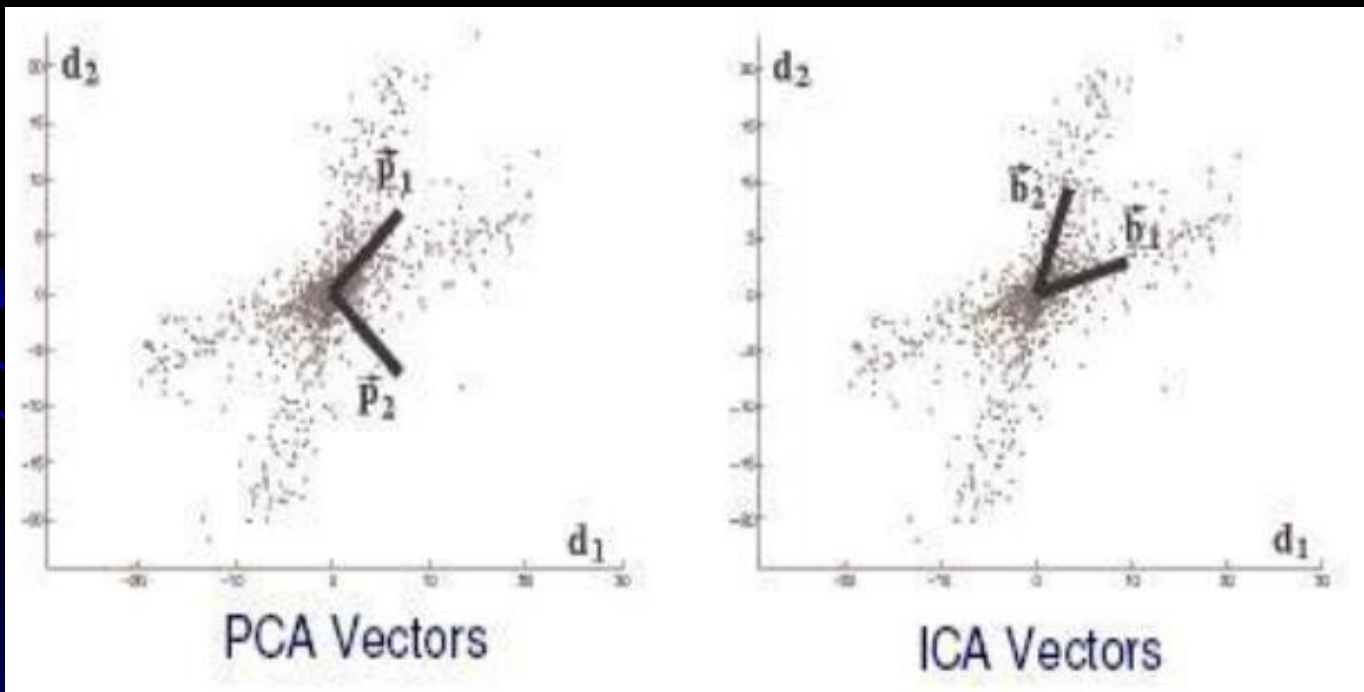


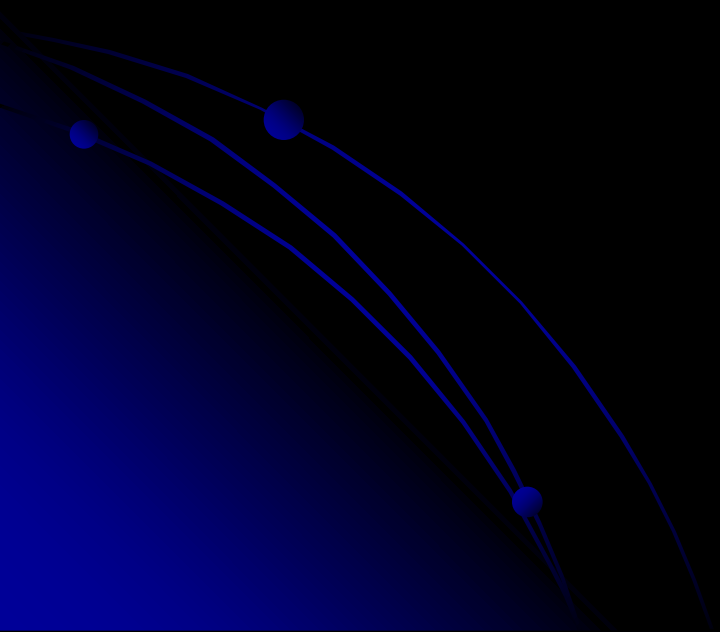
ICA (Independent Components Analysis)

Metóda nezávislých komponentov

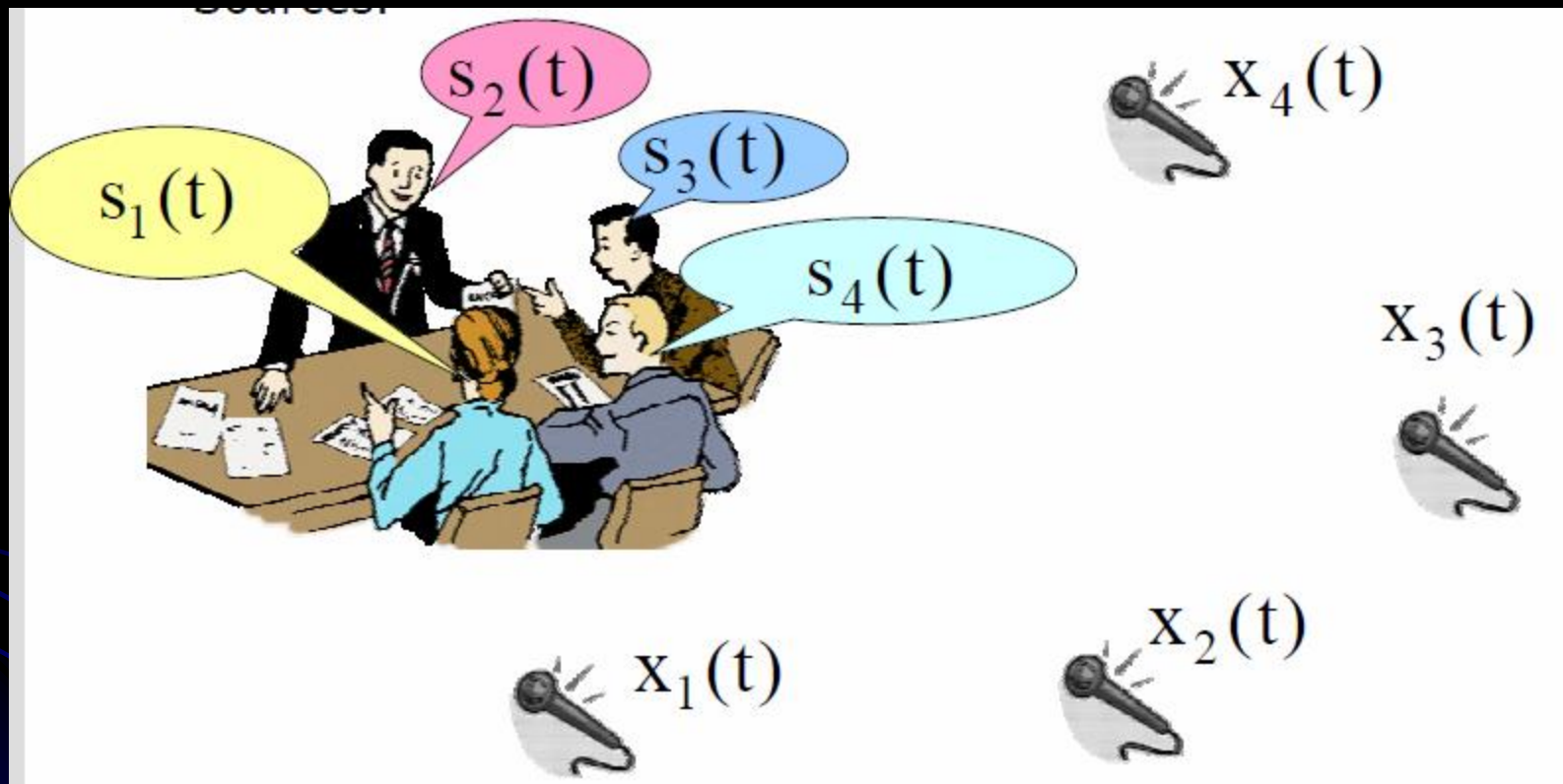
Nevyžaduje ortogonalitu, len linearitu



ICA - reprezentuje viacrozmerný vektor ako lineárnu kombináciu negaussovských náhodných premenných („nezávislé komponenty“), ktoré sú čo najnezávislejšie

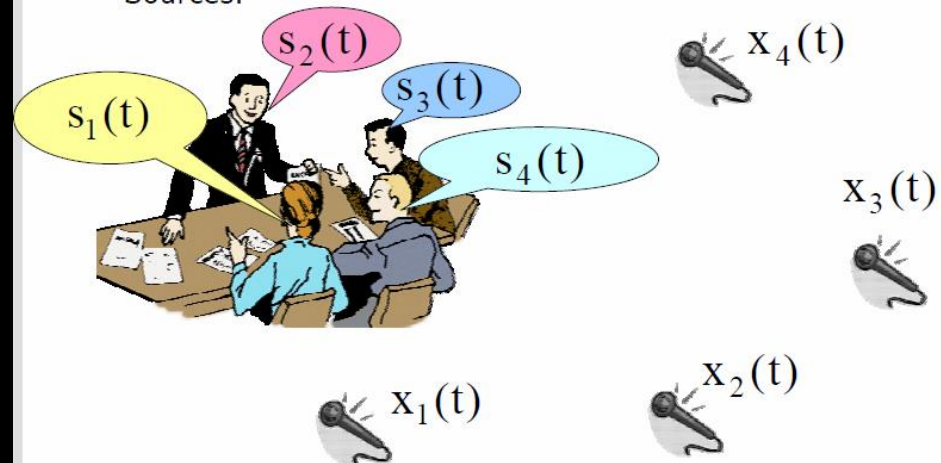


Cocktail party problem



Platí

$$\mathbf{x} = \mathbf{A}\mathbf{s} = \sum_{i=1}^n \mathbf{a}_i s_i$$

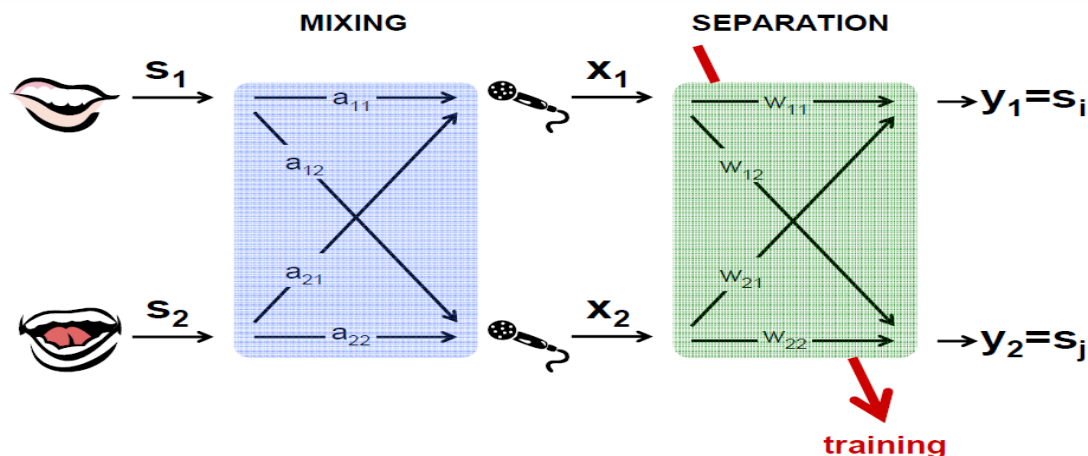


$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) + a_{13}s_3(t) + a_{14}s_4(t)$$

