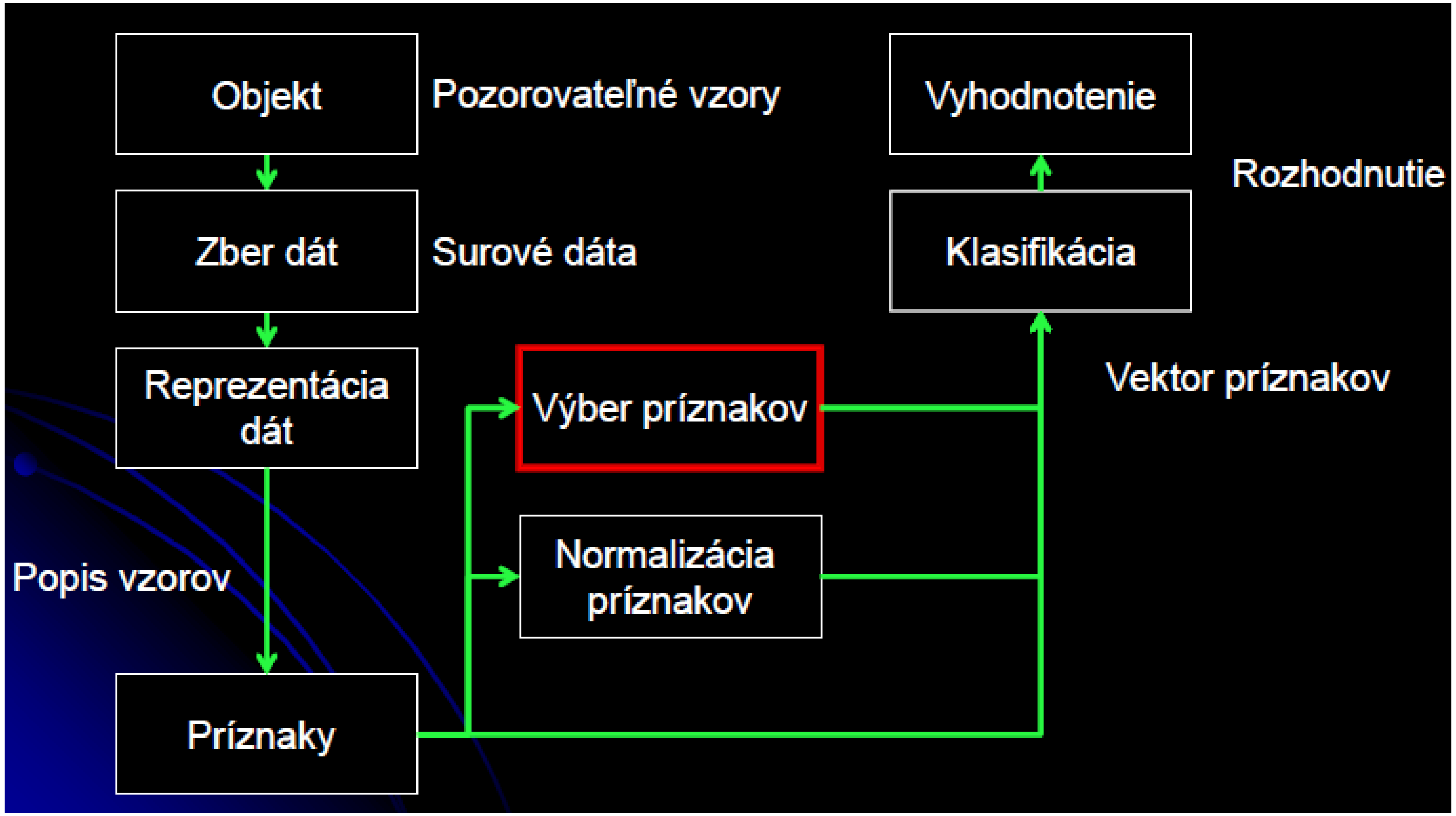
UNIVERZITA
KOMENSKÉHO
V BRATISLAVE

Rozpoznávanie Obrazcov
5. Prednáška – Redukcia príznakov

Ing. Viktor Kocur, PhD.
KAI FMFI

Výber príznakov



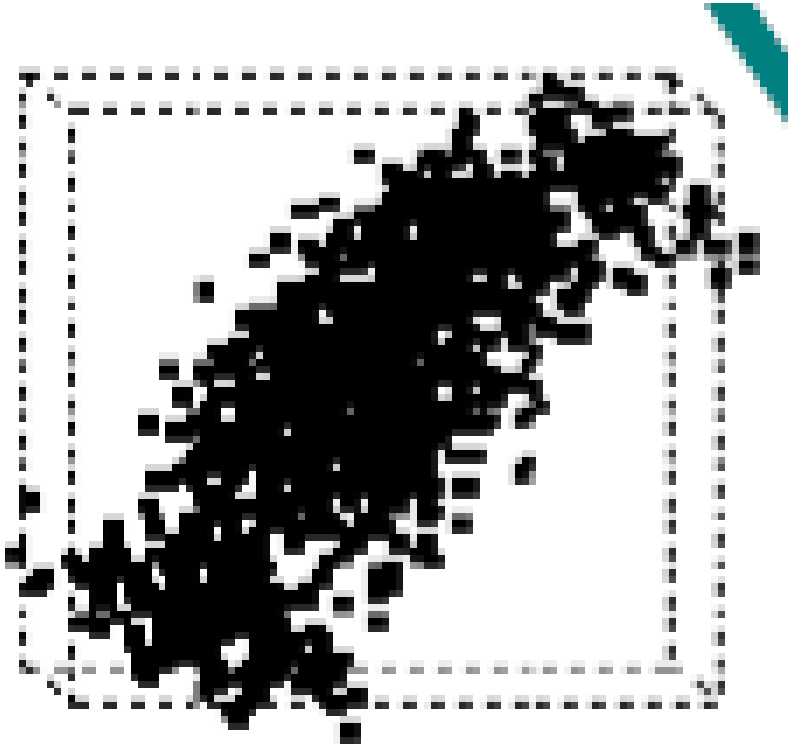
Dva prístupy

- Výber príznakov:
  - Vyberieme podmnožinu z originálnych príznakov

- Redukcia príznakov:
  - Transformujeme pôvodnú množinu príznakov do menej-dimenzionálnej

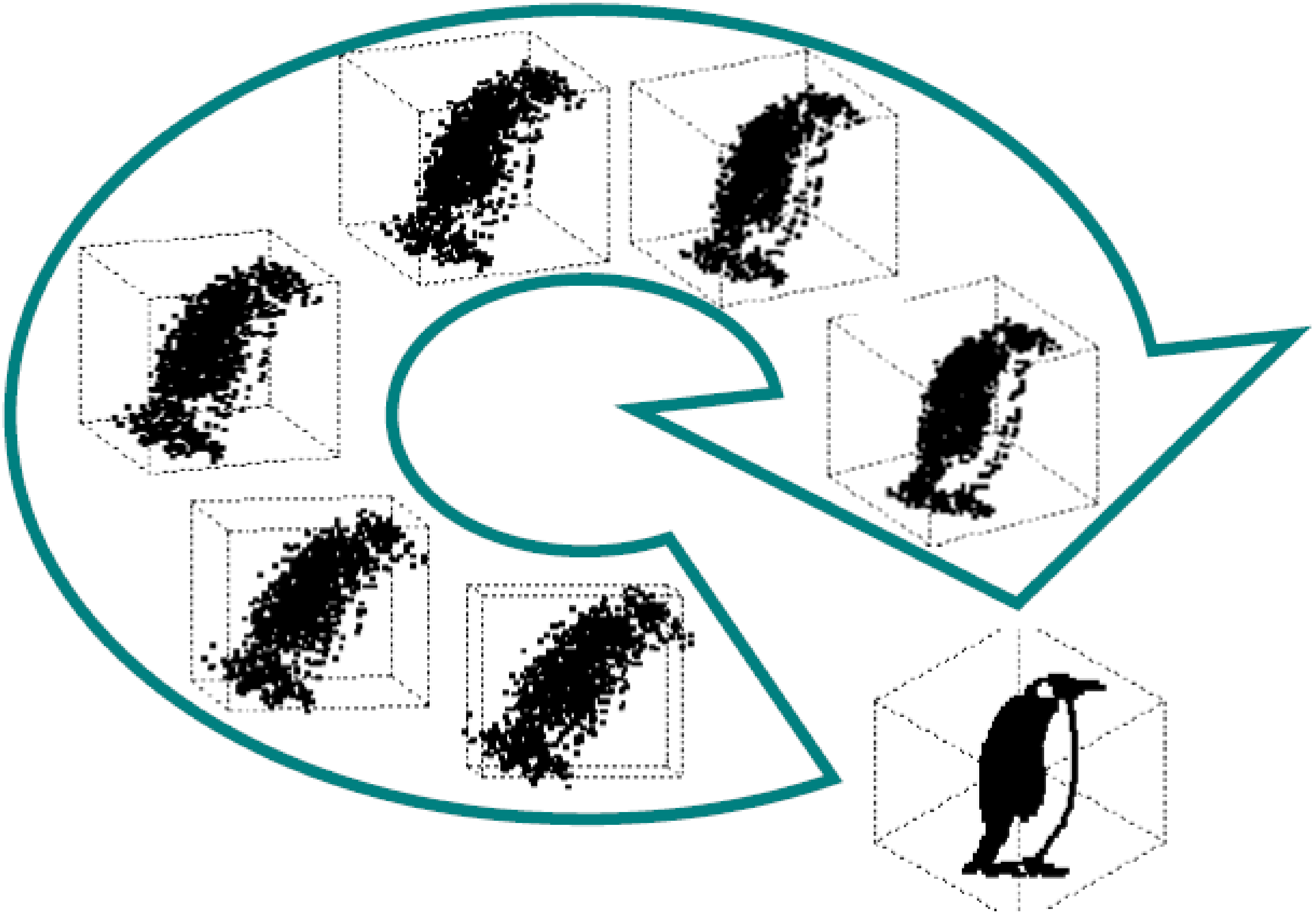
Motivácia

- Dáta v 3D priestore



- Zmysluplný smer premietania do 2D?

