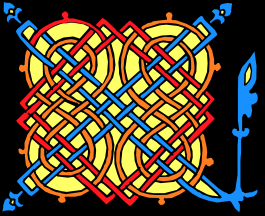


Diskriminačná analýza

predpokladáme, že klasifikačné triedy sú
lineárne separovateľné

existuje nadrovina rozdeľujúca priestor tak,
že vektory rôznych tried ležia v rôznych
polpriestoroch

Lineárny klasifikátor



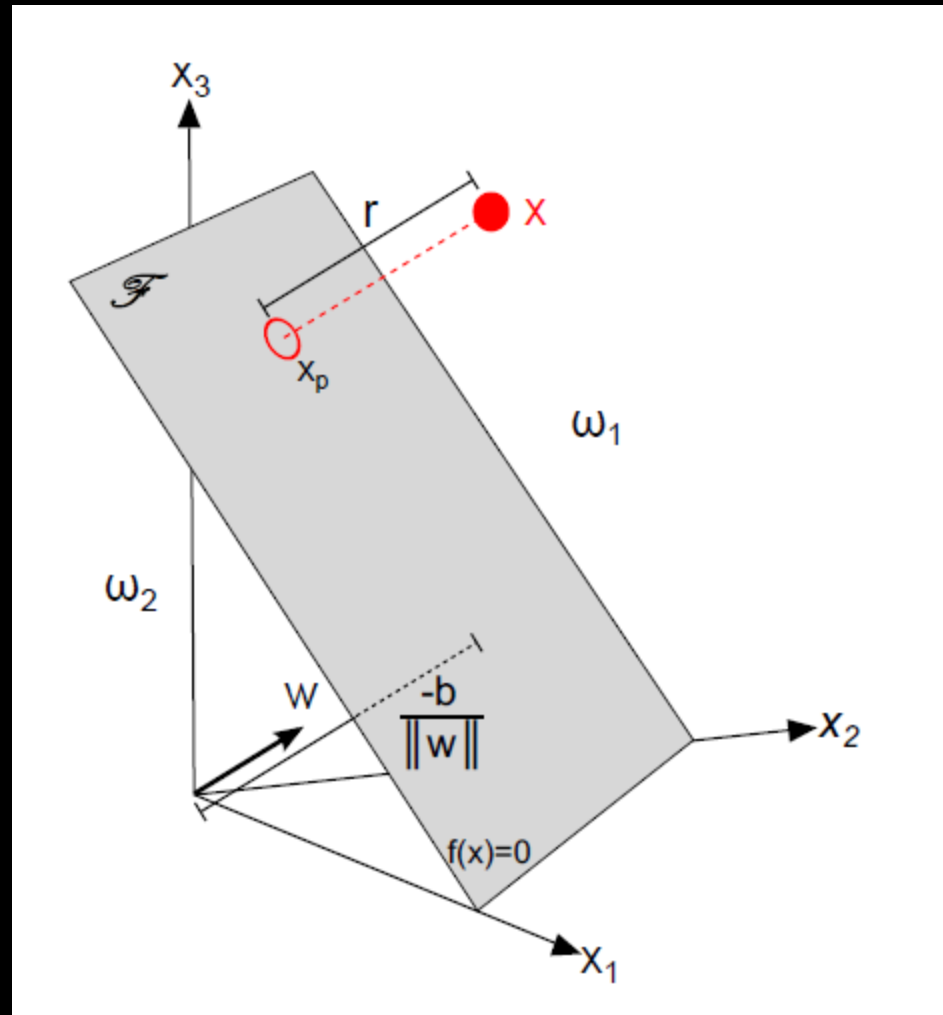
Klasifikácia do 2 tried

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b$$

nadrovina: $\mathbf{w}^T \mathbf{x} + b = 0$
 $\mathbf{w}$ váhový vektor

$$f(\mathbf{x}) \geq 0 \text{ pre } \mathbf{x} \in \omega_1$$

$$f(\mathbf{x}) < 0 \text{ pre } \mathbf{x} \in \omega_2$$



Zovšeobecnený lineárny klasifikátor

$$f(\mathbf{x}) = \mathbf{u}^T \mathbf{y},$$

$$\mathbf{u} = [b, \mathbf{w}^T]^T \text{ a } \mathbf{y} = [1, \mathbf{x}^T]^T$$

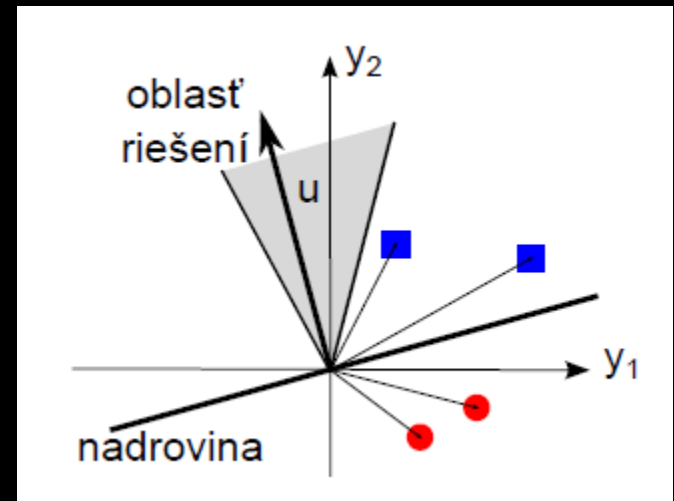