$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) + a_{23}s_3(t) + a_{24}s_4(t)$$

$$x_3(t) = a_{31}s_1(t) + a_{32}s_2(t) + a_{33}s_3(t) + a_{34}s_4(t)$$

$$x_4(t) = a_{41}s_1(t) + a_{42}s_2(t) + a_{43}s_3(t) + a_{44}s_4(t)$$



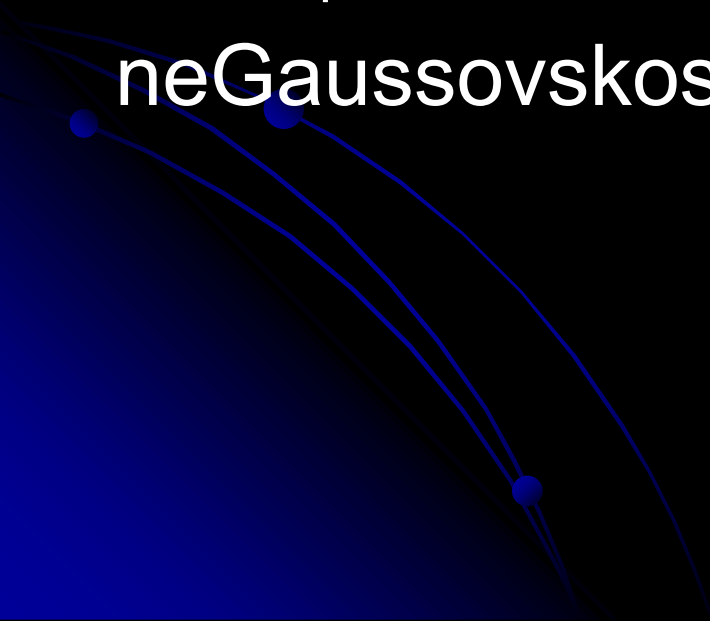
Predpoklad:

$$p(s_1, s_2, \dots, s_n) = p(s_1)p(s_2) \dots p(s_n)$$

$$E(s_i) = 0$$

$$\text{Var}(s_i) = 1$$

neGaussovskost'



ICA postup

Predspracovanie

Centrovanie $\mathbf{x}' = \mathbf{x} - \bar{\mathbf{x}}$

Bielenie $\tilde{\mathbf{x}} = \mathbf{B}\mathbf{x}'$, aby $\Sigma_{\tilde{\mathbf{x}}} = \mathbf{I}$

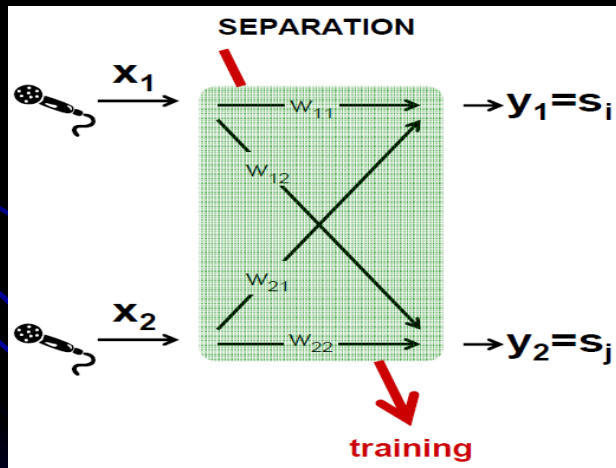
Postup $\tilde{\mathbf{x}} = \mathbf{V}\mathbf{D}\mathbf{V}^T \mathbf{x}'$

kde $\Sigma = \mathbf{V}\mathbf{S}\mathbf{V}^T$

$$d_{ij} = s_{ij}^{-1/2}$$

ICA postup

Hľadáme smery w_i , aby sme maximalizovali neGaussovskosť

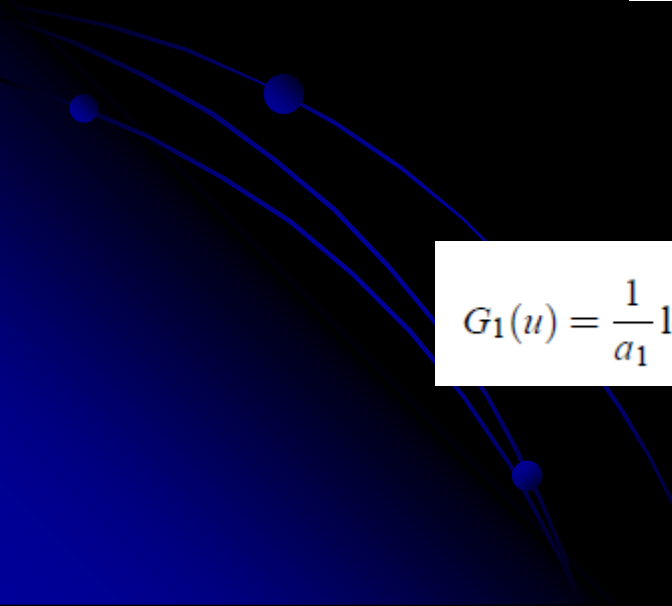


Miera neGaussovskosti:

Šikmost', špicatost' (3.,4. moment)

Negentropia $J(y) = H(y_G) - H(y)$

aproximácia $J(y) \propto (E(G(y)) - E(G(y_G)))^2$


$$G_1(u) = \frac{1}{a_1} \log \cosh a_1 u, \quad G_2(u) = -\exp(-u^2/2)$$

Hľadáme w , aby maximalizoval
neGaussovskosť

$$J(\mathbf{w}^T \mathbf{x}) \propto (E(G(\mathbf{w}^T \mathbf{x})) - E(G(\mathbf{y}_G)))^2$$

Za podmienky $\|\mathbf{w}\|^2 = 1$

Konštanta pre w

Lagrange: $L = E(G(\mathbf{w}^T \mathbf{x})) - \lambda(\mathbf{w}^T \mathbf{w} - 1)$

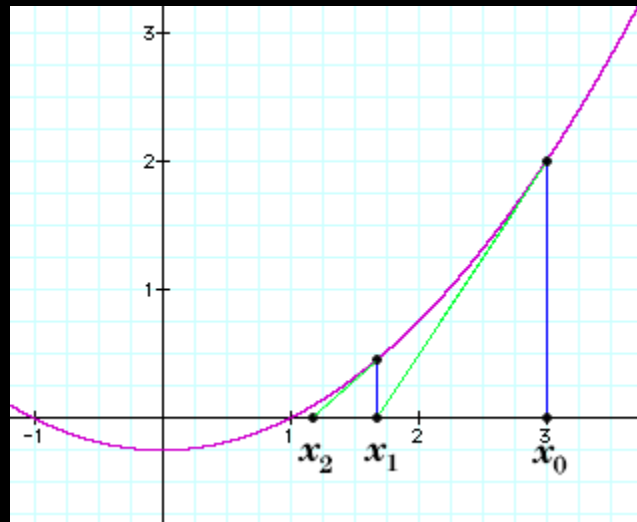
Derivácia: $L' = E(\mathbf{x} \cdot g(\mathbf{w}^T \mathbf{x})) - \lambda \mathbf{w} \equiv 0$

Riešenie použitím Newtonovej metódy

$f(\mathbf{w})$

Newtonova metóda hľadania koreňov

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



FastICA algoritmus pre 1 smer:

1. Zvoľ náhodný počiatočný vektor w

2.
$$w^+ = E\{xg(w^T x)\} - E\{g'(w^T x)\}w$$

3.
$$w = w^+ / \|w^+\|$$

4. Opakuj 2.,3. kým nenastane konvergencia

Pre viacero smerov:

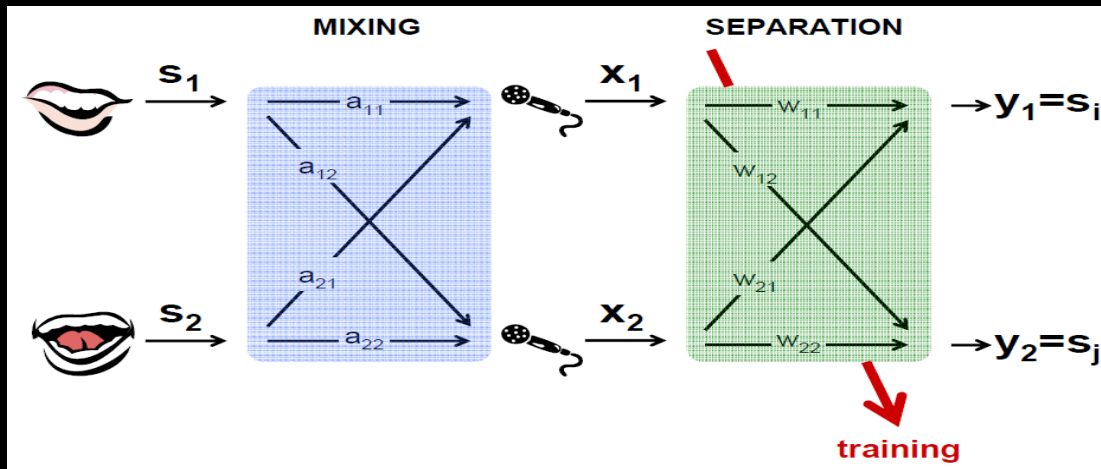
FastICA postupne pre každý smer, ale pri každej iterácii de Korelujeme ten nový smer:

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

FastICA všetky smery, potom symetrická de Korelácia:

$$\mathbf{W} = \mathbf{W} / \sqrt{\|\mathbf{W}\mathbf{W}^T\|}, \text{ iteruj } \mathbf{W} = \frac{3}{2}\mathbf{W} - \frac{1}{2}\mathbf{W}\mathbf{W}^T\mathbf{W}$$

Redukcia



$$X = AS$$

$$S = WX$$

$$[p \times n] = [p \times q][q \times n]$$

$$q \leq p$$

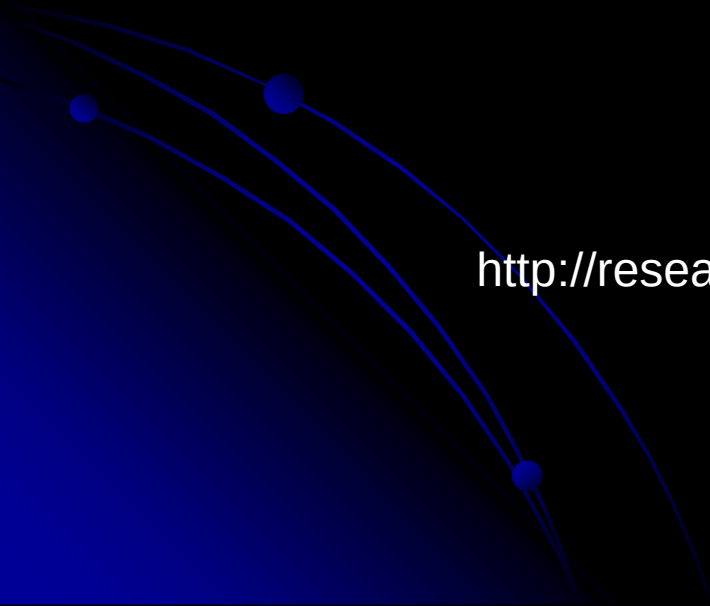
$$A = W^{-1}$$

Možnosti redukcie

1. Vypočítame ICA pre $q < p$
2. Pri bielení zachováme q PC
3. Vypočítame ICA pre $q = p$,
na základe matice A učíme q IC

Independent Component Analysis:
Algorithms and Applications
Aapo Hyvärinen and Erkki Oja
Neural Networks, 13(4-5):411-430, 2000