Motivácia II

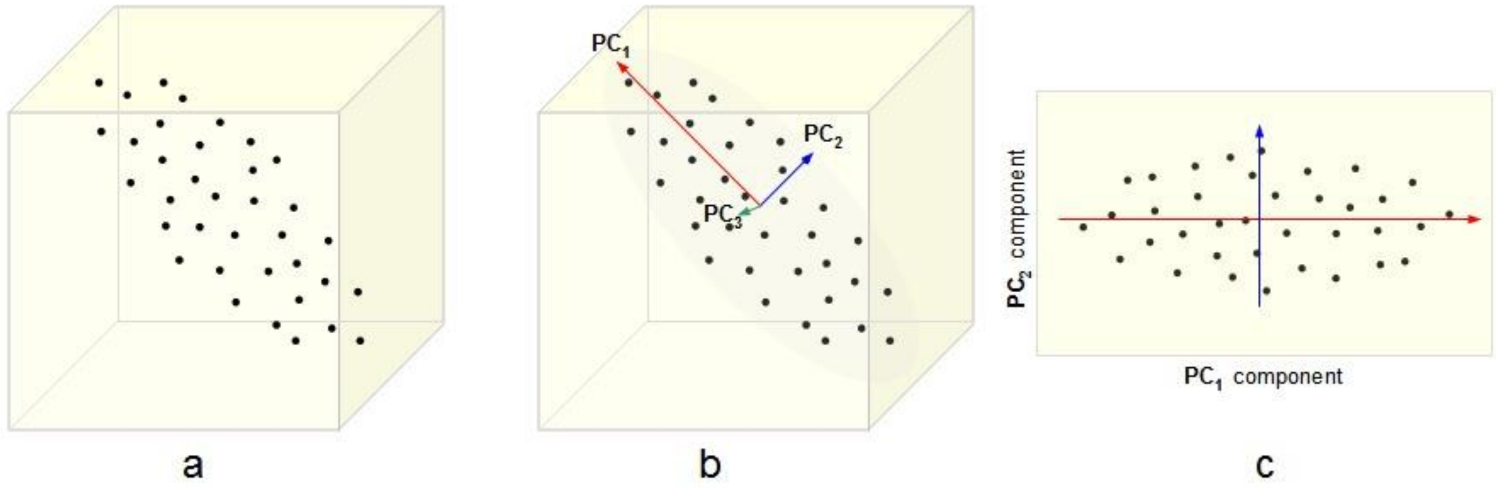


Metódy redukcie príznakov

- Neriadené (minimalizujú stratu informácie)
  - Principal Component Analysis (PCA)
  - Independent Component Analysis (ICA)
  - Latent Semantic Indexing (LSI)
- Riadené (maximalizujú medzitriedne rozdiely)
  - Linear Discriminant Analysis (LDA)
  - Canonical Correlation Analysis (CCA)
  - Partial Least Squares (PLS)

Principal Component Analysis (PCA)

- Metóda hlavných komponentov, alebo tiež Karhunen-Loeveho (K-L) metóda
- PCA hľadá „podpriestor“, ktorý zachytáva čo najviac variancie (rozptylu) v dátach



PCA II

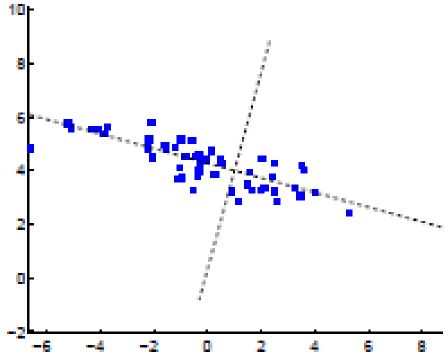
- Otočí a posunie súradnicovú sústavu tak, aby prvá os bola v smere najväčšej variability dát a ďalšie osi boli na ňu kolmé v smeroch najväčšej zvyšnej variability
- Máme N D-rozmerných vektorov príznakov $[\boldsymbol{x}_1, \dots, \boldsymbol{x}_N]$ a po aplikácii PCA dostaneme novú ortonormálnu bázu $\{\boldsymbol{b}_1, \dots, \boldsymbol{b}_D\}$, čiže bázové vektory sú jednotkové a na seba kolmé, takže platí $\boldsymbol{b}_i^T \boldsymbol{b}_j = \delta_{ij}$

PCA III

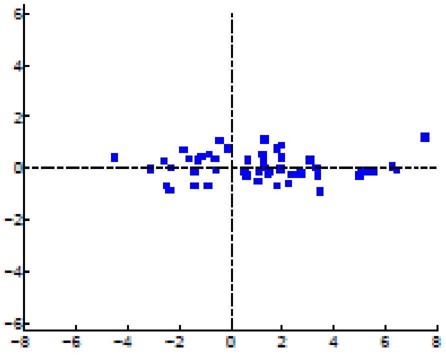
- Nová báza je lineárnou kombináciou pôvodnej bázy, vektory sú ortogonálne
- PCA predpokladá normálne rozdelenie jednotlivých príznakov
- Vychádza z toho, že príznaky s veľkým rozptylom odrážajú dynamiku dát v databáze, príznaky s malým rozptylom predstavujú šum
- PCA nevyužíva informáciu o triedach

PCA IV

- Ak pôvodné premenné (príznaky) boli možno korelované, v novej báze dostaneme lineárne nekorelované premenné, ktoré sa nazývajú hlavné komponenty
- PCA je citlivá na škálovanie pôvodných premenných



(a) Originálne dáta.



(b) Transformované dáta.

PCA V

- Cieľ PCA možno sformulovať aj takto: nájdite ortonormálnu maticu $\boldsymbol{P}$, ktorá zobrazí databázu $\boldsymbol{X}$ na $\boldsymbol{Y}$, t.j. $\boldsymbol{Y} = \boldsymbol{P}\boldsymbol{X}$, takú, že kovariančná matica $\frac{1}{N}\boldsymbol{Y}\boldsymbol{Y}^T$ bude diagonálna, teda mimo diagonály má samé 0, t.j. príznaky v matici $\boldsymbol{Y}$ sú nekorelované
- Riadky matice $\boldsymbol{P}$ sú potom hlavné komponenty databázy $\boldsymbol{X}$

PCA VI

- Priemet vektora $\boldsymbol{x}_i$ do smeru $\boldsymbol{b}_1$ je $x'_{i1} = \boldsymbol{b}_1^T \boldsymbol{x}_i$
- Priemer v stĺpci 1 je $\bar{x}'_1 = \boldsymbol{b}_1^T \overline{\boldsymbol{x}}$, kde $\overline{\boldsymbol{x}} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{x}_i$
- Variancia $Var_1 = \frac{1}{N} \sum_{i=1}^{N} (x'_{i1} - \bar{x}'_1)^2 =$

$$= \frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{b}_1^T \boldsymbol{x}_i - \boldsymbol{b}_1^T \overline{\boldsymbol{x}})^2 = \frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{b}_1^T (\boldsymbol{x}_i - \overline{\boldsymbol{x}}))^2$$

$$= \boldsymbol{b}_1^T \boldsymbol{\Sigma} \boldsymbol{b}_1$$

PCA VII

- Ak prvá os má byť v smere najväčšej variability
- Maximalizačná úloha s viazaným extrémom
- Maximalizujeme $\max_{\boldsymbol{b}_1} \boldsymbol{b}_1^T \boldsymbol{\Sigma} \boldsymbol{b}_1$ tak aby $\boldsymbol{b}_1^T \boldsymbol{b}_1 = 1$
- Lagrangeova funkcia $L = \boldsymbol{b}_1^T \boldsymbol{\Sigma} \boldsymbol{b}_1 - \lambda(\boldsymbol{b}_1^T \boldsymbol{b}_1 - 1)$
- Podľa KT podmienok $\frac{\partial L}{\partial \boldsymbol{b}_1} = 2\boldsymbol{\Sigma} \boldsymbol{b}_1 - 2\lambda \boldsymbol{b}_1 \equiv 0$
- Z toho $\boldsymbol{\Sigma} \boldsymbol{b}_1 = \lambda\, \boldsymbol{b}_1$ kde λ je vlastné číslo a platí $\boldsymbol{b}_1^T \boldsymbol{\Sigma} \boldsymbol{b}_1 = \lambda \boldsymbol{b}_1^T \boldsymbol{b}_1 = \lambda$
- Pre maximum musí byť λ najväčšie vlast. číslo

PCA VIII

- Pre druhý hlavný komponent počítame Var_2
- Maximalizujeme $\max_{\boldsymbol{b}_2} \boldsymbol{b}_2^T \boldsymbol{\Sigma} \boldsymbol{b}_2$ tak aby platilo $\boldsymbol{b}_2^T \boldsymbol{b}_2 = 1$ a súčasne $\boldsymbol{b}_1^T \boldsymbol{b}_2 = 0$
- $L = \boldsymbol{b}_2^T \boldsymbol{\Sigma} \boldsymbol{b}_2 - \lambda(\boldsymbol{b}_2^T 2 - 1) - \mu \boldsymbol{b}_1^T \boldsymbol{b}_2$
- Potom $\frac{\partial L}{\partial \boldsymbol{b}_2} = 2\boldsymbol{\Sigma} \boldsymbol{b}_2 - 2\lambda \boldsymbol{b}_2 - \mu \boldsymbol{b}_1 \equiv 0$
- Z toho sa potom $\mu = 0\ a\ \boldsymbol{\Sigma} \boldsymbol{b}_2 = \lambda\ \boldsymbol{b}_2$ kde λ je vlastné číslo, ktoré pre maximum Var_2 musí byť druhé najväčšie

PCA IX

- Smery nových bázových vektorov $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_D\}$
- Nech je počiatok súradnicovej sústavy p, potom $\boldsymbol{y}_i = \mathbf{p} + \sum_{i=1}^{N} y_{ij}\boldsymbol{b}_j$
- Chceme, aby suma štvorcov vzdialeností medzi pôvodnými a premietnutými vektormi bola minimálna

$$E = \sum_{i=1}^{N} ||\mathbf{x}_i - \mathbf{y}_i||^2 = \sum_{i=1}^{N} ||\mathbf{x}_i - (\mathbf{p} + \sum_{j=1}^{N} y_{ij}\mathbf{b}_j)||^2 =$$

- Potom

$$= \sum_{i=1}^{N} ||\mathbf{x}_i - \mathbf{p}||^2 - 2\sum_{j=1}^{N}\sum_{i=1}^{N} y_{ij}\mathbf{b}_j^T(\mathbf{x}_i - \mathbf{p}) + \sum_{j=1}^{N}\sum_{i=1}^{N} y_{ij}^2.$$

$$y_{ij} = \mathbf{b}_j^T(\mathbf{x}_i - \bar{\mathbf{x}}) \qquad \mathrm{p} = \bar{\mathrm{x}}.$$

Použitie PCA

- 1. Zoberieme databázu $\boldsymbol{X}$
- 2. Odčítame priemer od každého rozmeru prvku databázy
- 3. Vypočítame kovariančnú maticu
- 4. Vypočítame vlastné čísla a vlastné vektory tejto kovariančnej matice
- 5. Vyberieme komponenty a príznakový vektor
- 6. Vypočítame novú databázu

Použitie PCA II

- 3. Vypočítame kovariančnú maticu

Vypočítame maticu $\mathbf{\Sigma}$

$$\mathbf{\Sigma} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T = \frac{1}{N}\mathbf{X}\mathbf{X}^T,$$

kde $[\mathbf{X}]_{D\times N} = [\mathbf{x}_1 - \bar{\mathbf{x}}, \ldots \mathbf{x}_N - \bar{\mathbf{x}}]$

- 4. Vypočítame vlastné čísla a vlastné vektory tejto kovariančnej matice $\boldsymbol{\Sigma b}_j = \lambda \boldsymbol{b}_j$

Použitie PCA III

- 6. Vypočítame novú databázu

Vypočítame súradnice premietnutých vektorov

$$\mathbf{x}'_i = \sum_{j=1}^{D} \mathbf{b}_j^T (\mathbf{x}_i - \bar{\mathbf{x}}) = \mathbf{B}^T (\mathbf{x}_i - \bar{\mathbf{x}}) = \mathbf{B}^T \mathbf{X},$$

kde $\mathbf{B} = [\mathbf{b}_1, \ldots \mathbf{b}_D]$.