Vzorku y_i zaradíme

do triedy ω_1

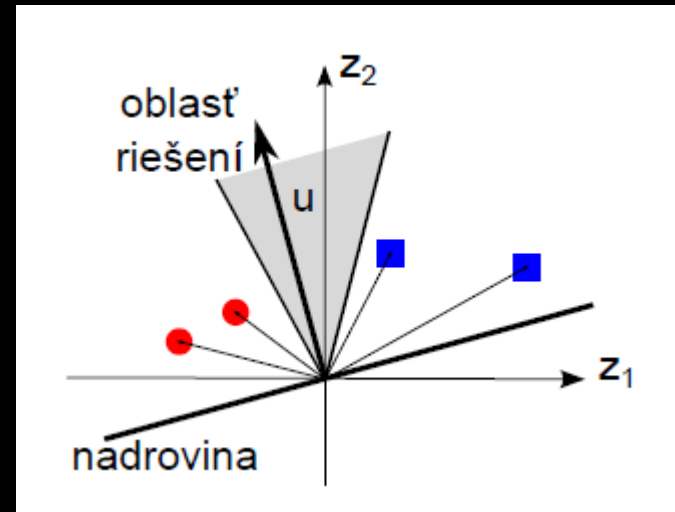
$$\mathbf{u}^T \mathbf{y}_i > 0$$

do triedy ω_2

$$\mathbf{u}^T \mathbf{y}_i < 0$$



$$\mathbf{z}_i = \begin{cases} \mathbf{y}_i & \text{pre } \mathbf{y} \in \omega_1, \\ -\mathbf{y}_i & \text{pre } \mathbf{y} \in \omega_2, \end{cases}$$



nájsť váhový vektor $\mathbf{u}$, aby platilo pre všetky $\mathbf{z}_i$

$$\mathbf{u}^T \mathbf{z}_i > 0$$

Gradientná metóda

skalárna účelová funkcia $O(\mathbf{u})$, ktorej hodnota je minimálna, keď $\mathbf{u}$ je riešením

náhodne zvolený vektor $\mathbf{u}_1$

gradient ∇O v bode $\mathbf{u}_1$

aktualizácia hodnoty $\mathbf{u}$ tak, že vyberieme vektor z okolia $\mathbf{u}_1$ v smere najväčšej negatívnej zmeny gradientu