<http://research.ics.aalto.fi/ica/icademo/>



Independent Component Analysis of Textures in Angiography Images

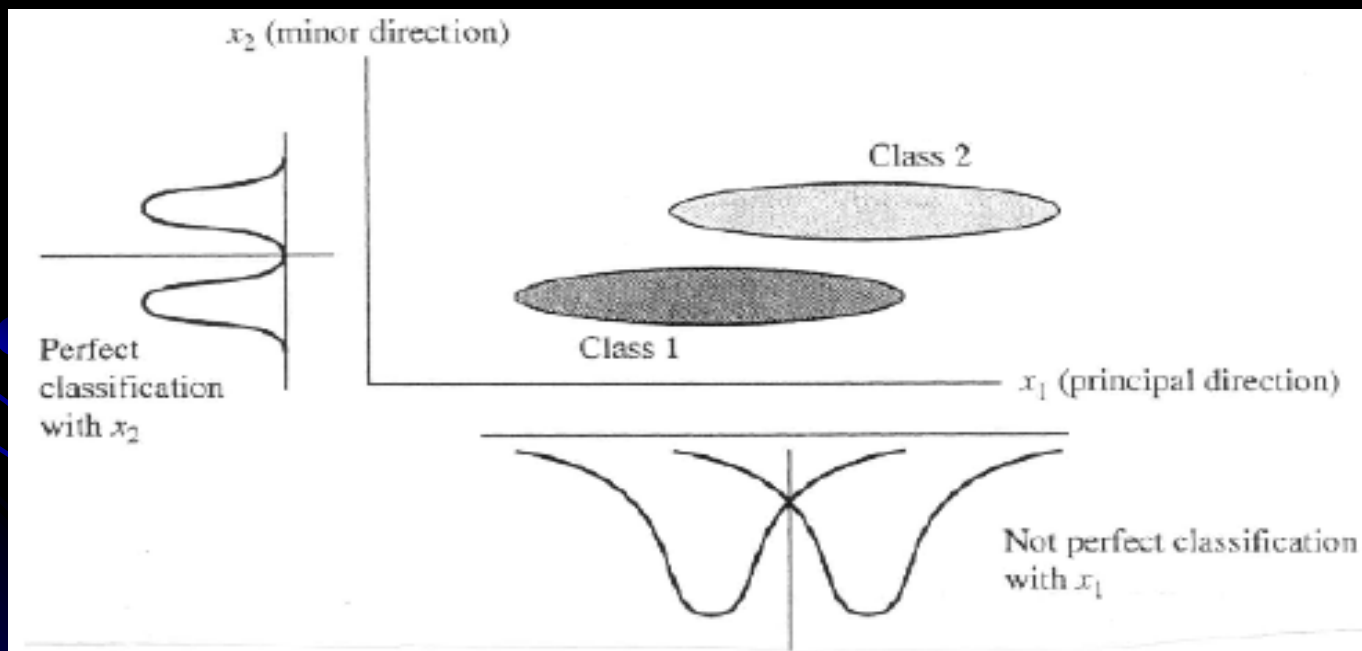
<http://www.ia.pw.edu.pl/~wkasprza/PAP/ICCVG04c.pdf>

Independent Component Analysis of Textures

<https://pdfs.semanticscholar.org/a734/1ead68514c5eb39cc2e907df62fd280e7f87.pdf>

PCA – nie vždy optimálne pre klasifikáciu

PCA – neberie do úvahy príslušnosť dát k jednotlivým triedam



Linear Discriminant Analysis (LDA)

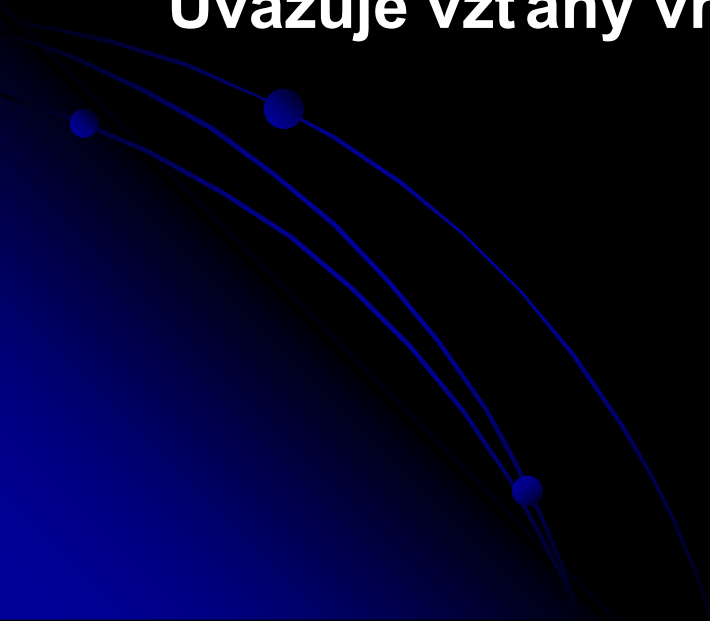
Lineárna diskriminačná analýza

Na čo slúži LDA?

Zníženie dimenzie, zachovanie separability tried

Smer, pri ktorom sú triedy najlepšie oddelené

Uvažuje vzťahy vnútri tried aj medzi triedami



Fisherova lineárna disriminačná analýza

riadená metóda

využíva informáciu o klasifikačných triedach

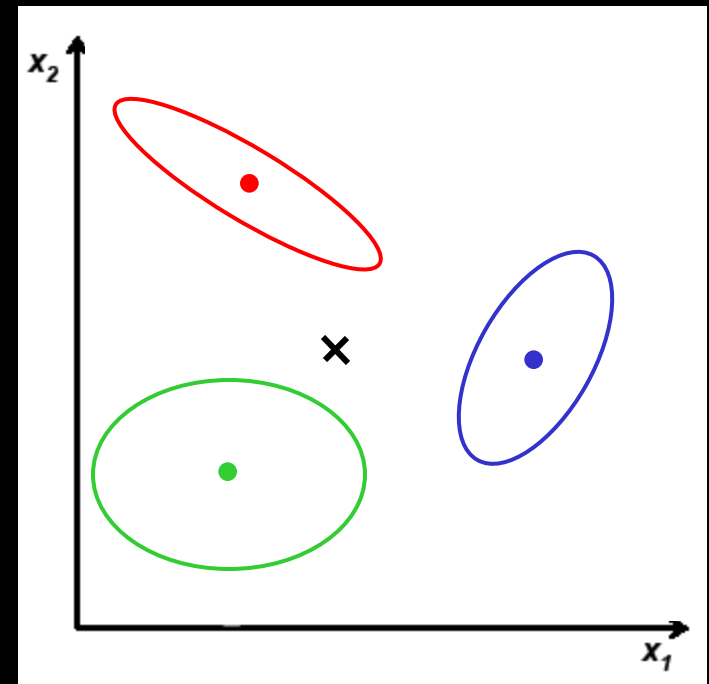
N D -rozmerných vektorov príznačkov $\mathbf{x}_1, \dots, \mathbf{x}_N$

$\{\omega_j\}_{j=1}^C$

triede ω_j patrí N_j príznačkov.

$$\bar{\mathbf{x}}_j = \frac{1}{|\omega_j|} \sum_{\mathbf{x} \in \omega_j} \mathbf{x} = \frac{1}{N_j} \sum_{\mathbf{x} \in \omega_j} \mathbf{x}$$

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i$$



Celková variabilita v dátach je

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

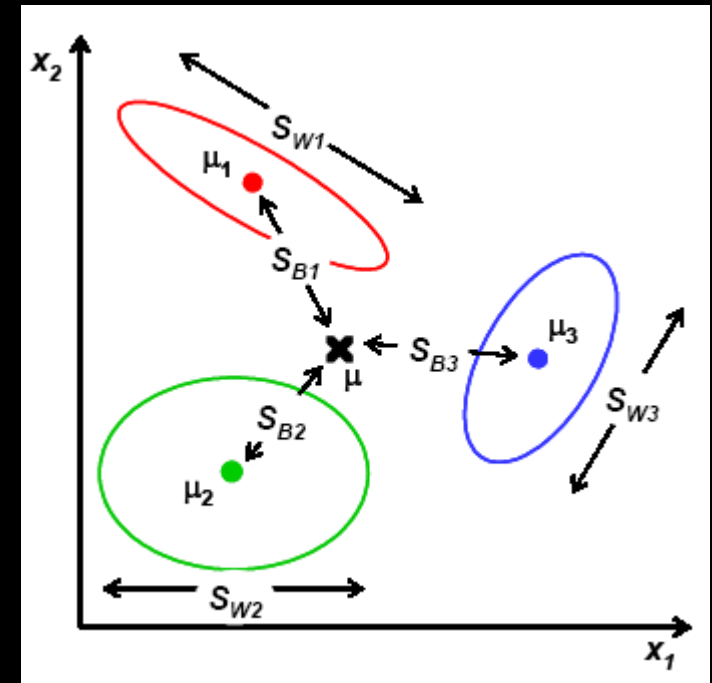
$$S = S_M + S_V$$

$$S_M = \sum_{j=1}^C N_j (\bar{\mathbf{x}}_j - \bar{\mathbf{x}})(\bar{\mathbf{x}}_j - \bar{\mathbf{x}})^T$$

$$S_V = \sum_{j=1}^C \sum_{\mathbf{x} \in \omega_j} (\mathbf{x} - \bar{\mathbf{x}}_j)(\mathbf{x} - \bar{\mathbf{x}}_j)^T = \sum_{j=1}^C S_j$$

Variabilita v jednotlivých triedach je

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



smer $\mathbf{w}$, ktorý by nám pomohol klasifikovať príznaky do jednotlivých tried.

$$x'_i = \mathbf{w}^T \mathbf{x}_i$$

$$\bar{x}'_j = \mathbf{w}^T \bar{\mathbf{x}}_j$$

Variability premietnutých príznakov

$$Q = Q_M + Q_V$$

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

$$Q_M = \sum_{j=1}^C N_j (\mathbf{w}^T \bar{\mathbf{x}}_j - \mathbf{w}^T \bar{\mathbf{x}}) (\mathbf{w}^T \bar{\mathbf{x}}_j - \mathbf{w}^T \bar{\mathbf{x}})^T = \mathbf{w}^T \mathbf{S}_M \mathbf{w}$$

$$Q_V = \sum_{j=1}^C \sum_{\mathbf{x} \in \omega_j} (\mathbf{w}^T \mathbf{x} - \mathbf{w}^T \bar{\mathbf{x}}_j) (\mathbf{w}^T \mathbf{x} - \mathbf{w}^T \bar{\mathbf{x}}_j)^T = \mathbf{w}^T \mathbf{S}_V \mathbf{w}.$$

$$\mathbf{S} = \mathbf{S}_M + \mathbf{S}_V$$

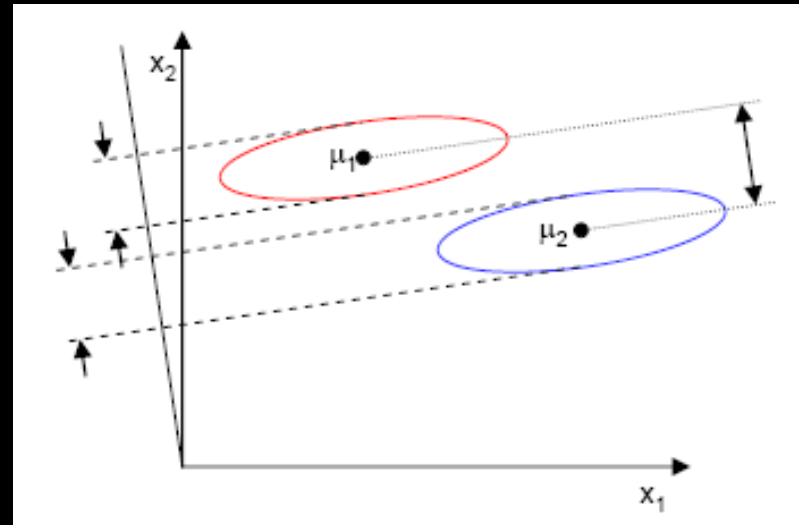
$$\mathbf{S}_M = \sum_{j=1}^C N_j (\bar{\mathbf{x}}_j - \bar{\mathbf{x}}) (\bar{\mathbf{x}}_j - \bar{\mathbf{x}})^T$$

$$\mathbf{S}_V = \sum_{j=1}^C \sum_{\mathbf{x} \in \omega_j} (\mathbf{x} - \bar{\mathbf{x}}_j) (\mathbf{x} - \bar{\mathbf{x}}_j)^T = \sum_{j=1}^C \mathbf{S}_j$$