- Treba dať pozor na neporovnateľné jednotky (výška, váha, teplota) v databáze

Ak počet pozorovaní $N \leq D$

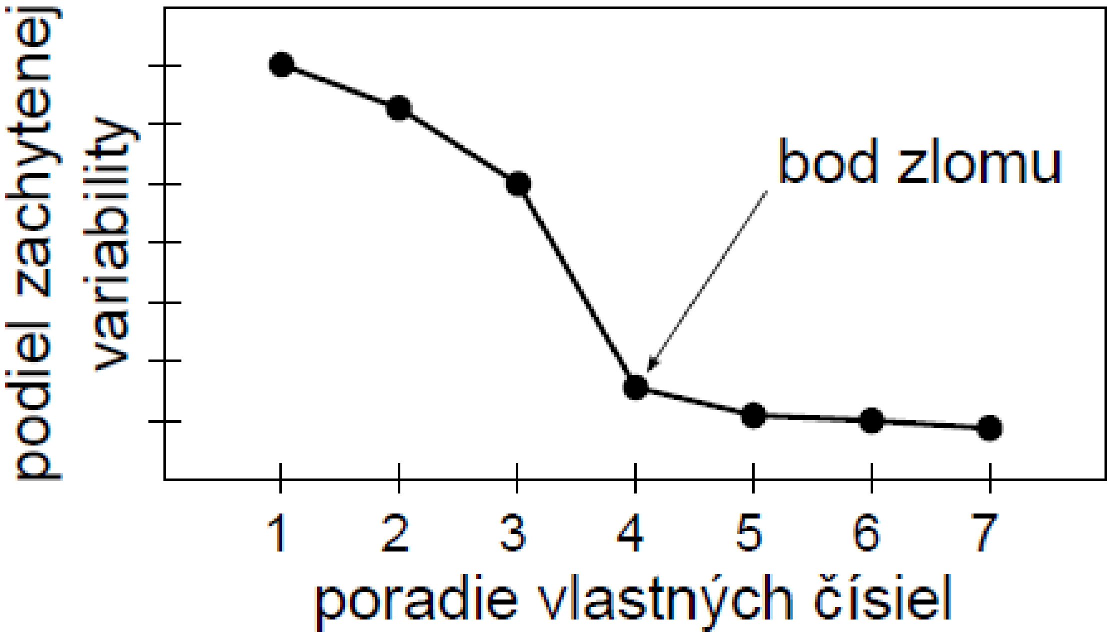
- Potom je najviac r vlastných hodnôt nenulových a $r = rank(\boldsymbol{X}) = rank(\boldsymbol{\Sigma}) \leq N$
- Kovariančná matica $\boldsymbol{\Sigma} = \frac{1}{N}\boldsymbol{X}\boldsymbol{X}^T$ má rozmer $D \times D$ a zložitosť výpočtu $O(D^3)$
- Platí $\boldsymbol{\Sigma}\boldsymbol{b}_j = \lambda_j \boldsymbol{b}_j$ t.j. $\frac{1}{N}\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{b}_j = \lambda_j \boldsymbol{b}_j$
- Potom $\frac{1}{N}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{b}_j = \lambda_j \boldsymbol{X}^T\boldsymbol{b}_j$ a z toho $\frac{1}{N}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{p}_j = \lambda_j \boldsymbol{p}_j$

Ak počet pozorovaní $N \leq D$ II

- kde $\boldsymbol{p}_j = \boldsymbol{X}^T \boldsymbol{b}_j$ je vlastný vektor $N \times N$ matice $\frac{1}{N} \boldsymbol{X}^T \boldsymbol{X}$ pričom časová zložitosť je $O(N^3)$
- Ak $\frac{1}{N} \boldsymbol{X}\boldsymbol{X}^T \boldsymbol{X}\boldsymbol{p}_j = \lambda_j \boldsymbol{X}\boldsymbol{p}_j$ tak $\boldsymbol{\Sigma}\boldsymbol{X}\boldsymbol{p}_j = \lambda_j \boldsymbol{X}\boldsymbol{p}_j$
- Z toho potom $\boldsymbol{b}_j \propto \boldsymbol{X}\boldsymbol{p}_j$ je vlastný vektor matice $\boldsymbol{\Sigma}$ pre dané λ
- Potom hľadáme také $\boldsymbol{b}_j$ že $\|\boldsymbol{b}_j\| = 1$ a $\boldsymbol{b}_j \propto \boldsymbol{X}\boldsymbol{p}_j$

Zníženie počtu príznakov

- Latencia pri PCA odráža variabilitu príznakov – na obrázku v bode zlomu, od štvrtého najväčšieho, začínajú vlastné čísla, zachycujúce málo variability

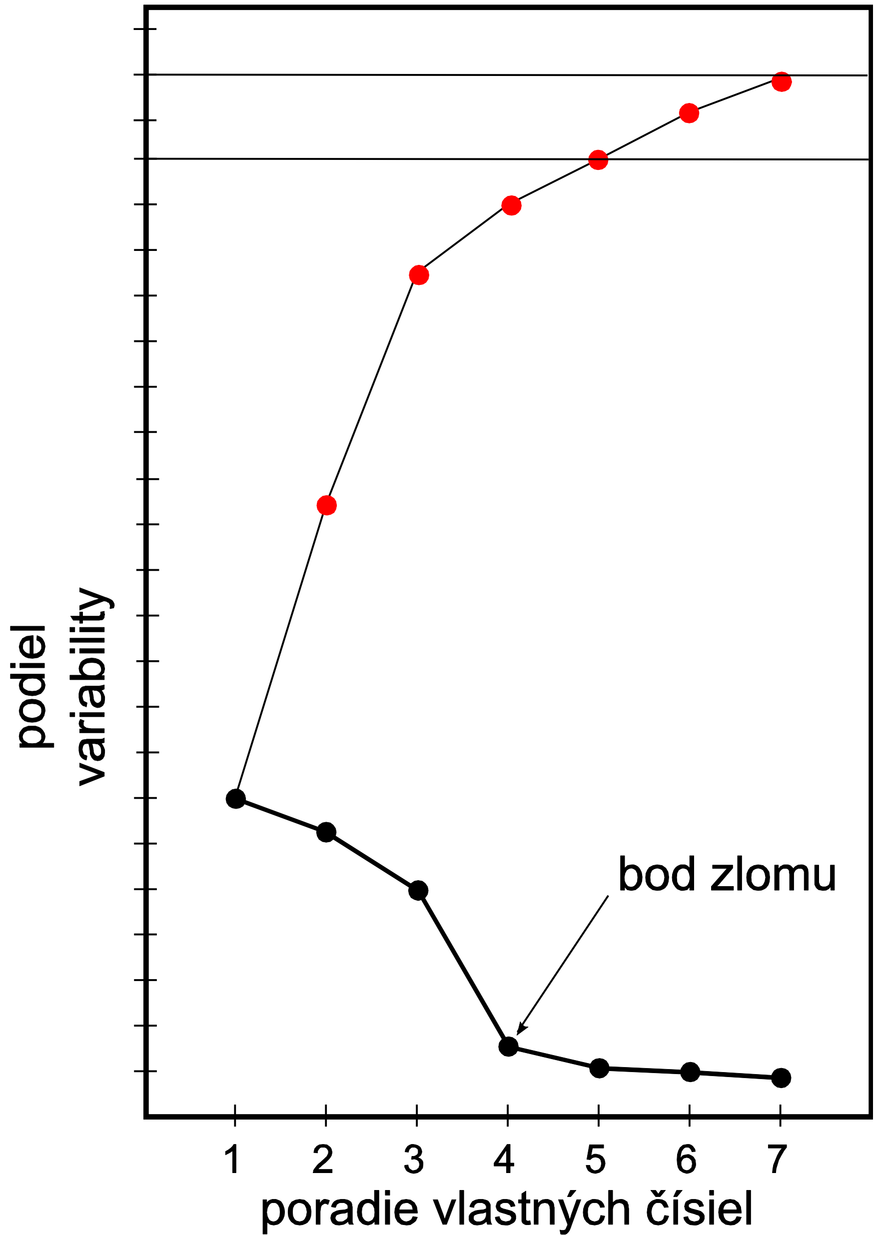


Zníženie počtu príznakov II

- Podiel variability v j-tom hlavnom komponente $\frac{\lambda_j}{\sum_{j=1}^{D} \lambda_j}$,

- Kumulatívny pomer achytenej variability

$$\frac{\sum_{j=1}^{K} \lambda_j}{\sum_{j=1}^{D} \lambda_j} > 0.9 \text{ alebo } 0.95.$$



Andersonov test

- Testovanie hypotézy

$$\begin{aligned} \mathrm{H}_0 &: \quad \lambda_{K+1} = \lambda_{K+2} = \ldots = \lambda_D \\ \mathrm{H}_1 &: \quad \text{neplatí } \mathrm{H}_0. \end{aligned}$$

Testovacia štatistika je $V = n(D-K)\ln\frac{a}{c}$:

$$a = \frac{\sum_{j=K+1}^{D}\lambda_j}{D-K} \qquad c = \left(\prod_{j=K+1}^{D}\lambda_j\right)^{\frac{1}{D-K}}$$

χ^2 rozdelenie s $0.5(D-K+2)(D-K-1)$ stupňami voľnosti

SVD

- Naspäť k problému $N < D$
- Singulárny rozklad matice A – každá matica sa rozložiť na súčin $\mathbf{A} = \mathbf{U}\mathbf{S}\boldsymbol{V}^T$
  $[\mathbf{A}]_{N\times D} = [\mathbf{U}]_{N\times N}[\mathbf{S}]_{N\times D}[\mathbf{V}]_{D\times D}$
  $\mathbf{U} = [\boldsymbol{u}_1, \dots, \boldsymbol{u}_N]$ a $\mathbf{V} = [\boldsymbol{v}_1, \dots, \boldsymbol{v}_D]$
  $\mathbf{S} = \mathrm{diag}(\sigma_1, \dots, \sigma_{rank(A)})$, kde
  $\boldsymbol{v}$ je vlastný vektor matice $\mathbf{A}^T\mathbf{A}$
  $\boldsymbol{u}$ je vlastný vektor matice $\mathbf{A}\mathbf{A}^T$ a $\lambda = \sigma^2$