$$\mathbf{u}_{i+1} = \mathbf{u}_i - \eta(i) \nabla O|_{\mathbf{u}_i}$$

kde $\eta(i)$ je (pozitívna) rýchlosť učenia v kroku i

účelová funkcia je diferencovateľná

Taylorov rozvoj

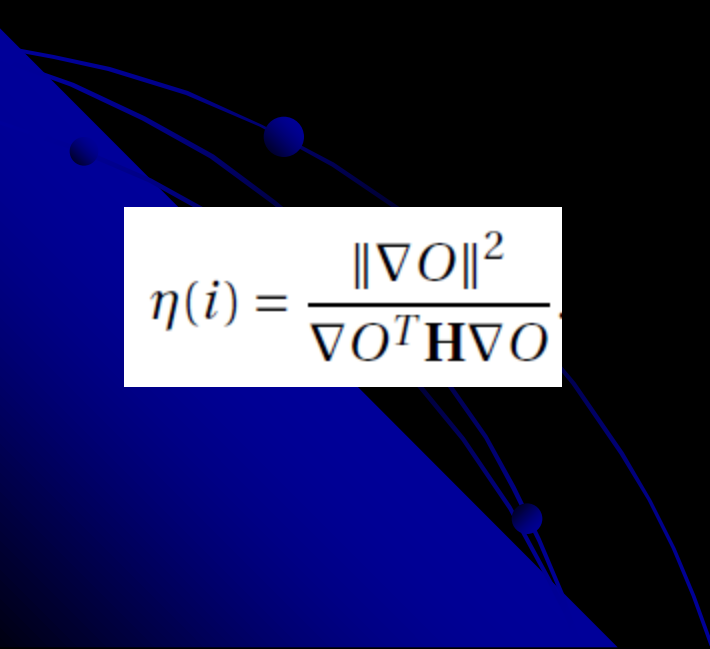
$O(\mathbf{a})$ je v okolí bodu $\mathbf{u}_i$ aproximovaná

$$O(\mathbf{a}) \approx O(\mathbf{u}_i) + \nabla O^T (\mathbf{a} - \mathbf{u}_i) + \frac{1}{2} (\mathbf{a} - \mathbf{u}_i)^T \mathbf{H} (\mathbf{a} - \mathbf{u}_i)$$

kde $\mathbf{H}$ je matica druhých parciálnych derivácií funkcie O (Hesseho matica).

$$\mathbf{u}_{i+1} = \mathbf{u}_i - \eta(i) \nabla O|_{\mathbf{u}_i}$$

$$O(\mathbf{u}_{i+1}) \approx O(\mathbf{u}_i) - \eta(i) \|\nabla O\|^2 + \frac{1}{2} \eta^2(i) \nabla O^T \mathbf{H} \nabla O$$


$$\eta(i) = \frac{\|\nabla O\|^2}{\nabla O^T \mathbf{H} \nabla O}$$

Účelová funkcia

po častiach lineárna perceptrónová funkcia

$$O_P = \sum_{\mathbf{z} \in \mathcal{Z}} (-\mathbf{u}^T \mathbf{z})$$

$\mathcal{Z}$ – množina nesprávne klasifikovaných vzoriek

$$\nabla O_P = \sum_{\mathbf{z} \in \mathcal{Z}} (-\mathbf{z})$$

$$\mathbf{u}_{i+1} = \mathbf{u}_i + \eta(i) \sum_{\mathbf{z} \in \mathcal{Z}_i} \mathbf{z}$$