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_M \mathbf{w}}{\mathbf{w}^T \mathbf{S}_V \mathbf{w}}$$

2 klasifikačné triedy

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_M \mathbf{w}}{\mathbf{w}^T \mathbf{S}_V \mathbf{w}}$$



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

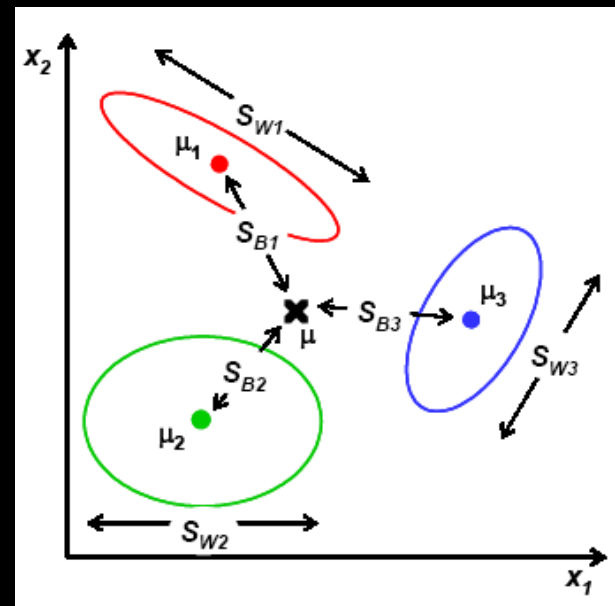
$$\mathbf{S}_V^{-1} \mathbf{S}_M \mathbf{w} = J \mathbf{w}$$

$$\mathbf{S}_M \mathbf{w} = J \mathbf{S}_V \mathbf{w}$$

C klasifikačných tried

$$\mathbf{W} = [\mathbf{w}_1 | \mathbf{w}_2 | \dots | \mathbf{w}_{C-1}]$$

$$J(\mathbf{w}) = \frac{|\mathbf{W}^T \mathbf{S}_M \mathbf{W}|}{|\mathbf{W}^T \mathbf{S}_V \mathbf{W}|}$$



Riešením FDLA sú vektory $\mathbf{w}_j$, ktoré sú zovšeobecnenými¹ vlastnými vektormi prislúchajúcimi zovšeobeným vlastným číslam matíc $\mathbf{S}_M$ a $\mathbf{S}_V$, a teda spĺňajú

$$(\mathbf{S}_M - \mathbf{S}_V \lambda) \mathbf{w} = 0.$$

Maticu $\mathbf{S}_V$, ktorá je štvorcová a symetrická rozložíme na $\mathbf{S}_V = \mathbf{U}\Phi\mathbf{U}^T$, kde $\Phi = \text{diag}(\{\varphi_i\}_{i=1}^D)$ má na diagonále vlastné čísla a $\mathbf{U}$ sú ortonormálne vlastné vektory matice $\mathbf{S}_V$. Zadefinujeme matice

$$\mathbf{S}_V^{1/2} = \mathbf{U}\Phi^{1/2}\mathbf{U}^T, \text{ kde } \Phi^{1/2} = \text{diag}\left(\{\varphi_i^{1/2}\}_{i=1}^D\right), \quad (12.40)$$

$$\mathbf{S}_V^{-1/2} = \mathbf{U}\Phi^{-1/2}\mathbf{U}^T, \text{ kde } \Phi^{-1/2} = \text{diag}\left(\{\varphi_i^{-1/2}\}_{i=1}^D\right), \quad (12.41)$$

pre ktoré platí $\mathbf{S}_V = \mathbf{S}_V^{1/2}\mathbf{S}_V^{1/2}$, $\mathbf{S}_V^{-1} = \mathbf{S}_V^{-1/2}\mathbf{S}_V^{-1/2}$ a $\mathbf{S}_V^{-1/2}\mathbf{S}_V^{1/2} = \mathbf{I}$. Preto

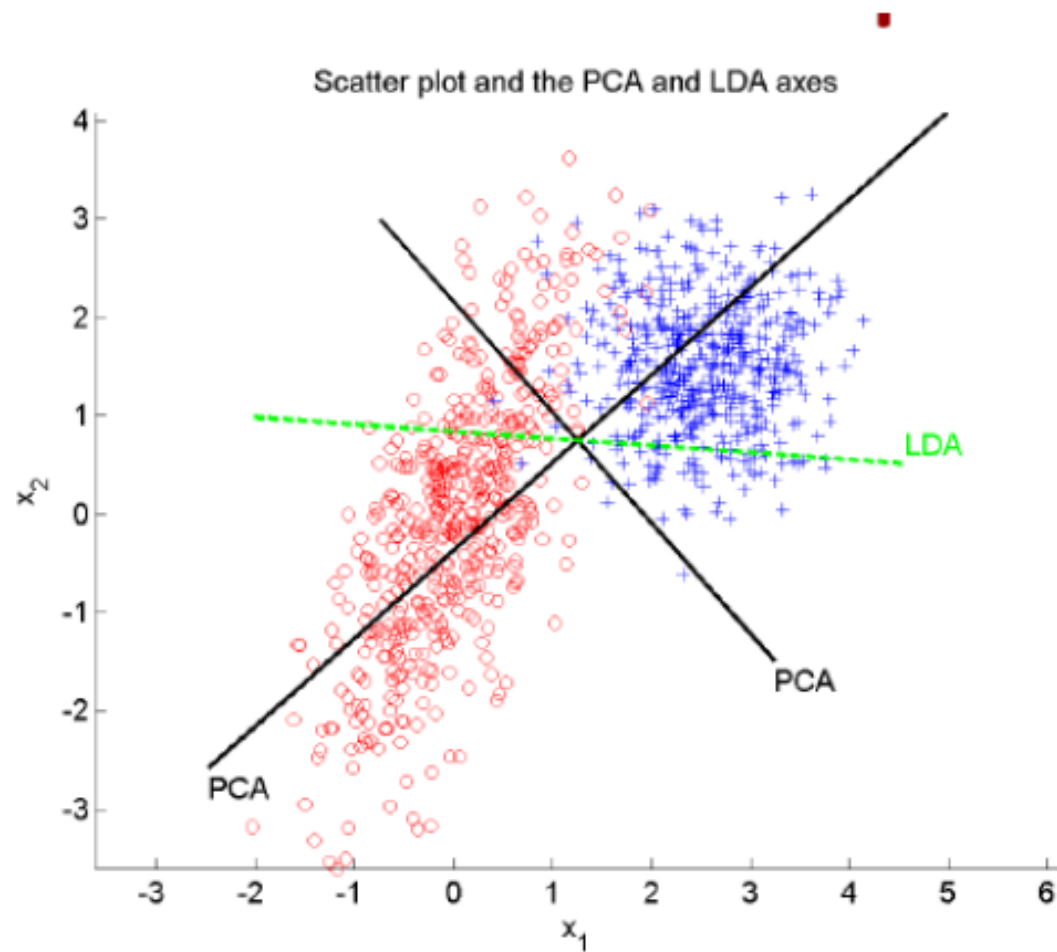
$$\mathbf{S}_M\mathbf{w} - \mathbf{S}_V\lambda\mathbf{w} = 0, \quad (12.42)$$

$$\mathbf{S}_M(\mathbf{S}_V^{-1/2}\mathbf{S}_V^{1/2})\mathbf{w} - \lambda\mathbf{S}_V(\mathbf{S}_V^{-1/2}\mathbf{S}_V^{1/2})\mathbf{w} = 0 \quad | \quad \mathbf{S}_V^{-1/2}, \quad (12.43)$$

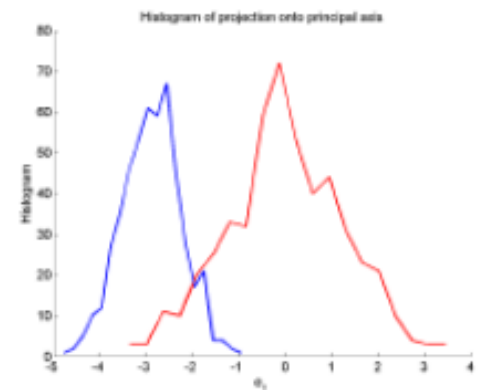
$$\mathbf{S}_V^{-1/2}\mathbf{S}_M\mathbf{S}_V^{-1/2}\mathbf{c} - \lambda\mathbf{c} = 0, \quad (12.44)$$

kde $\mathbf{c} = \mathbf{S}_V^{1/2}\mathbf{w}$. Matica $\mathbf{S}_V^{-1/2}\mathbf{S}_M\mathbf{S}_V^{-1/2}$ je symetrická, preto nájdenie vlastných čísel je jednoduchší problém.

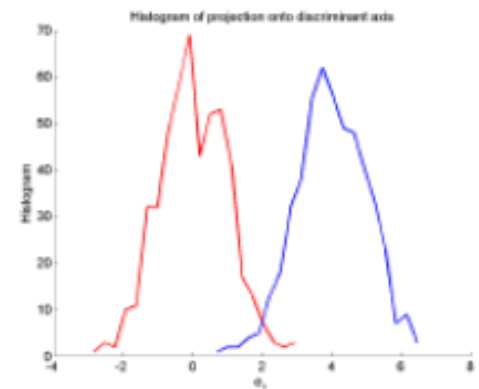
PCA vs. LDA



(a) Scatter plot.



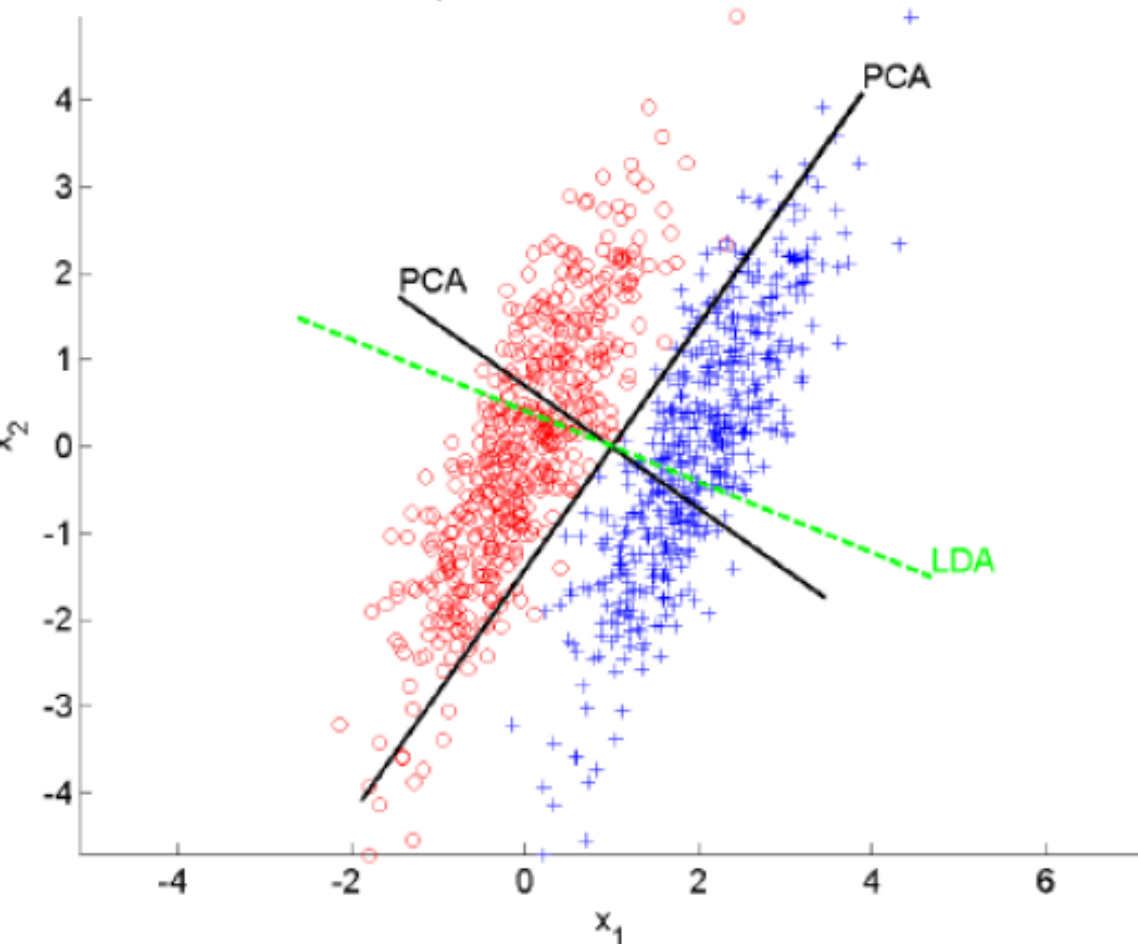
(b) Projection onto the first PCA axis.