Vzťah medzi PCA a SVD

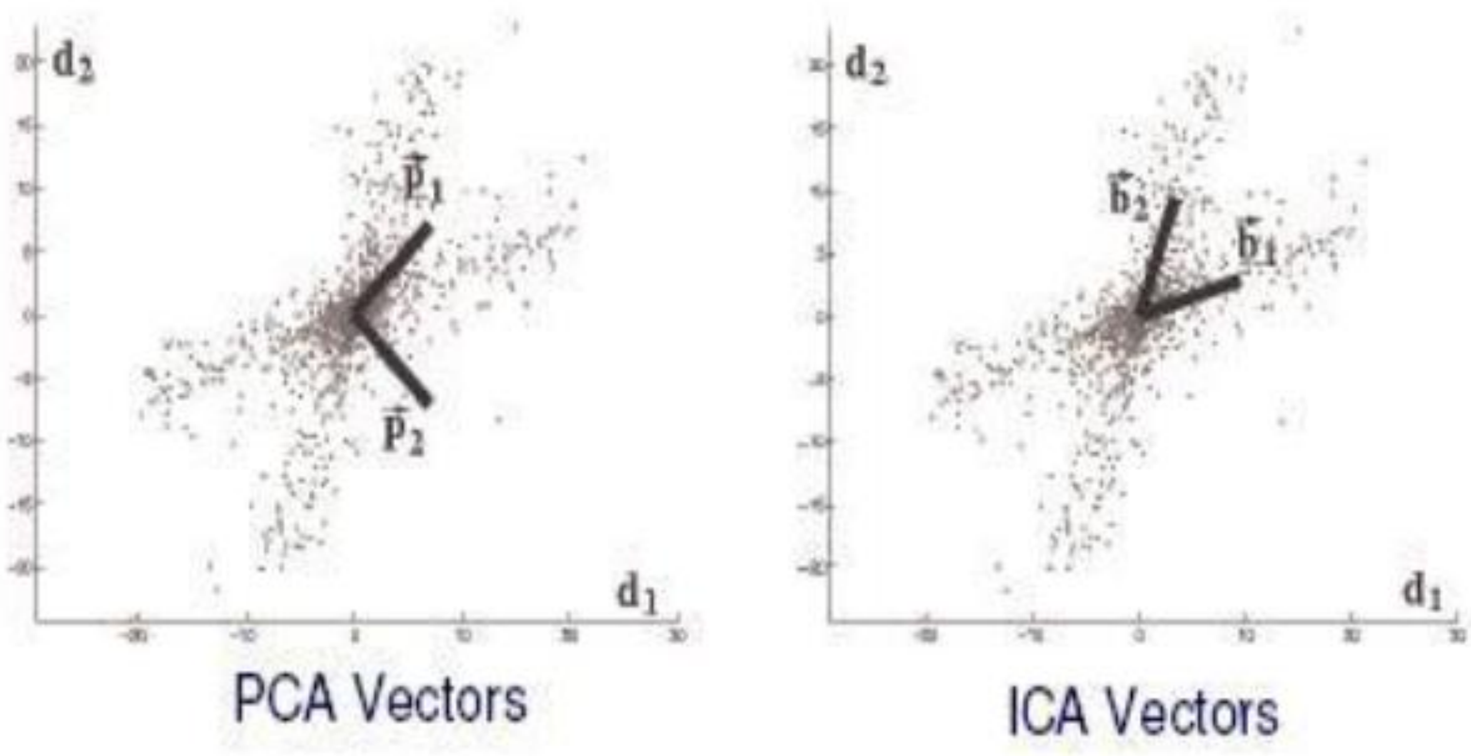
- $\mathbf{X} = \ [\boldsymbol{x}_1 - \overline{\boldsymbol{x}}, \dots, \boldsymbol{x}_N - \overline{\boldsymbol{x}}]$
- $\mathbf{Y} = \ \frac{1}{\sqrt{N}} \mathbf{X}^T \Longrightarrow \mathbf{Y}^T \mathbf{Y} = \frac{1}{N} \mathbf{X} \mathbf{X}^T = \mathbf{\Sigma}$
- Pre $\mathbf{Y} = \mathbf{U S V}^T$ sú $\mathbf{V}$ vlastné vektory matice $\mathbf{Y}^T \mathbf{Y} = \mathbf{\Sigma}$
- Často používame SVD namiesto PCA
- SVD algoritmus je numericky stabilnejší ako výpočet PCA z predchádzajúcich slidov

Sila a slabosť PCA

- Je to neparametrická analýza, odpoveď je jednoznačná bez vstupu používateľa
- Pri parametrickom prístupe sa využije znalosť o charaktere dát a najprv ich transformujeme nelineárnou transformáciou – ide o tzv. kernel PCA
- Ak upustíme od ortogonality osí, resp. od Gaussovského charakteru dát, dostaneme ICA

ICA (Independent CA)

- Metóda nezávislých komponentov
- Nevyžaduje ortogonalitu, len linearitu

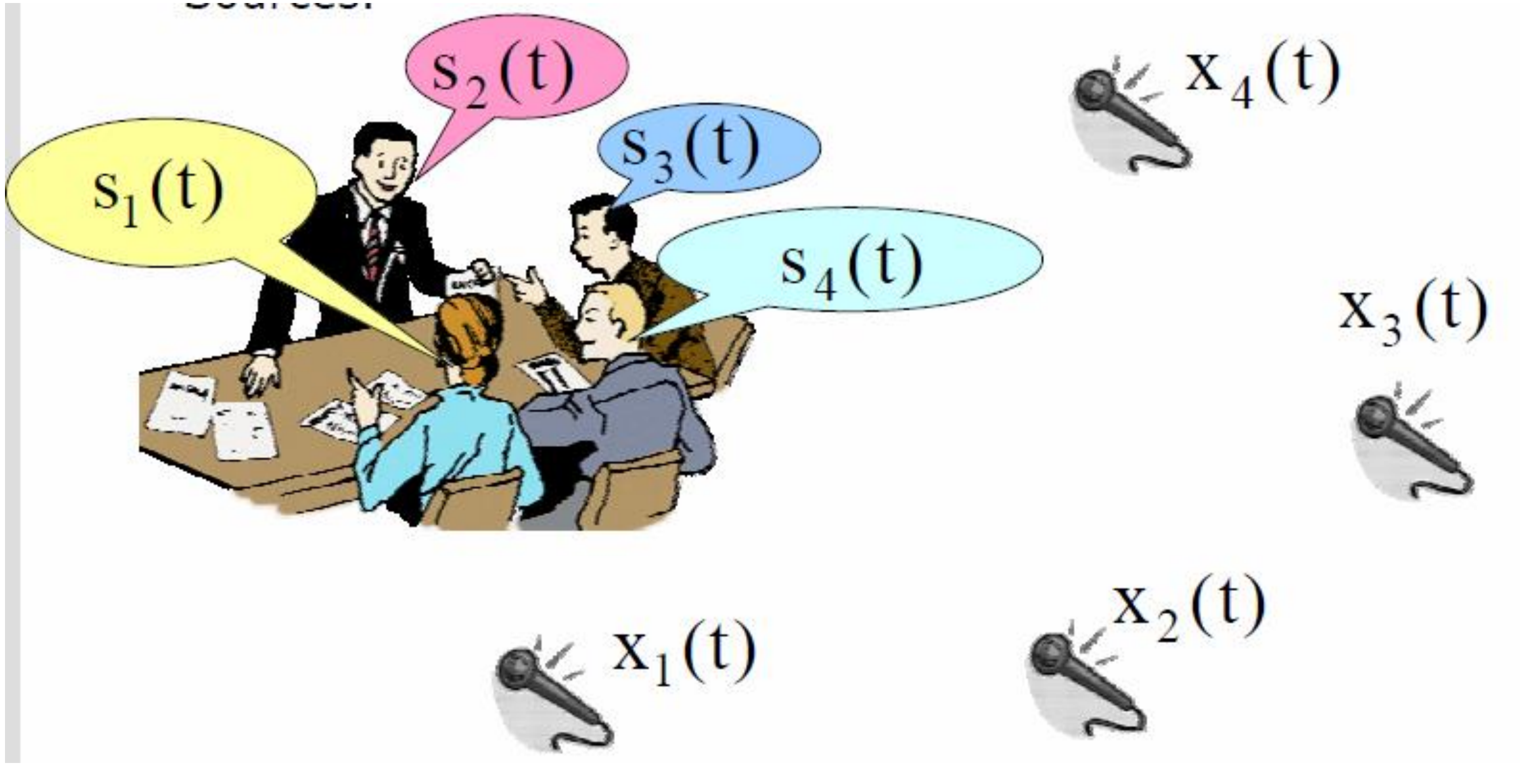


ICA II

- ICA – reprezentuje vektor $\boldsymbol{X}$ ako lineárnu kombináciu negaussovských náhodných premenných, ktoré sú čo najnezávislejšie
- Opäť sa snažíme nájsť maticu s novou bázou tak, aby kovariančná matica nového vektora bola diagonálna
- Naviac chceme, aby nové premenné boli nezávislé, t.j. $P\left(\boldsymbol{y}_i, \boldsymbol{y}_j\right) = P(\boldsymbol{y}_i)P(\boldsymbol{y}_j)$

Príklad

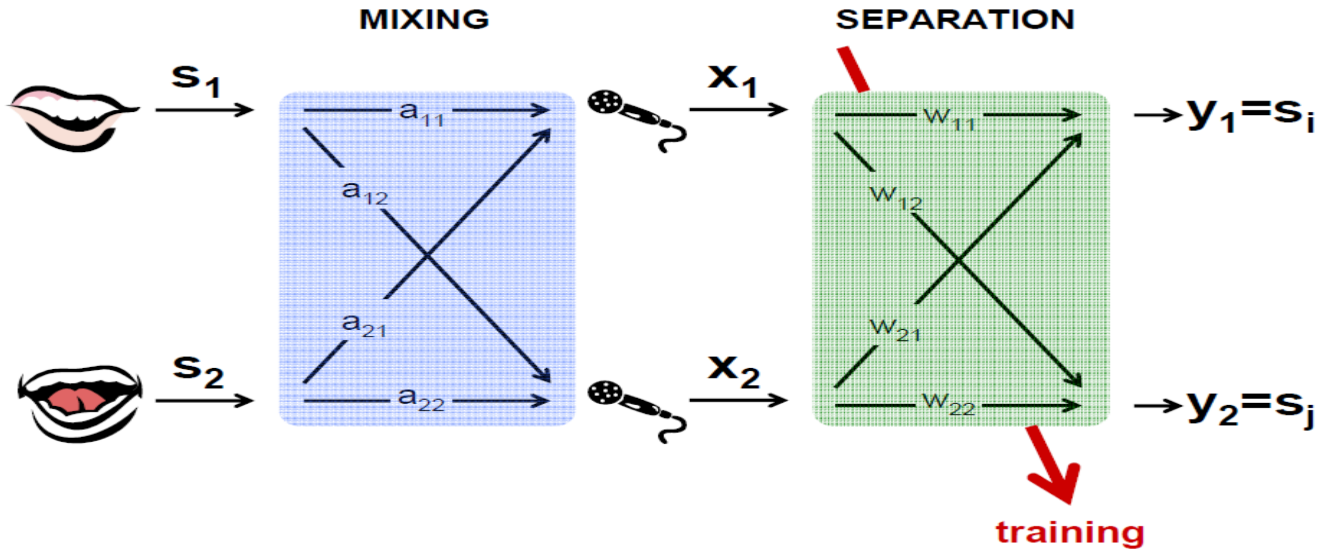
- Cocktail party problem – pozorované vstupy do mikrofónov, treba nájsť zdroje



Príklad II

- $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s}$

$$
\begin{aligned}
x_1(t) &= a_{11}s_1(t) + a_{12}s_2(t) + a_{13}s_3(t) + a_{14}s_4(t) \\
x_2(t) &= a_{21}s_1(t) + a_{22}s_2(t) + a_{23}s_3(t) + a_{24}s_4(t) \\
x_3(t) &= a_{31}s_1(t) + a_{32}s_2(t) + a_{33}s_3(t) + a_{34}s_4(t) \\
x_4(t) &= a_{41}s_1(t) + a_{42}s_2(t) + a_{43}s_3(t) + a_{44}s_4(t)
\end{aligned}
$$