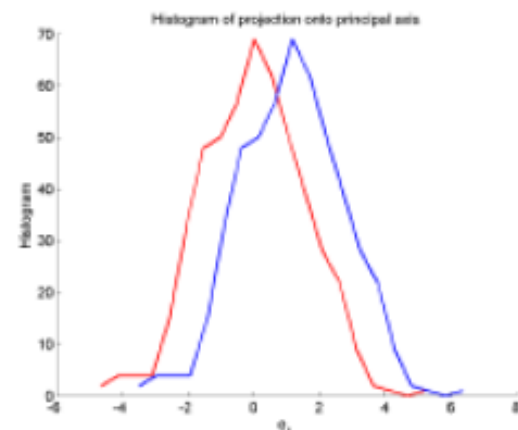
(c) Projection onto the first LDA axis.

PCA vs. LDA

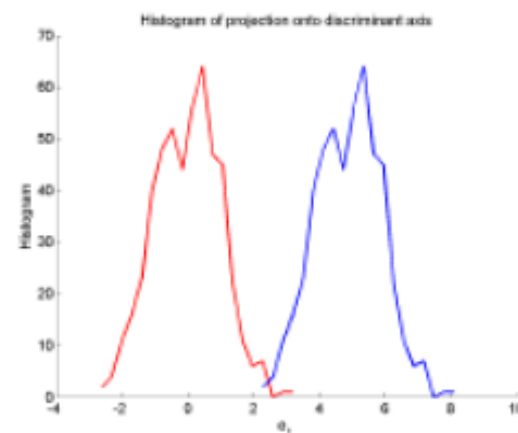
Scatter plot and the PCA and LDA axes



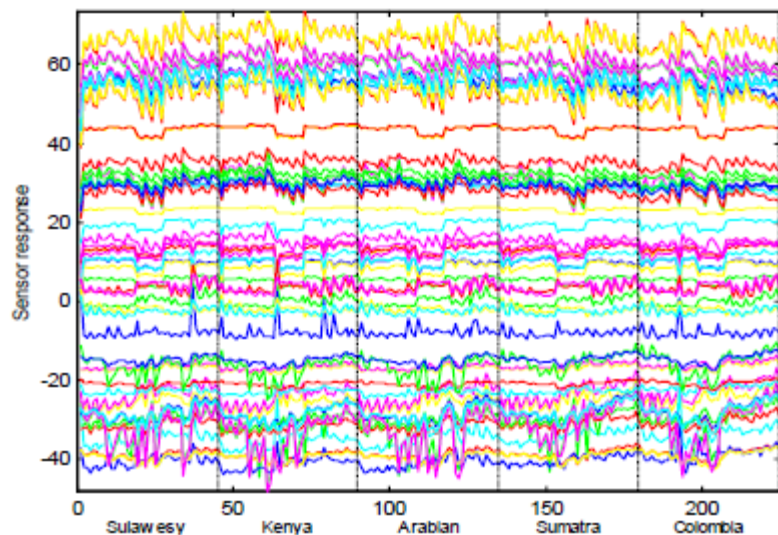
(a) Scatter plot.



(b) Projection onto the first PCA axis.

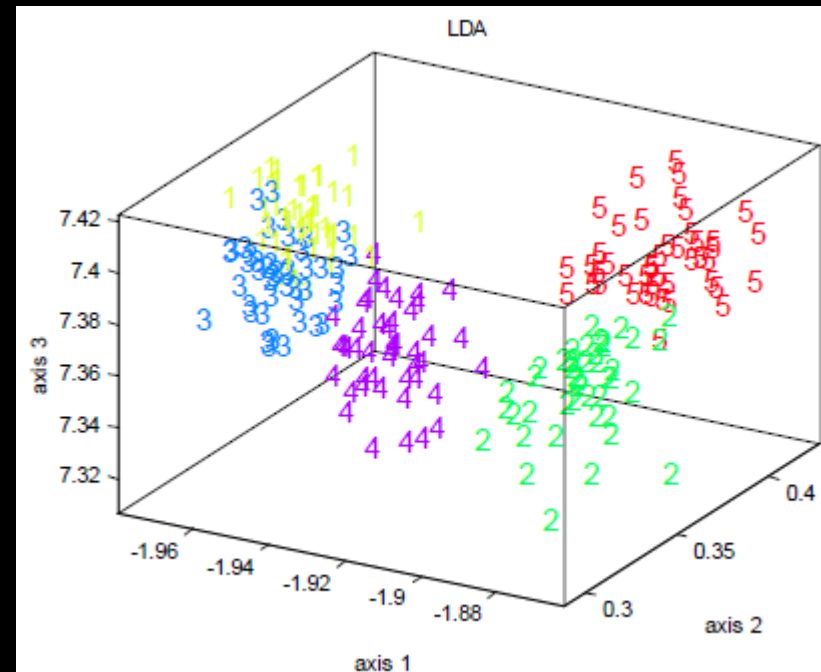
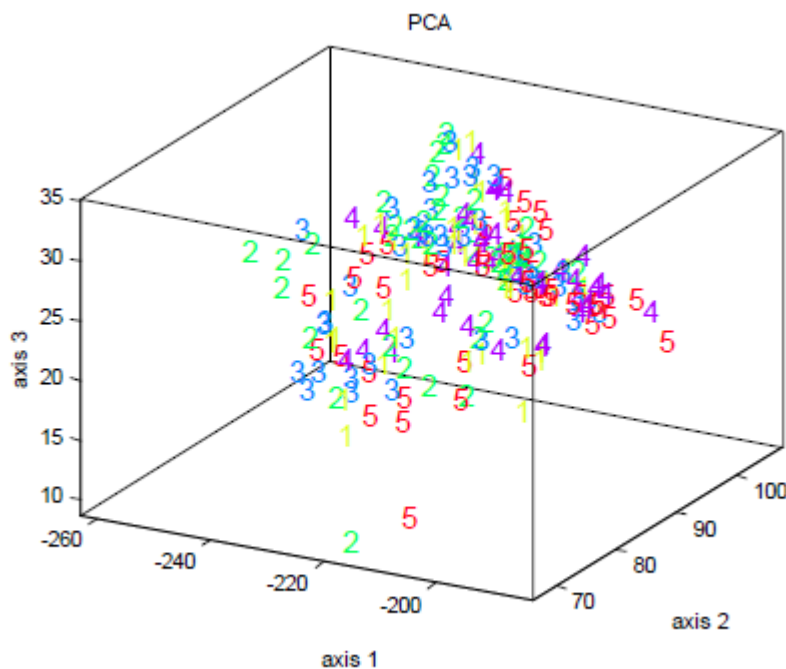


(c) Projection onto the first LDA axis.



Five types of coffee beans were presented to an array of gas sensors

For each coffee type, 45 "sniffs" were performed and the response of the gas sensor array was processed in order to obtain a 60-dimensional feature vector

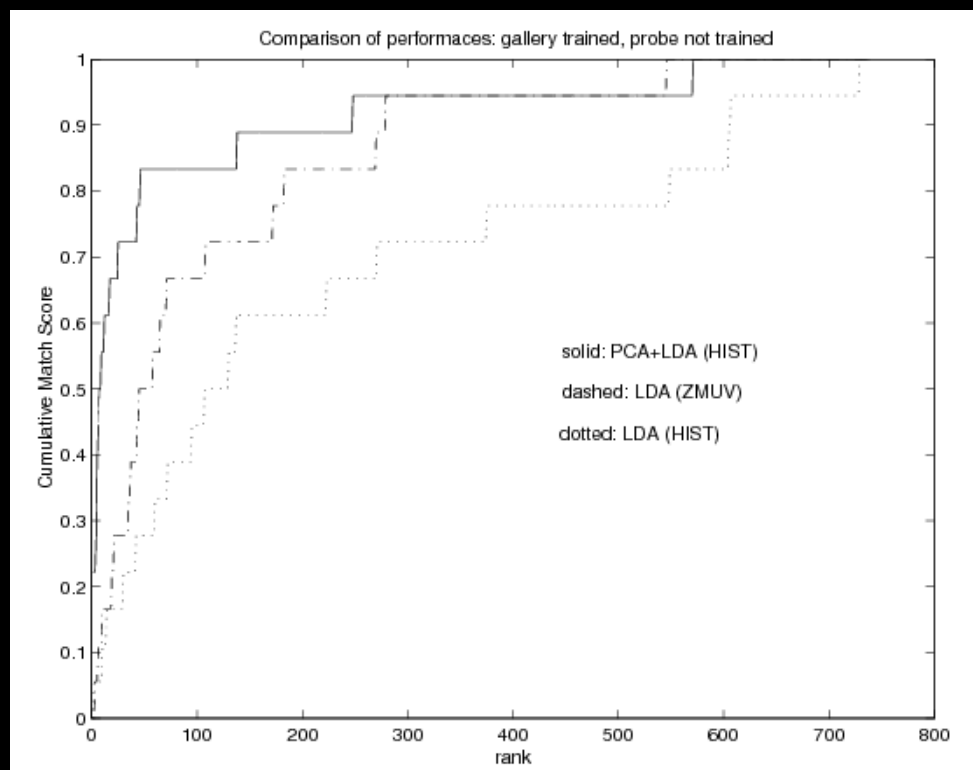


Využijeme PCA na zníženie dimenzie

$$\begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_N \end{bmatrix} \xrightarrow{\text{PCA}} \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_K \end{bmatrix}$$

Aplikujeme LDA aby sme našli diskriminatívne smery

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_K \end{bmatrix} \xrightarrow{\text{LDA}} \begin{bmatrix} z_1 \\ z_2 \\ \dots \\ z_{C-1} \end{bmatrix}$$

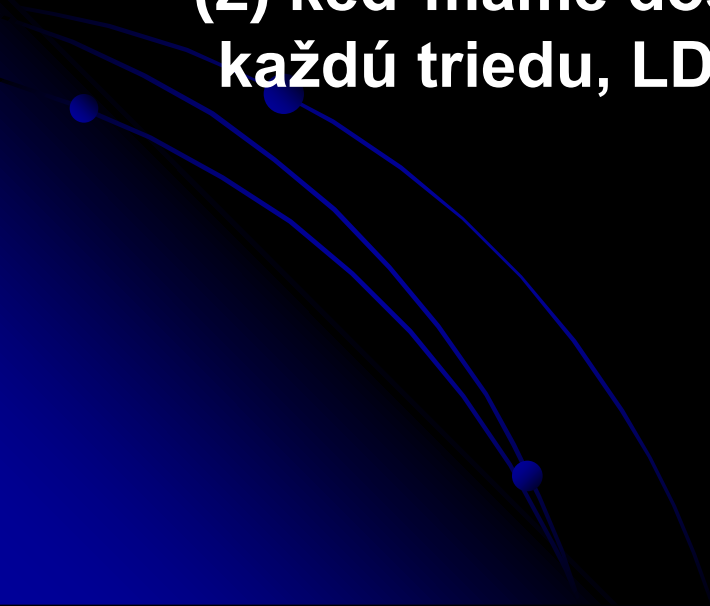


PCA versus LDA

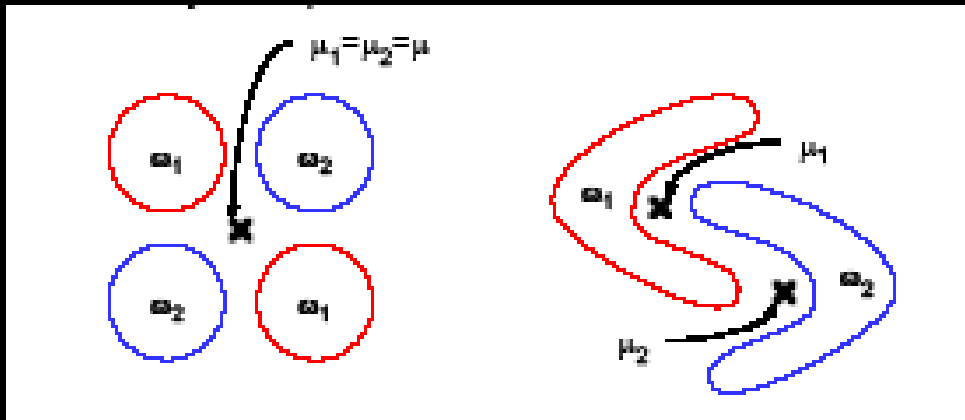
A. Martinez, A. Kak, "PCA versus LDA", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228-233, 2001.

(1) pri malých trénovacích množinách, PCA dáva lepšie výsledky ako LDA

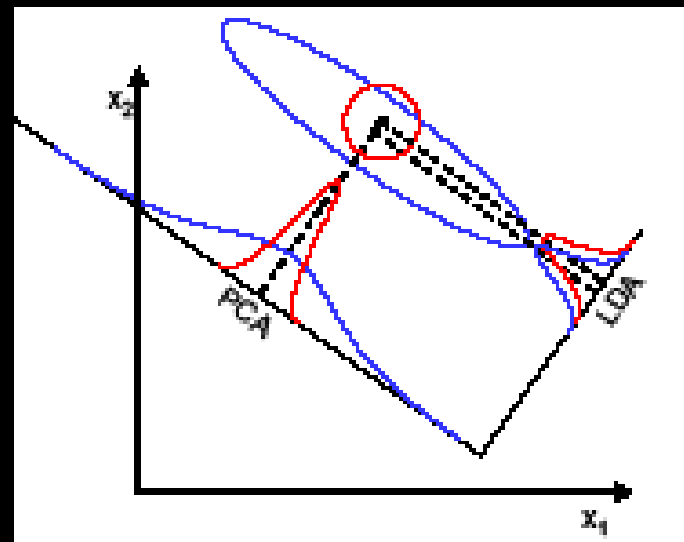
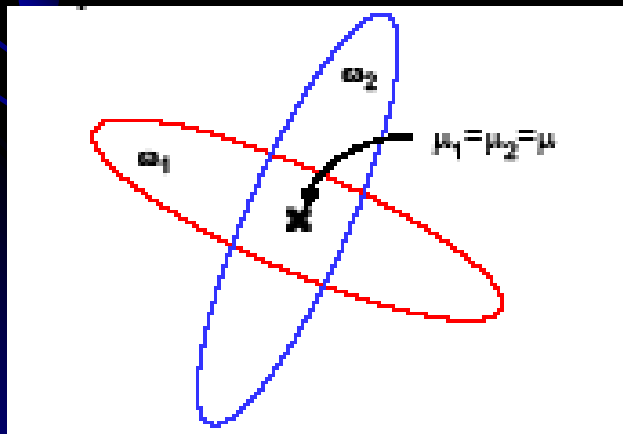
(2) keď máme dostatočný počet trénovacích dát pre každú triedu, LDA dáva lepšie výsledky ako PCA



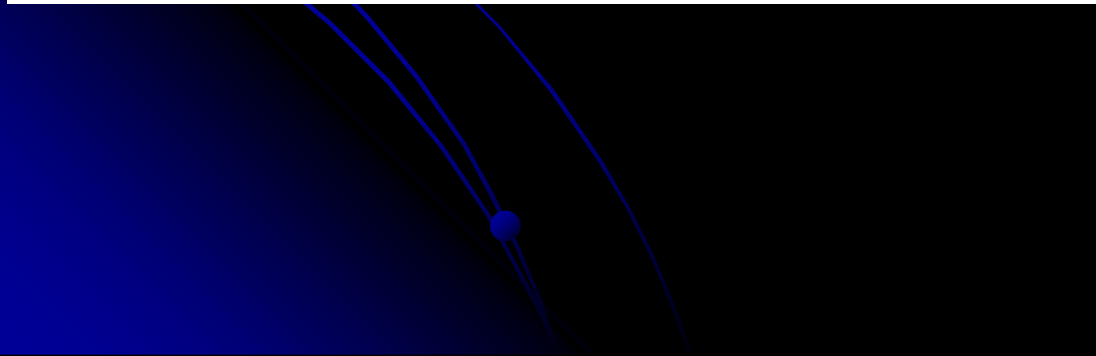
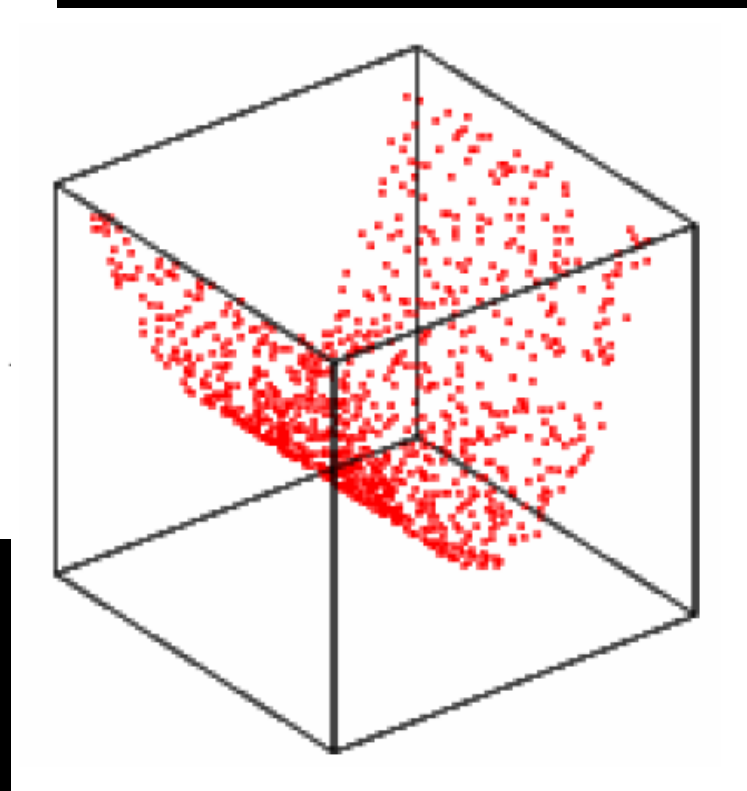
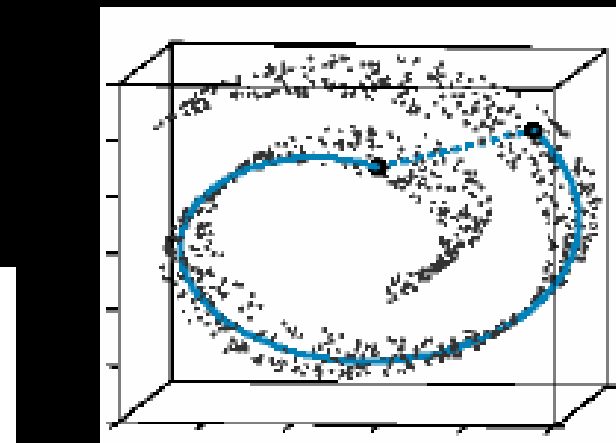
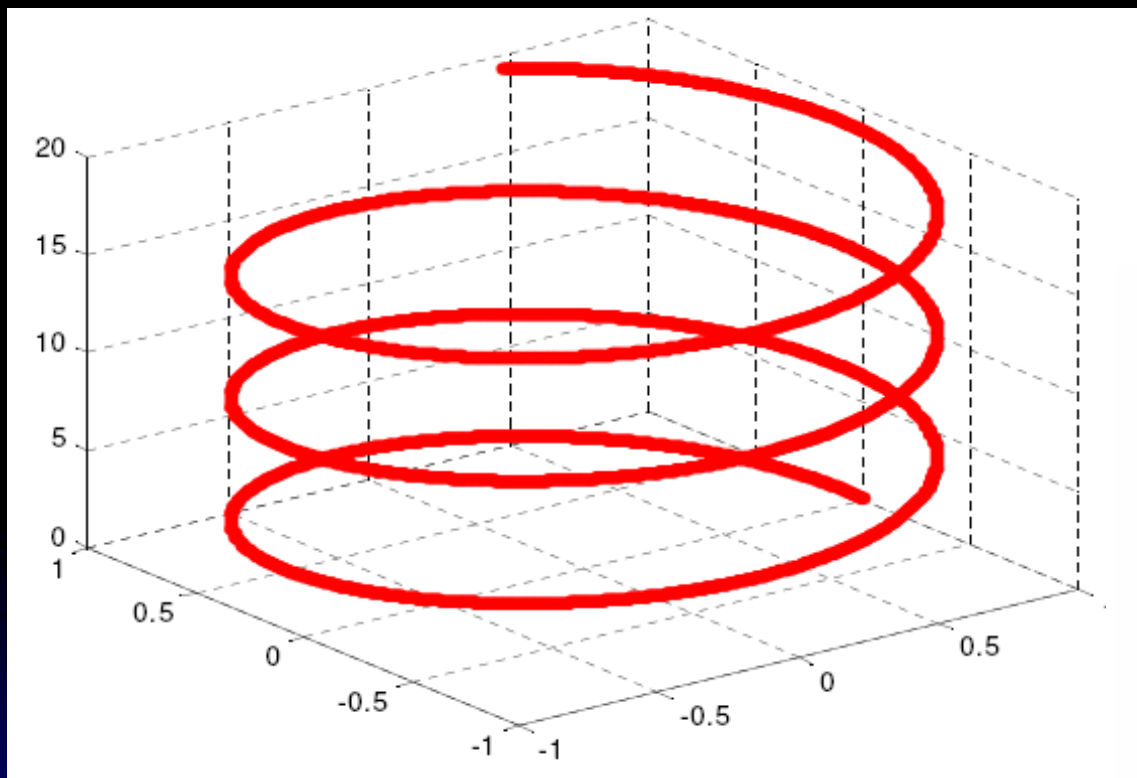
LDA predpokladá normálne rozdelenie dát