Príklad III

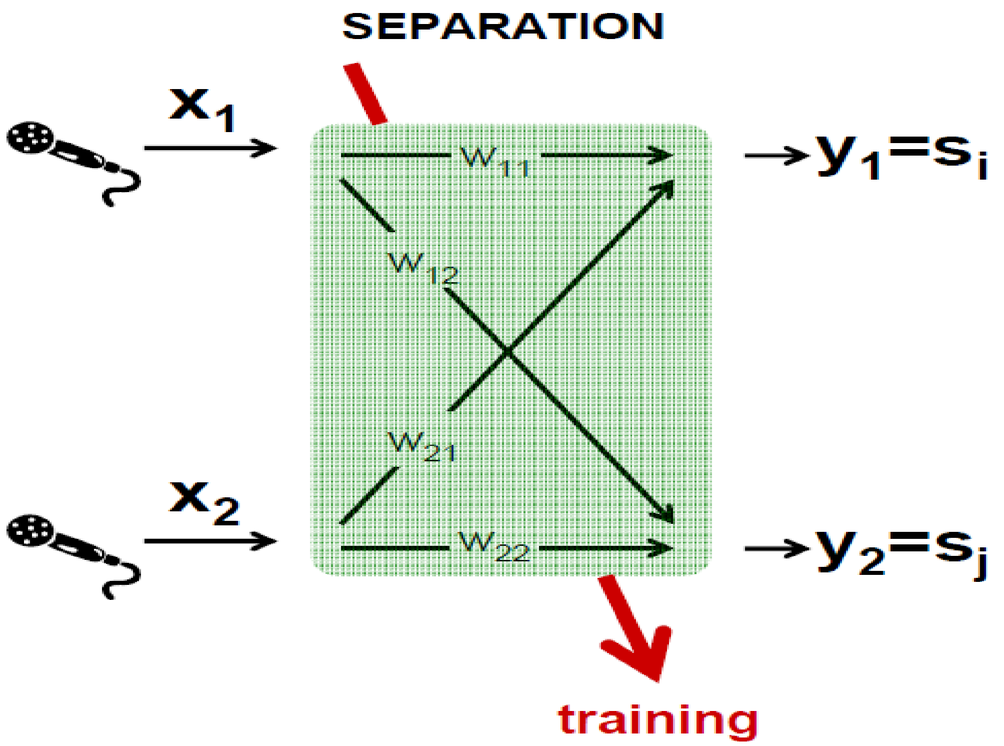
- Predpoklad:
- $P(s_1, s_2, \ldots, s_n) = P(s_1)P(s_2) \ldots P(s_n)$
- $E(s_i) = 0$
- $Var(s_i) = 1$
- neGaussovskosť dát
- Nezávislosť → nekorelovanosť → diagonálna kovariančná matica

ICA postup

- Centrovanie $\boldsymbol{x}' = \boldsymbol{x} - \overline{\boldsymbol{x}}$
- Bielenie $\boldsymbol{y} = \boldsymbol{B}\boldsymbol{x}'$, aby $\Sigma_{\boldsymbol{y}} = \boldsymbol{I}$
  - Postup $\boldsymbol{y} = \mathbf{V}\mathbf{D}\mathbf{V}^T\boldsymbol{x}'$
  - kde $\Sigma = \mathbf{V}\mathbf{S}\mathbf{V}^T$ a $d_{ij} = \sqrt{s_{ij}}$
- $\mathbf{V}$ a $\mathbf{D}$ sú vlastné vektory a diagonálna matica vlastných čísel z PCA

ICA postup II

- Hľadáme smery $\boldsymbol{w}_i$, aby sme maximalizovali neGaussovskosť



ICA postup III

- Miera neGaussovskosti: šikmosť a špicatosť (3. a 4. moment)
- Negentropia $J(y) = S(y_G) - S(y)$
  aproximácia $J(y) \propto (E(G(y) - E(G(y_G))^2$
  $$G_1(u) = \frac{1}{a_1} \log cosh a_1 u \quad G_2(u) = -\exp(-\frac{u^2}{2})$$
- Tieto funkcie sa rýchlo rátajú, na rozdiel od momentov
  $1 < a_1 < 2$

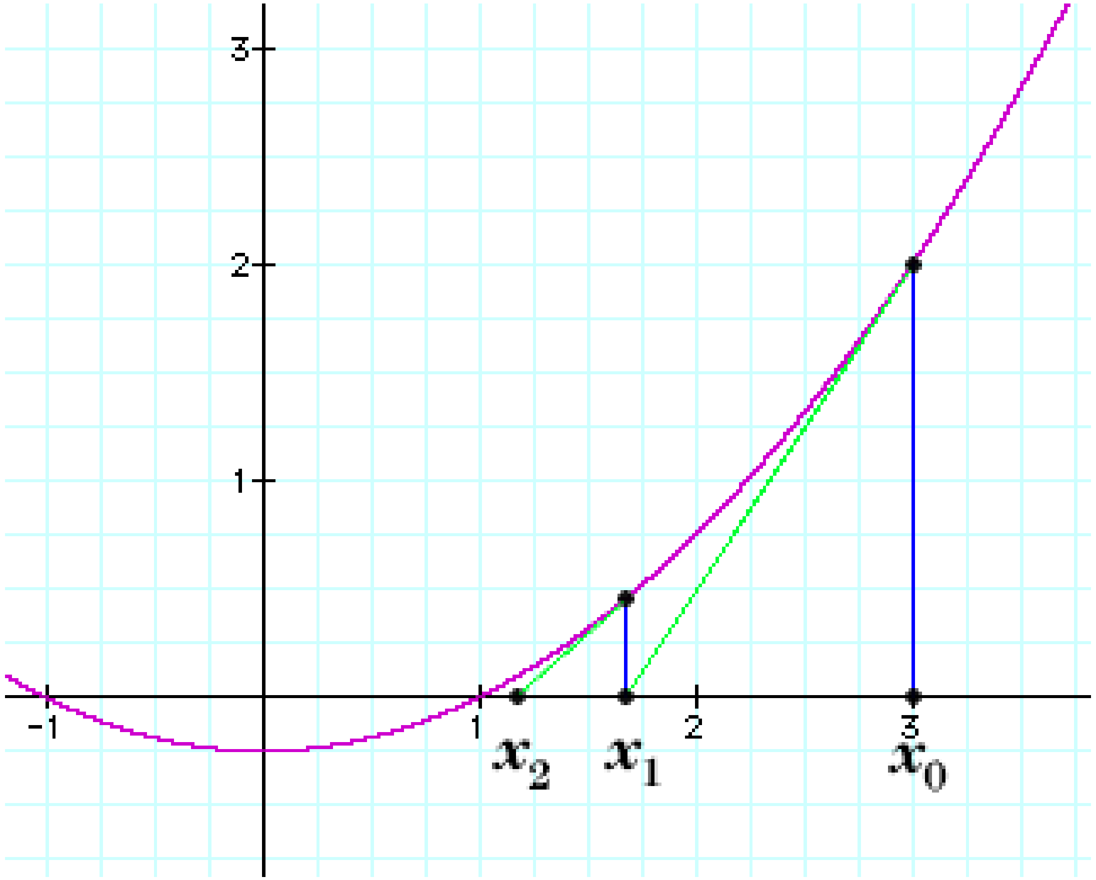
ICA postup IV

- Hľadáme $\boldsymbol{w}$, aby maximalizoval $J(\boldsymbol{w}^T\boldsymbol{x}) \propto (E\left(G(\boldsymbol{w}^T\boldsymbol{x})\right) - E(G(\boldsymbol{y}_G))^2$ za podmienky $||w||^2 = 1$ konštanta pre $\boldsymbol{w}$
- Lagrange $L = E\left(G(\boldsymbol{w}^T\boldsymbol{x})\right) - \lambda(\boldsymbol{w}^T\boldsymbol{w} - 1)$
- $\frac{\partial L}{\partial \boldsymbol{w}} = f(\boldsymbol{w}) = E\left(\boldsymbol{x}.g(\boldsymbol{w}^T\boldsymbol{x})\right) - \lambda\boldsymbol{w} \equiv 0$
- Riešenie použitím Newtonovej metódy

ICA postup V

- Newtonova metóda hľadania koreňov

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$



FastICA

- Algoritmus pre jeden smer:

1. Zvoľ náhodný počiatočný vektor $\boldsymbol{w}$
2. $\boldsymbol{w}^{+} = E\{\boldsymbol{x}g(\boldsymbol{w}^{T}\boldsymbol{x})\} - E\{g^{´}(\boldsymbol{w}^{T}\boldsymbol{x})\}\boldsymbol{w}$
3. $\boldsymbol{w} = \boldsymbol{w}^{+}/||\boldsymbol{w}^{+}||$
4. Opakuj kroky 2 a 3, kým nenastane konvergencia

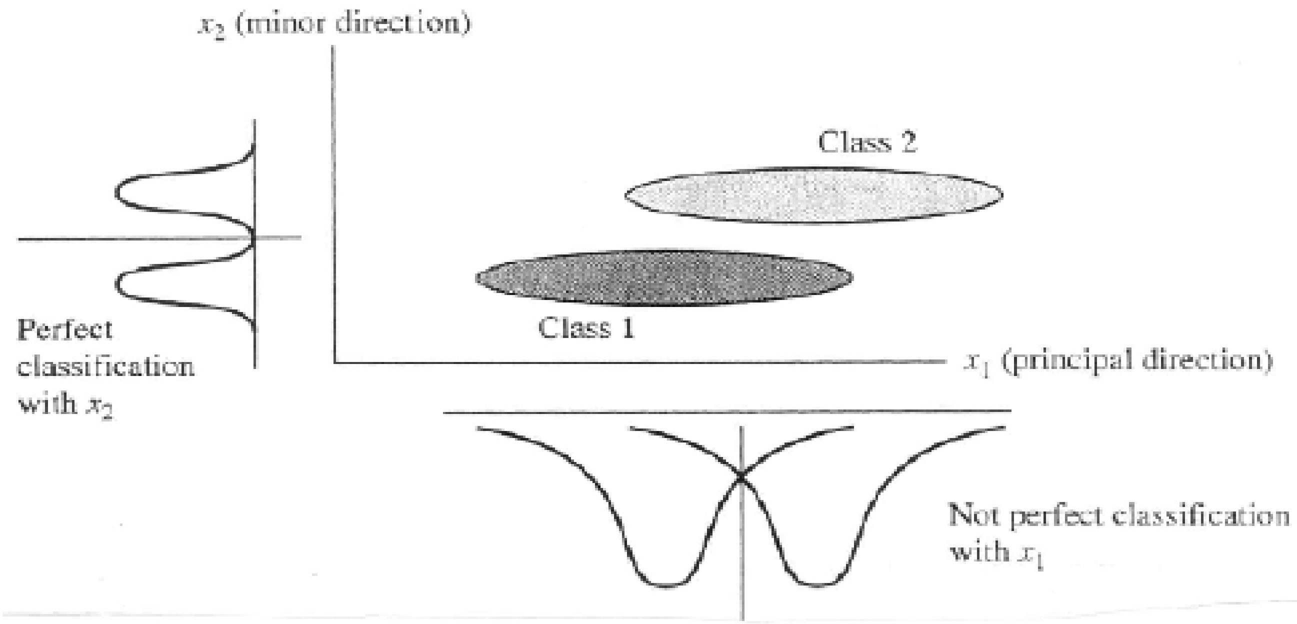
FastICA II

- Algoritmus pre viacero smerov:
- FastICA postupne pre každý smer, ale pri každej iterácii dekorelujeme ten nový smer

$$\mathbf{w}_{p+1} = \mathbf{w}_{p+1} - \sum_{j=1}^{p} \mathbf{w}_{p+1}^T \mathbf{w}_j \mathbf{w}_j$$
$$\mathbf{w}_{p+1} = \mathbf{w}_{p+1} / \sqrt{\mathbf{w}_{p+1}^T \mathbf{w}_{p+1}}$$