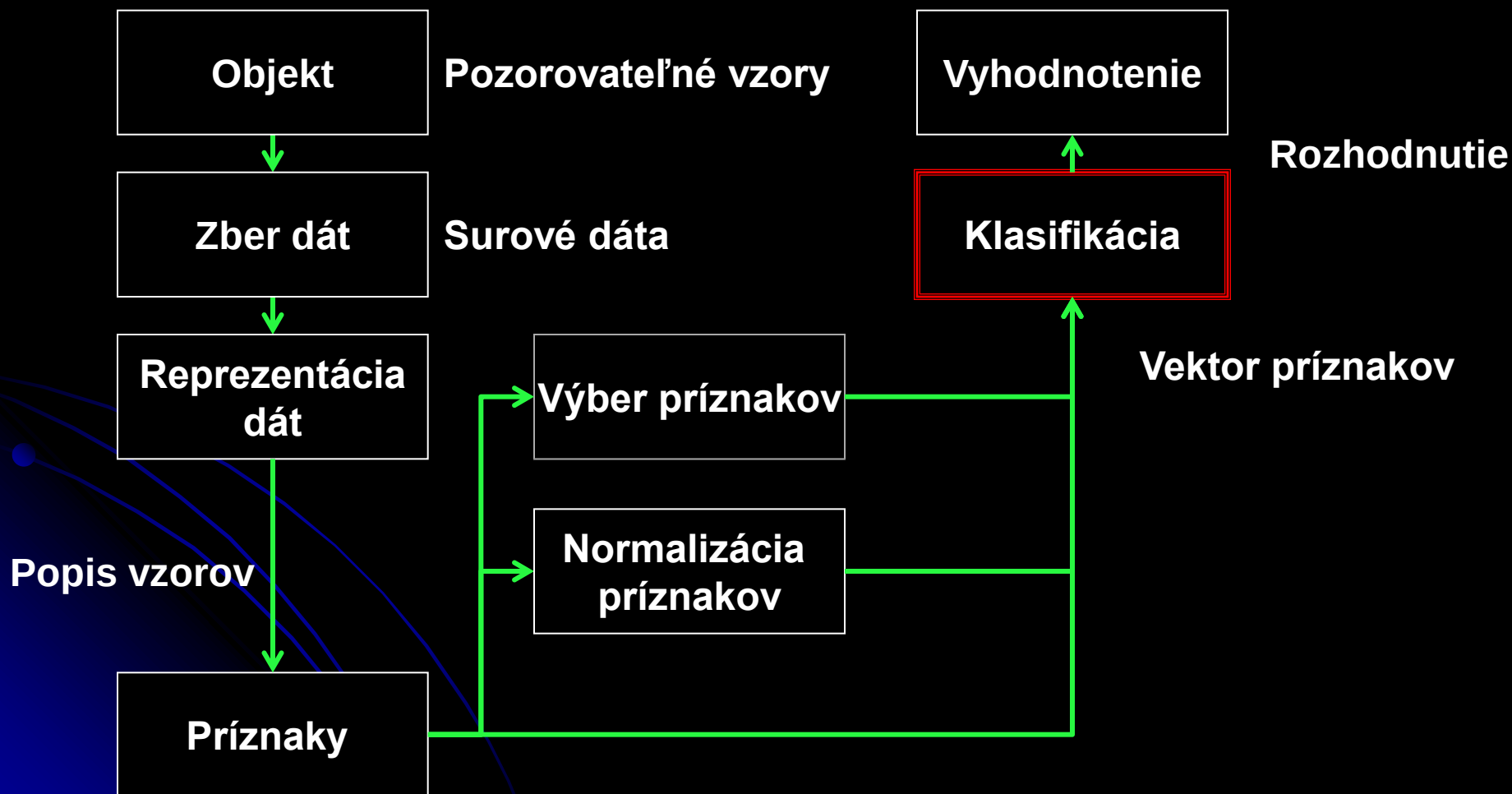
Ak je rozdiel medzi triedami najmä vo variancii a nie v priemere, LDA tieto dáta neseparuje správne



Nelineárne metódy



Rozpoznávanie



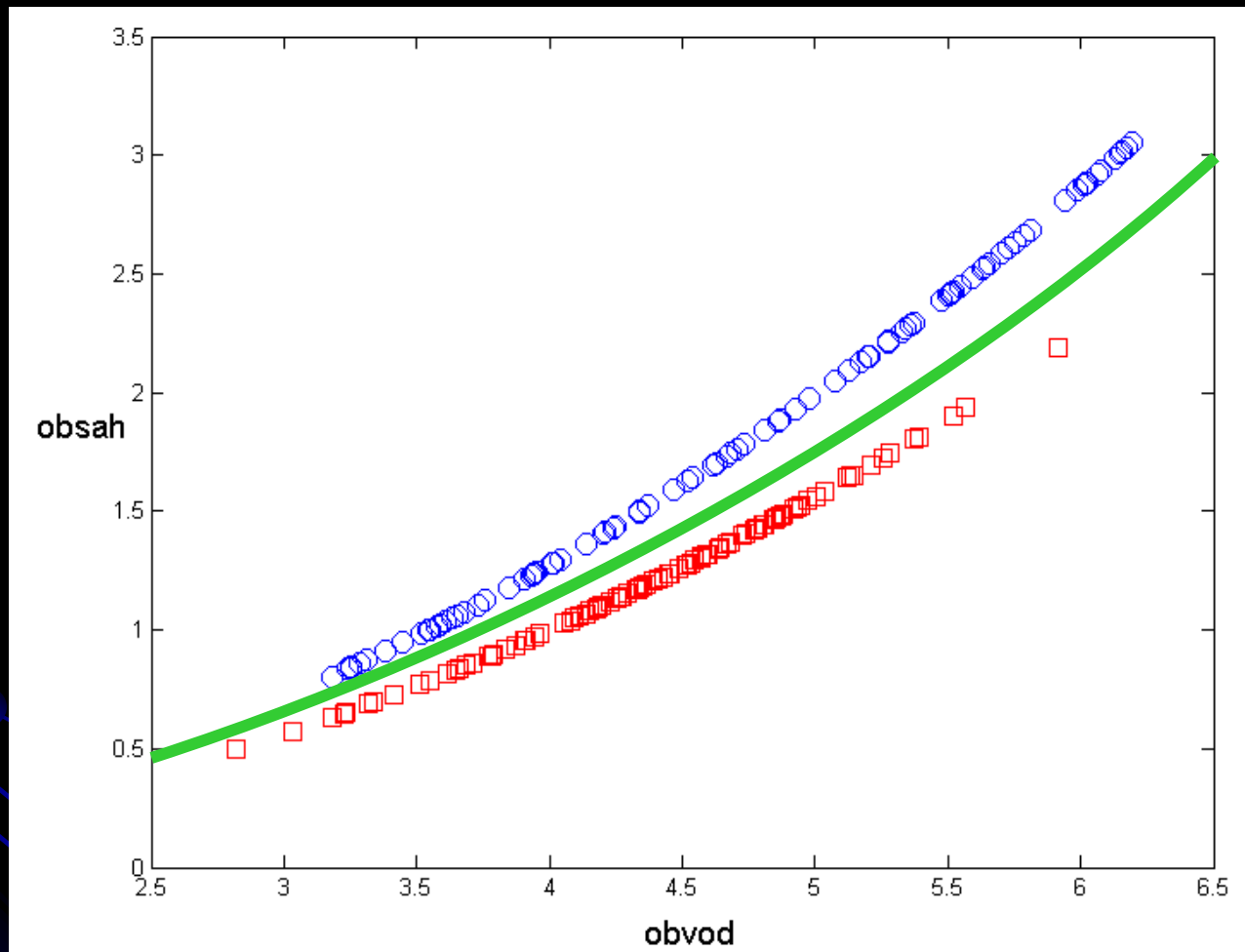
Klasifikácia

Obchodný dom = oddelenia (typ/druh tovaru)

alternatívne kritériá výberu príznakov vedú
k rôznemu výsledku klasifikácie

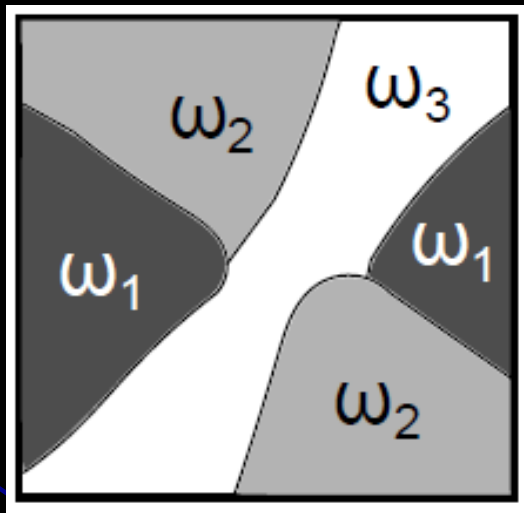
určenie triedy objektu na základe príznakov
získaných z objektu





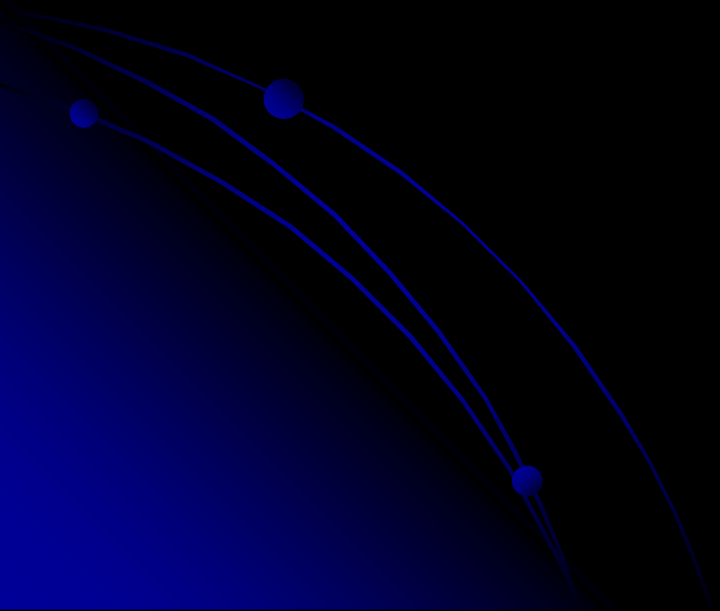
$$A = P^2 / k, \text{ where } 4\pi < k < 16$$

Rozhodovacie oblasti



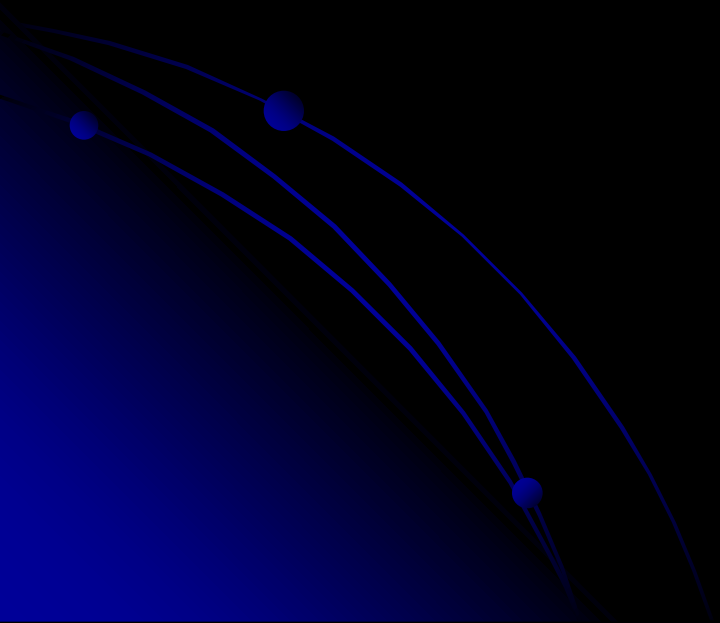
Príznakové metódy

- štatistické metódy
- vzdialenosti medzi príznakmi
- biologicky inšpirované metódy
- pravidlá (aj nemetrické príznaky)



Riadené metódy

Jedna premenná, ktorá sa nazýva cieľová (klasifikačná trieda), riadi proces učenia tak, aby ostatné premenné (prediktory), čo najlepšie predpovedali (v trénovacej vzorke známu) hodnotu cieľovej premennej.