Ešte raz slabosti PCA

- PCA nie je vždy optimálne pre klasifikáciu
- PCA neberie do úvahy príslušnosť k triedam



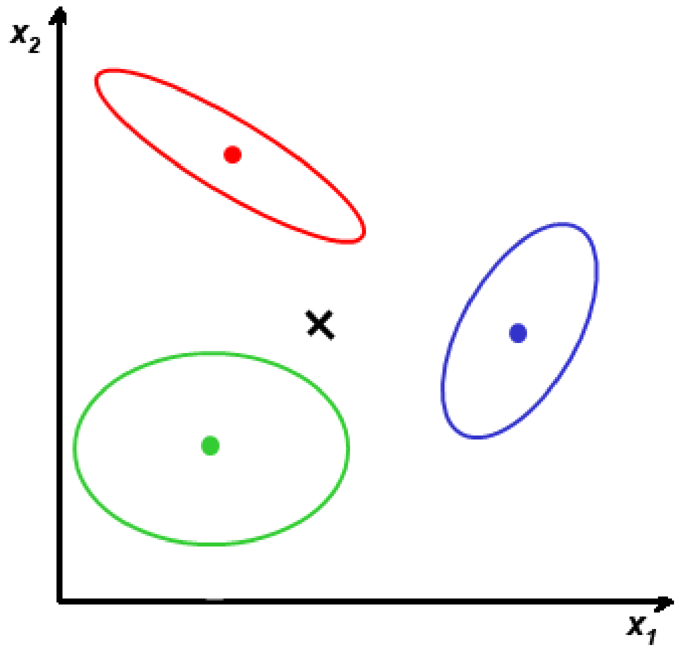
Linear Discriminant Analysis (LDA)

- Načo slúži LDA – lineárna diskriminančná analýza?
- Na zníženie dimenzie (počtu príznakov) pri zachovaní separovateľnosti tried
  - zníži dimenziu na počet tried - 1
- Nájde smer, pri ktorom sú triedy najlepšie oddelené
- Uvažuje vzťahy vo vnútri tried aj medzi triedami

Fischerova lineárna diskriminačná analýza

- Je to riadená metóda – využíva informáciu o klasifikačných triedach ω_j
- Máme N D-rozmerných príznakových vektorov $\boldsymbol{x}_1, \dots, \boldsymbol{x}_N$
- $\{\omega_j\}_{j=1}^{C}$ a do ω_j patrí N_j vektorov

$$\bar{\mathbf{x}}_j = \frac{1}{|\omega_j|} \sum_{\mathbf{x} \in \omega_j} \mathbf{x} = \frac{1}{N_j} \sum_{\mathbf{x} \in \omega_j} \mathbf{x} \qquad \bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i$$



LDA III

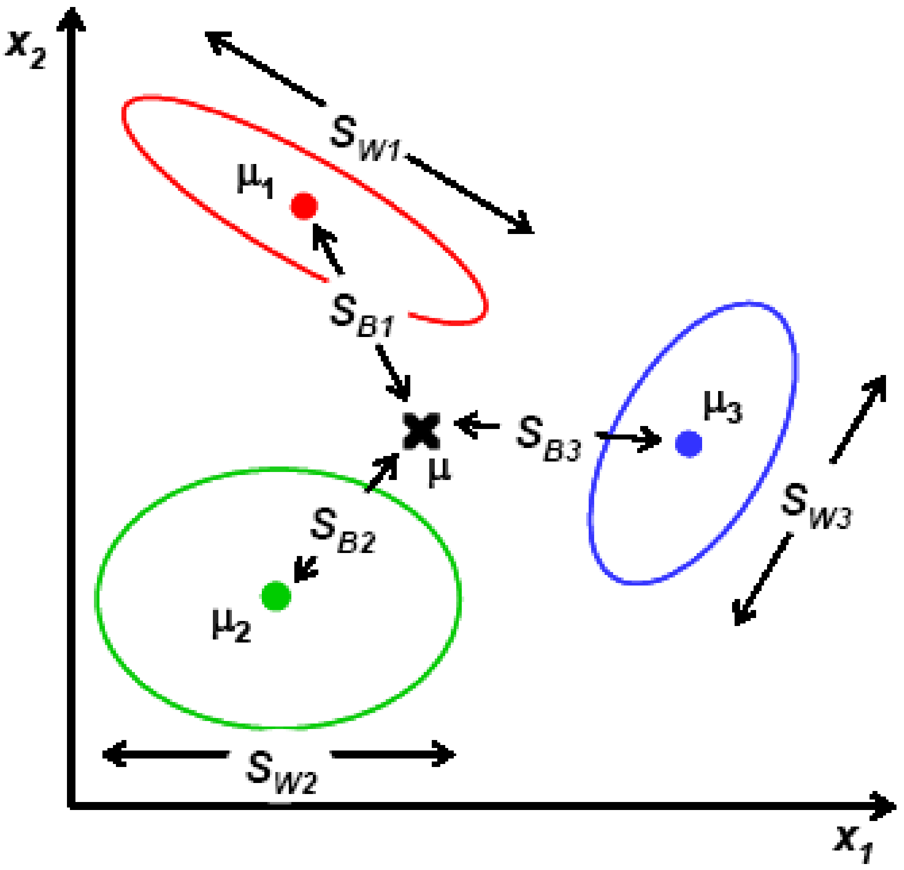
Celková variabilita v dátach je

$$\mathbf{S} = \sum_{i=1}^{N} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T.$$

$$\begin{aligned}
\mathbf{S} &= \mathbf{S}_M + \mathbf{S}_V \\
\mathbf{S}_M &= \sum_{j=1}^{C} N_j (\bar{\mathbf{x}}_\mathbf{j} - \bar{\mathbf{x}})(\bar{\mathbf{x}}_\mathbf{j} - \bar{\mathbf{x}})^T \\
\mathbf{S}_V &= \sum_{j=1}^{C} \sum_{\mathbf{x} \in \omega_j} (\mathbf{x} - \bar{\mathbf{x}}_j)(\mathbf{x} - \bar{\mathbf{x}}_j)^T = \sum_{j=1}^{C} \mathbf{S}_j
\end{aligned}$$

Variabilita v jednotlivých triedach je

$$\mathbf{S}_j = \sum_{\mathbf{x} \in \omega_j} (\mathbf{x} - \bar{\mathbf{x}}_j)(\mathbf{x} - \bar{\mathbf{x}}_j)^T.$$



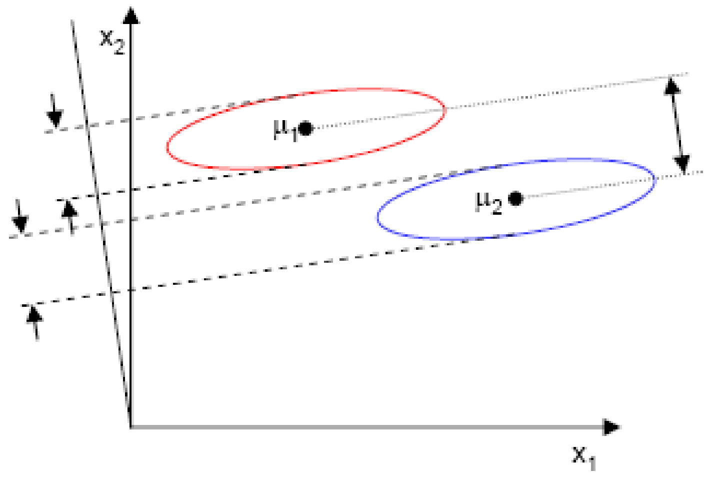
LDA IV

- Hľadáme smer $\boldsymbol{w}$, ktorý by nám pomohol najlepšie klasifikovať príznaky do jednotlivých tried – ten transformuje
- $x'_i = \boldsymbol{w}^T \boldsymbol{x}_i$ a $\bar{x}'_j = \boldsymbol{w}^T \overline{\boldsymbol{x}}_j$
- Potom variability premietnutých príznakov sú
- $Q = Q_M + Q_V$
- $Q = \sum_{i=1}^{N} (\boldsymbol{w}^T \boldsymbol{x}_i - \boldsymbol{w}^T \overline{\boldsymbol{x}})(\boldsymbol{w}^T \boldsymbol{x}_i - \boldsymbol{w}^T \overline{\boldsymbol{x}})^T = \boldsymbol{w}^T S \boldsymbol{w}$

LDA V

- $Q_M = \sum_{j=1}^{C} N_j\left(\boldsymbol{w}^T\bar{x}_j - \boldsymbol{w}^T\overline{\boldsymbol{x}}\right)\left(\boldsymbol{w}^T\bar{x}_j - \boldsymbol{w}^T\overline{\boldsymbol{x}}\right)^T = \boldsymbol{w}^T S_M \boldsymbol{w}$
- $Q_V = \sum_{j=1}^{C} \sum_{x\in\omega_j}\left(\boldsymbol{w}^T\boldsymbol{x} - \boldsymbol{w}^T\overline{\boldsymbol{x}}_j\right)\left(\boldsymbol{w}^T\boldsymbol{x} - \boldsymbol{w}^T\overline{\boldsymbol{x}}_j\right)^T = \boldsymbol{w}^T S_V \boldsymbol{w}$
- Potom definujeme Fischerovo kritérium J, ktoré je skalárom pre konkrétny vektor $\boldsymbol{w}$
- $J(\boldsymbol{w}) = \frac{\boldsymbol{w}^T S_M \boldsymbol{w}}{\boldsymbol{w}^T S_V \boldsymbol{w}}$
- Hľadáme $\boldsymbol{w}$, ktoré maximalizuje medzitriednu a minimalizuje vnútrotriednu variabilitu

LDA VI

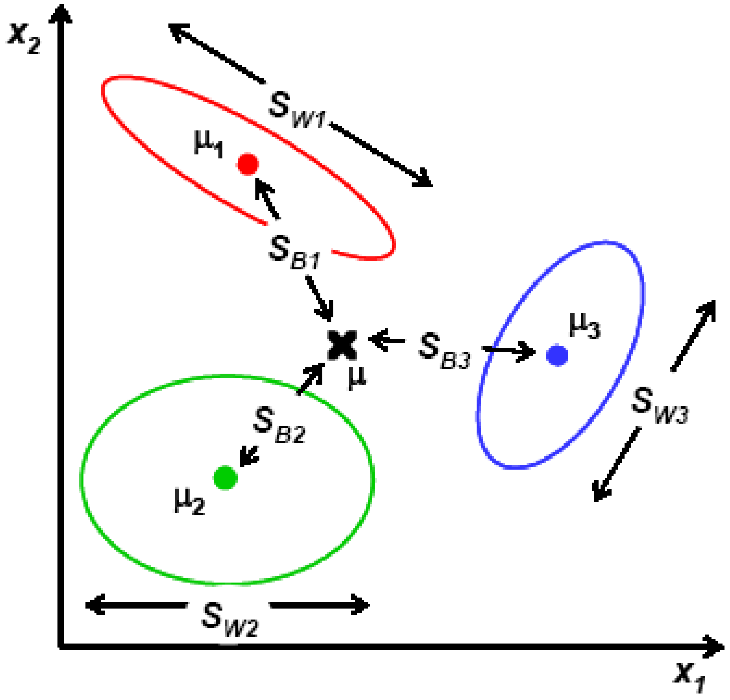


- Pre dve klasifikačné triedy:

$$\frac{\mathrm{d}J}{\mathrm{d}\mathbf{w}} = (\mathbf{w}^T \mathbf{S}_M \mathbf{w}) 2\mathbf{S}_V \mathbf{w} - (\mathbf{w}^T \mathbf{S}_V \mathbf{w}) 2\mathbf{S}_M \mathbf{w} \equiv 0$$