Riadené metódy klasifikácie

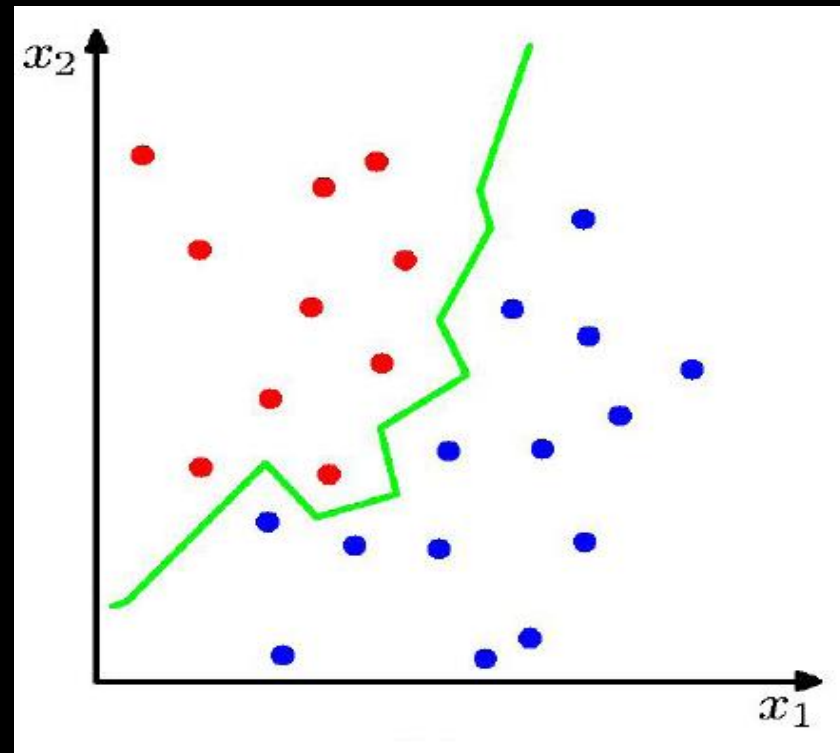
trénovacia množina

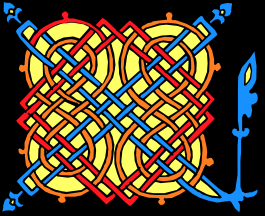
N pozorovaní $(x_1, \dots, x_N)$, $x_i \in \mathbb{R}^d$

triedy $(y_1, \dots, y_N)$, $y_i \in \{-1, 1\}$

klasifikačný problém:

určiť $f(x)$ tak, aby $f(x_i) = y_i$





Bayesovský klasifikátor

objekt je priradený do triedy, ktorá je najpravdepodobnejšia
vzhľadom na príznaky

- podmienené rozdelenie pravdepodobnosti $P(x|\omega_i)$ pre každú klasifikačnú triedu ω_i
- apriórna pravdepodobnosť $P(\omega_i)$ pre každú ω_i (pred pozorovaním vektora x)
- nepodmienené rozdelenie pravdepodobnosti $P(x)$ – určuje pravdepodobnosť vektora x bez ohľadu na triedu

Naivný Bayesovský klasifikátor

vychádza z predpokladu, že podmienené pravdepodobnosti jednotlivých príznakov sú štatisticky nezávislé, t.j. hodnota každého príznaku vektora x v danej triede, nie je ovplyvnená hodnotami ostatných príznakov

$$P(x_1, x_2, \dots, x_d | \omega_i) = P(x_1 | \omega_i) P(x_2 | \omega_i) \dots P(x_d | \omega_i)$$

Bayesovský klasifikátor priradí objekt triede, kde $P(\omega_i|\mathbf{x})$ je maximálne

$$P(\omega_i|\mathbf{x}) = \frac{P(\mathbf{x}|\omega_i)P(\omega_i)}{P(\mathbf{x})}$$

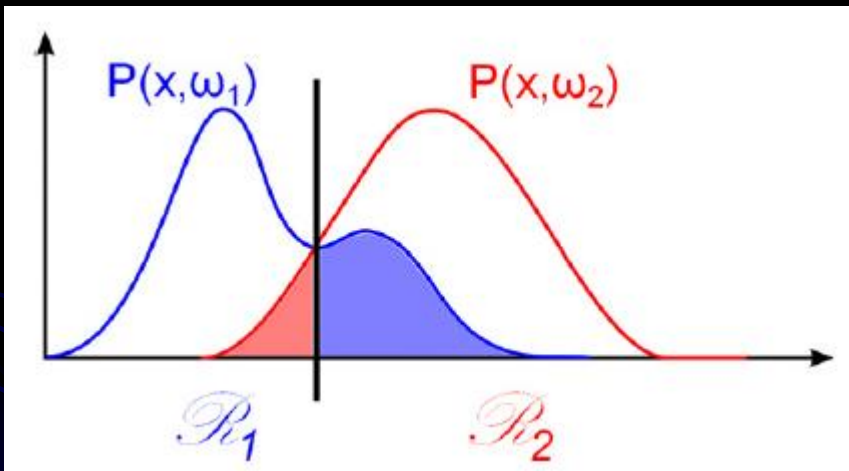
$$\frac{P(\mathbf{x}|\omega_i)P(\omega_i)}{P(\mathbf{x})} \geq \frac{P(\mathbf{x}|\omega_j)P(\omega_j)}{P(\mathbf{x})}$$

Rozhodovacia funkcia

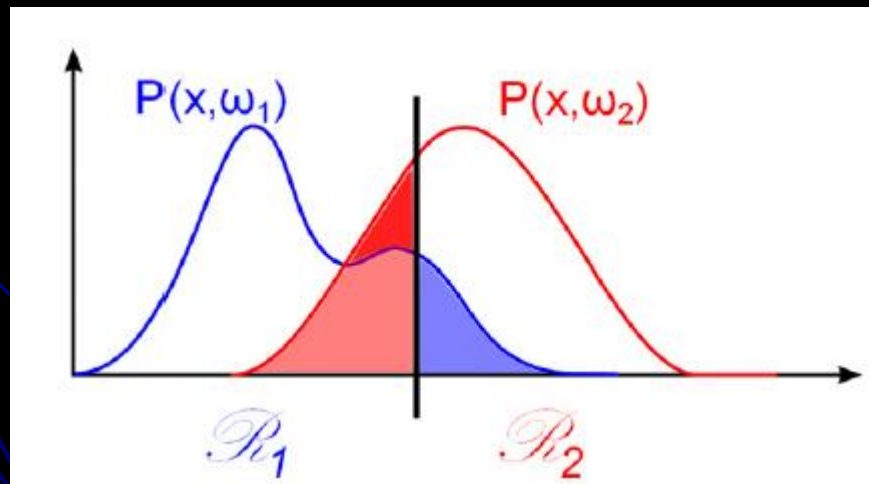
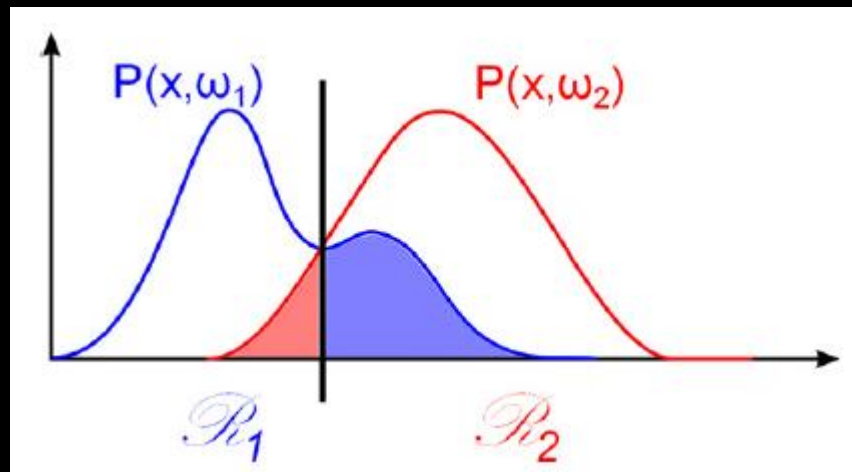
$$d_i(\mathbf{x}) = P(\mathbf{x}|\omega_i)P(\omega_i)$$

Optimalita Bayesovho klasifikátora

$$P(\text{chyba}) = \int_{\mathcal{R}_1} P(\mathbf{x}, \omega_2) d\mathbf{x} + \int_{\mathcal{R}_2} P(\mathbf{x}, \omega_1) d\mathbf{x},$$

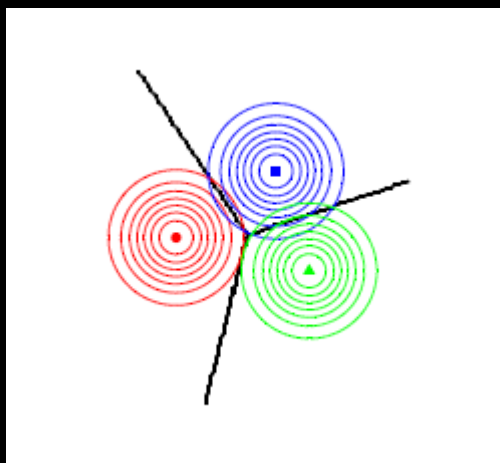


$$\frac{P(\mathbf{x}|\omega_i)P(\omega_i)}{P(\mathbf{x})} \geq \frac{P(\mathbf{x}|\omega_j)P(\omega_j)}{P(\mathbf{x})}$$

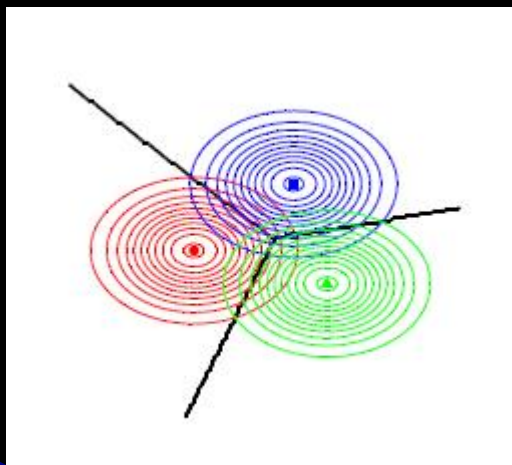


Rozhodovacie oblasti

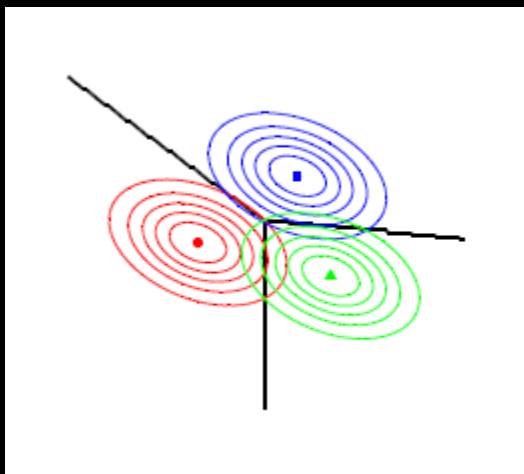
štatisticky nezávislé
rovnaká variancia



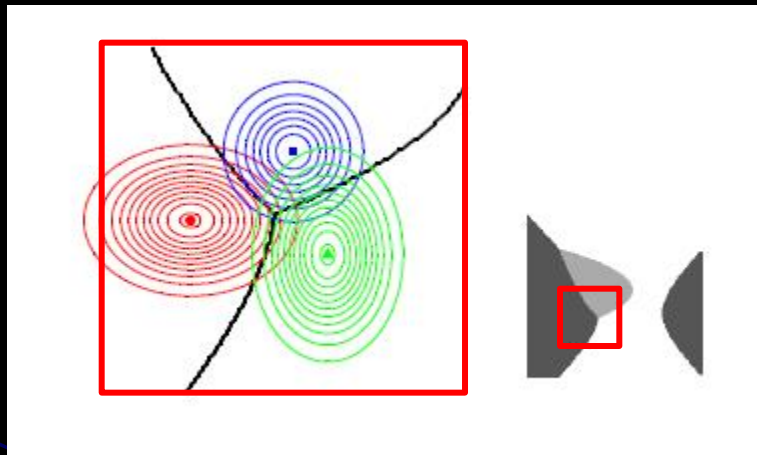
Rovnaká kovariančná matica
Rôzne variancie



Rovnaká kovariančná matica
Nie diagonálna



Diagonálna kovariančná matica, rôzna



Rôzna kovariančná matica