- $\boldsymbol{S}_M \boldsymbol{w} = J \boldsymbol{S}_V \boldsymbol{w}$
- $\boldsymbol{S}_V^{-1} \boldsymbol{S}_M \boldsymbol{w} = J \boldsymbol{w}$
- Ak $\boldsymbol{S}_V$ je regulárna, potom J je vlastné číslo matice $\boldsymbol{S}_V^{-1} \boldsymbol{S}_M$

LDA VII

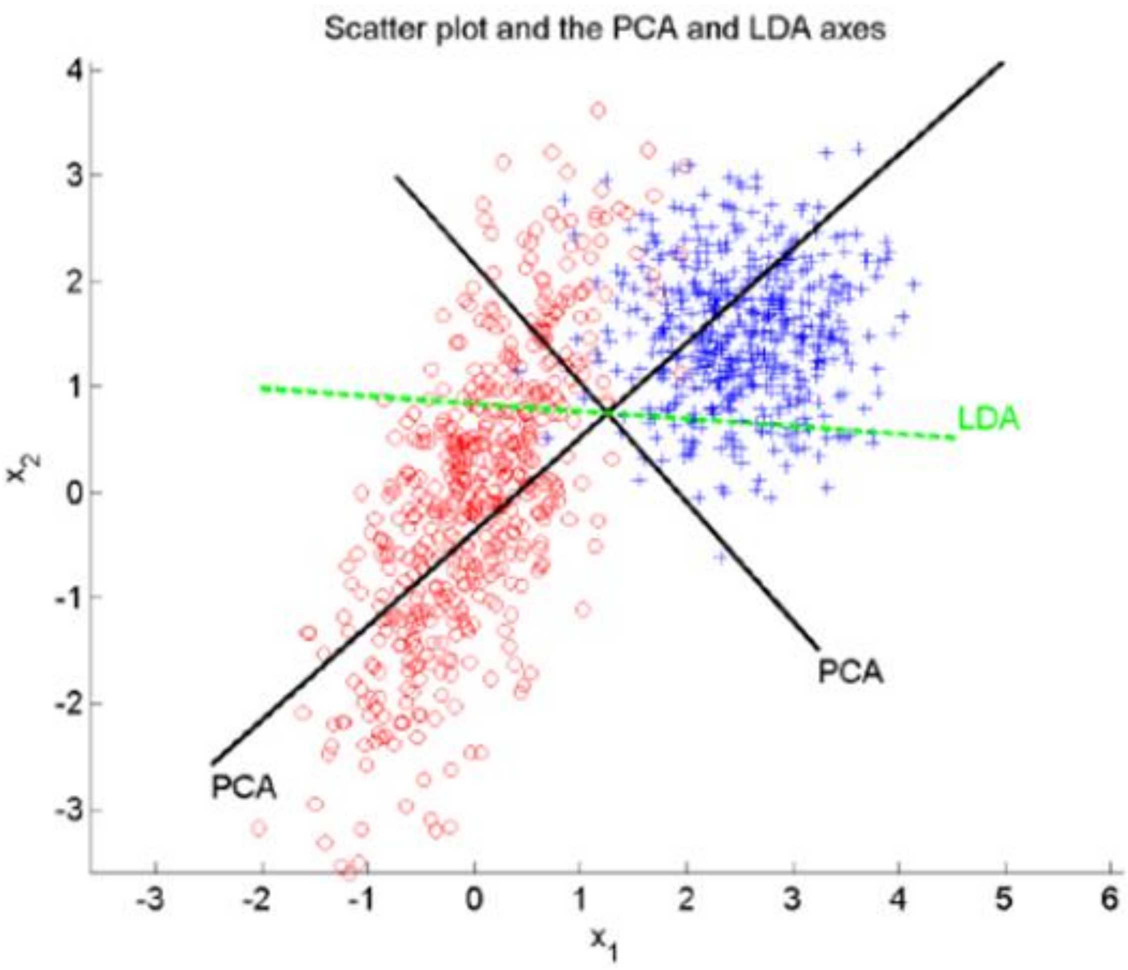


- Pre C klasifikačných tried:
- $\boldsymbol{W} = [\boldsymbol{w}_1 | \boldsymbol{w}_2 | \dots | \boldsymbol{w}_{C-1}]$
- $J(\boldsymbol{w}) = \frac{|\boldsymbol{W}^T S_M \boldsymbol{W}|}{|\boldsymbol{W}^T S_V \boldsymbol{W}|}$
- Riešením sú vektory $\boldsymbol{w}_j$, ktoré sú zovšeobecnenými vlastnými vektormi prislúchajúcimi vlastným číslam matíc $\boldsymbol{S}_M$ a $\boldsymbol{S}_V$ a teda spĺňajú $(\boldsymbol{S}_M - \boldsymbol{S}_V \lambda)\boldsymbol{w} = 0$ (dajú sa vypočítať ako vlastné čísla a vektory matice $\boldsymbol{S}_V^{-1} \boldsymbol{S}_M$)

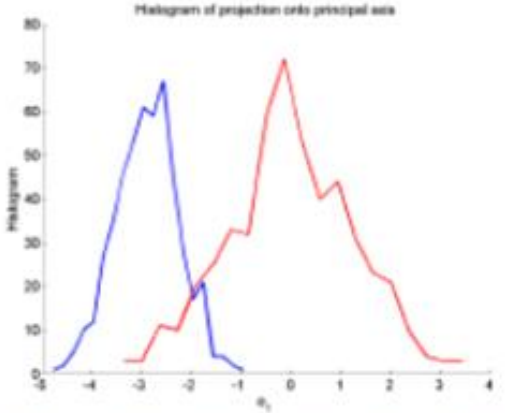
LDA VIII

- Vlastné čísla matice $\boldsymbol{S}_V^{-1}\boldsymbol{S}_M$ sa počítajú zložito, lebo matica je nesymetrická
- Preto maticu $\boldsymbol{S}_V$, ktorá je štvorcová a symetrická rozložíme pomocou SVD na $\boldsymbol{S}_V = \boldsymbol{U}\boldsymbol{\Phi}\boldsymbol{U}^T$ a definujeme $\boldsymbol{S}_V^{1/2} = \boldsymbol{U}\boldsymbol{\Phi}^{1/2}\boldsymbol{U}^T$ a tiež na $\boldsymbol{S}_V^{-1/2} = \boldsymbol{U}\boldsymbol{\Phi}^{-1/2}\boldsymbol{U}^T$, kde platí $\boldsymbol{S}_V = \boldsymbol{S}_V^{1/2}\boldsymbol{S}_V^{1/2}$ a $\boldsymbol{S}_V^{-1} = \boldsymbol{S}_V^{-1/2}\boldsymbol{S}_V^{-1/2}$
- Získame symetrickú maticu $\boldsymbol{S}_V^{-1/2}\boldsymbol{S}_M\boldsymbol{S}_V^{-1/2}$

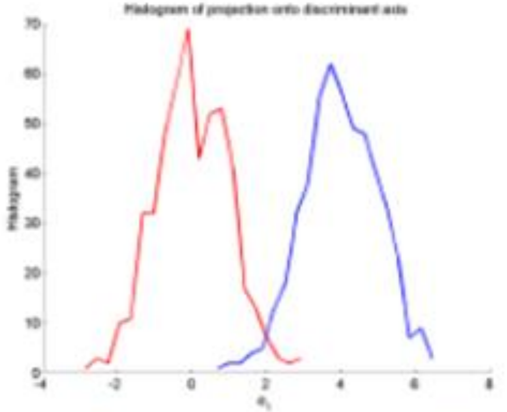
PCA vs. LDA



(a) Scatter plot.

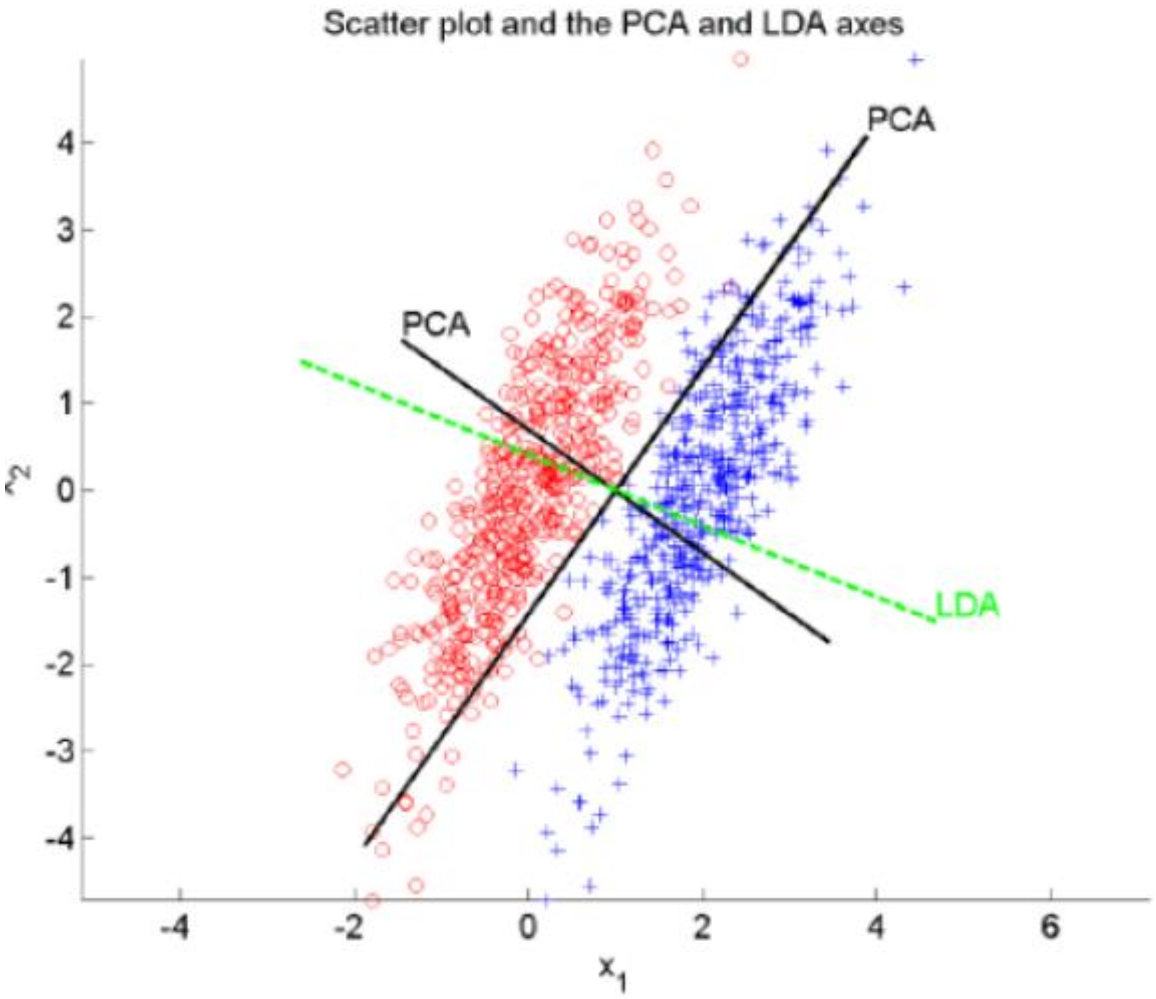


(b) Projection onto the first PCA axis.

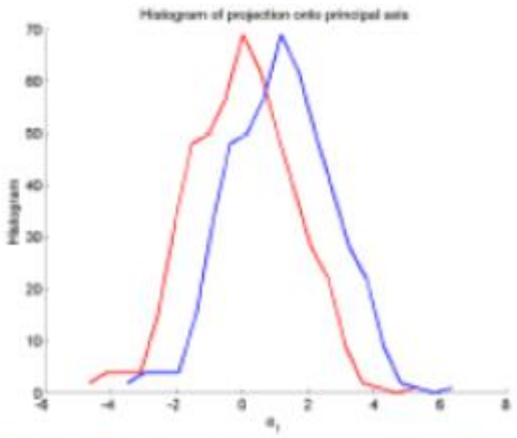


(c) Projection onto the first LDA axis.

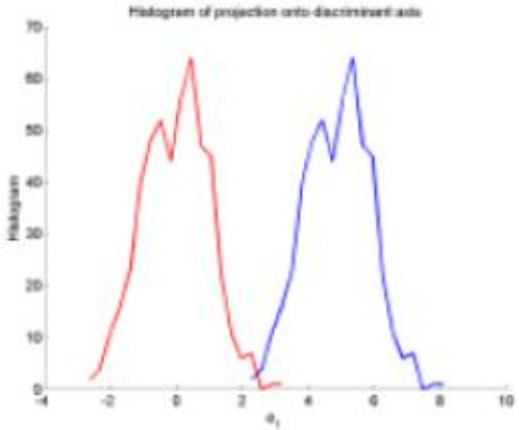
PCA vs. LDA II



(a) Scatter plot.



(b) Projection onto the first PCA axis.



(c) Projection onto the first LDA axis.

PCA vs. LDA III

- Niekedy použijeme obe metódy:
- Využijeme PCA na zníženie dimenzie

$$\begin{bmatrix} x_1 \\ \dots \\ x_D \end{bmatrix} \longrightarrow PCA \longrightarrow \begin{bmatrix} y_1 \\ \dots \\ y_K \end{bmatrix}$$

- Aplikujeme LDA, aby sme našli diskriminatívne smery

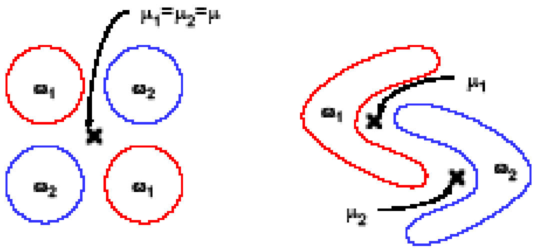
$$\begin{bmatrix} y_1 \\ \dots \\ y_K \end{bmatrix} \longrightarrow LDA \longrightarrow \begin{bmatrix} z_1 \\ \dots \\ z_{C-1} \end{bmatrix}$$

PCA vs. LDA IV

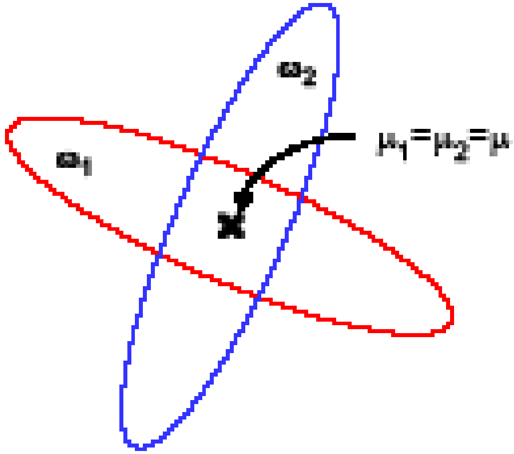
- Všeobecné hodnotenie oboch metód:
- 1) Pri malých trénovacích množinách dáva PCA lepšie výsledky ako LDA
- 2) Keď máme dostatočný počet trénovacích dát pre každú triedu, LDA dáva lepšie výsledky ako PCA

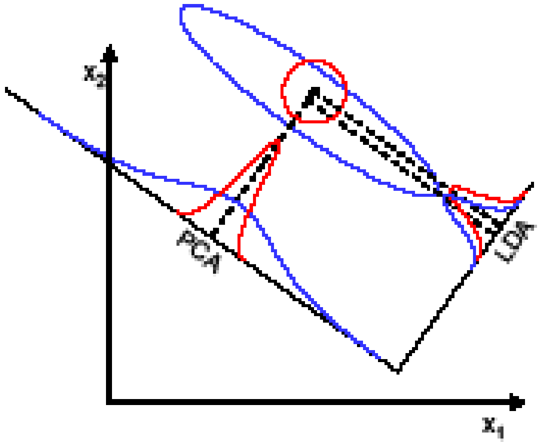
PCA vs. LDA V

- LDA predpokladá normálne rozdelenie dát

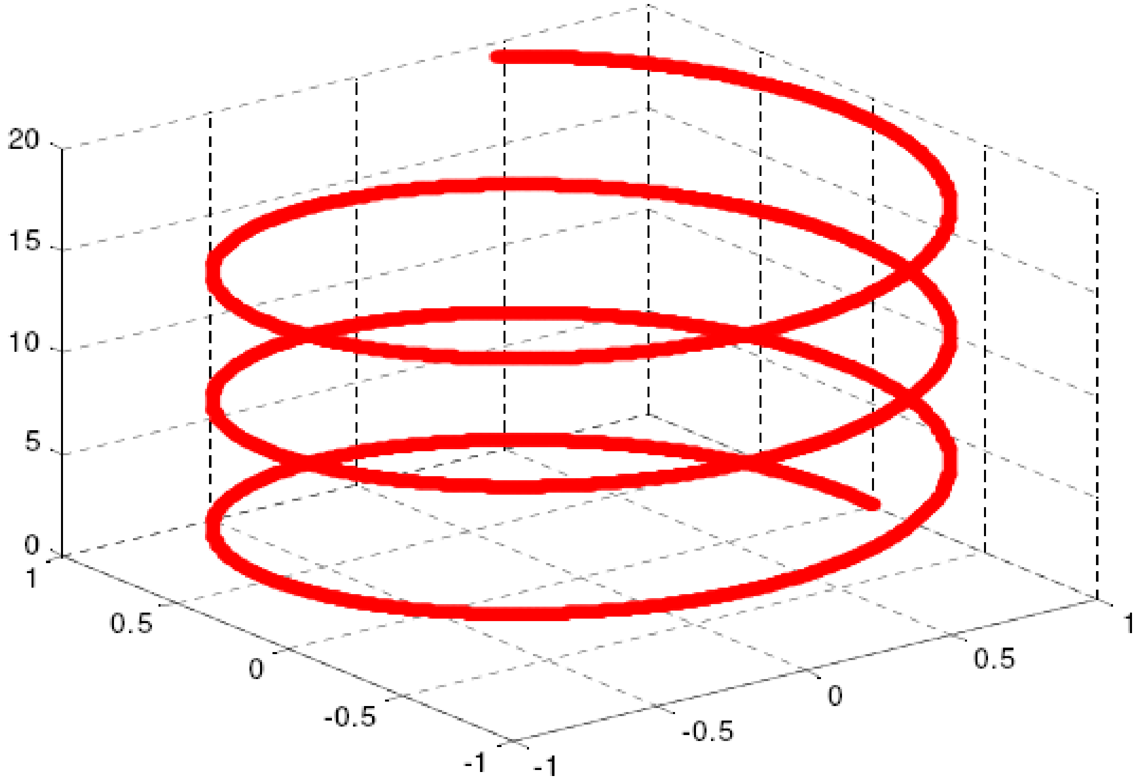


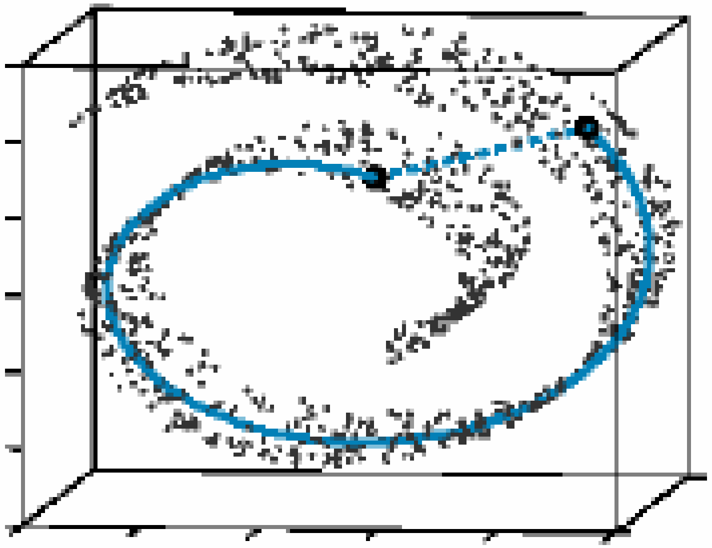
- Ak je rozdiel medzi triedami najmä vo variancii a nie v priemere, LDA tieto dáta neseparuje správne

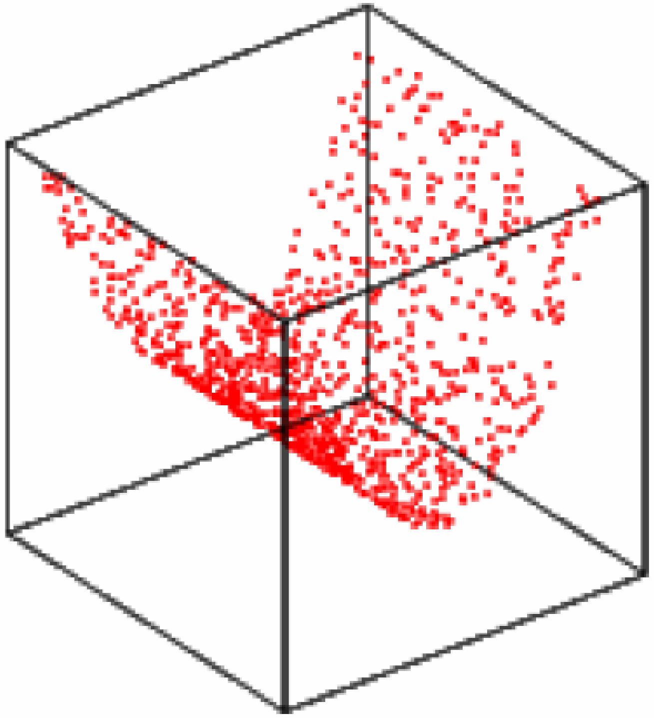




Nelineárne metódy







Rozpoznávanie