vyhodnocovanie vzoriek po jednej

inicializuj $\mathbf{u}_0 = \mathbf{1}$, $i=0$, $k=0$

opakuj

ak $\mathbf{z}_k$ je nesprávne klasifikovaný

$$\mathbf{u}_{i+1} = \mathbf{u}_i + \mathbf{z}_k$$

$$i=i+1$$

koniec ak

$$k=(k+1) \bmod N$$

kým existuje nejaký nesprávne klasifikovaný vektor $\mathbf{z}_j$

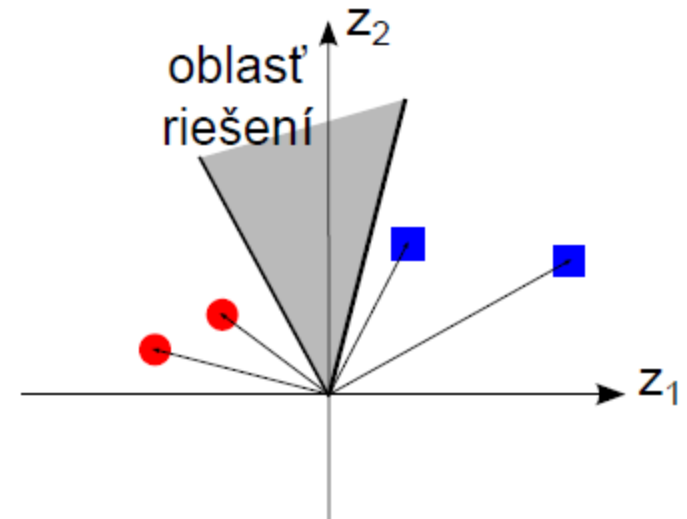
kvadratická funkcia

$$O_Q = \sum_{\mathbf{z} \in \mathcal{X}_i} (\mathbf{u}^T \mathbf{z})^2$$

spojitý gradient

príliš hladká

môže konvergovať na
hranici oblasti riešení

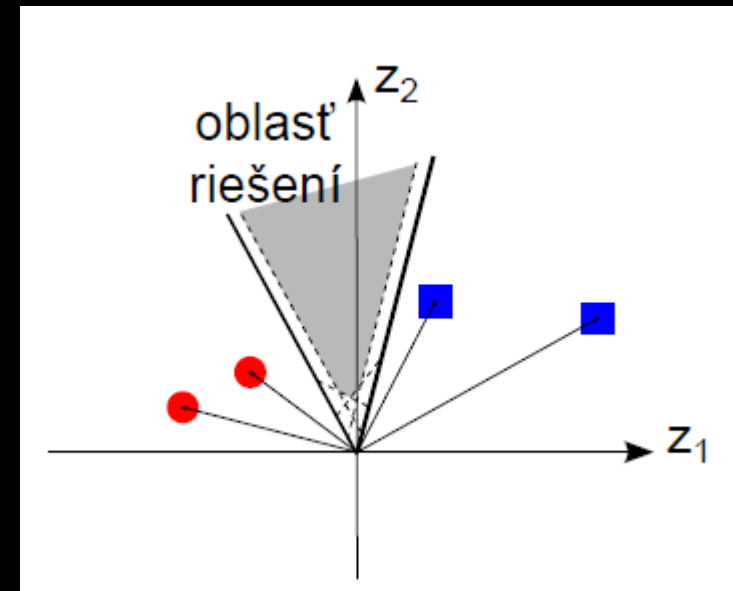


normalizačný člen

člen určujúci okraj v oblasti riešení

$$O_R = \frac{1}{2} \sum_{\mathbf{z} \in \mathcal{Z}_i} \frac{(\mathbf{u}^T \mathbf{z} - \epsilon)^2}{\|\mathbf{z}\|^2}$$

$$\nabla O_R = \sum_{\mathbf{z} \in \mathcal{Z}_i} \frac{\mathbf{u}^T \mathbf{z} - \epsilon}{\|\mathbf{z}\|^2} \mathbf{z}$$



inicializuj $\mathbf{u}_0 = 0$, $i=0$, $k=0$

opakuj

$k=(k+1) \bmod N$

ak $\mathbf{u}^T \mathbf{z}_k \leq \epsilon$

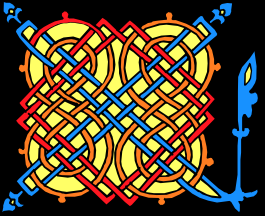
$\mathbf{z} = \mathbf{z}_k$

$\mathbf{u}_{i+1} = \mathbf{u}_i + \eta(i) \frac{\epsilon - \mathbf{u}^T \mathbf{z}}{\|\mathbf{z}\|^2} \mathbf{z}$

$i=i+1$

koniec ak

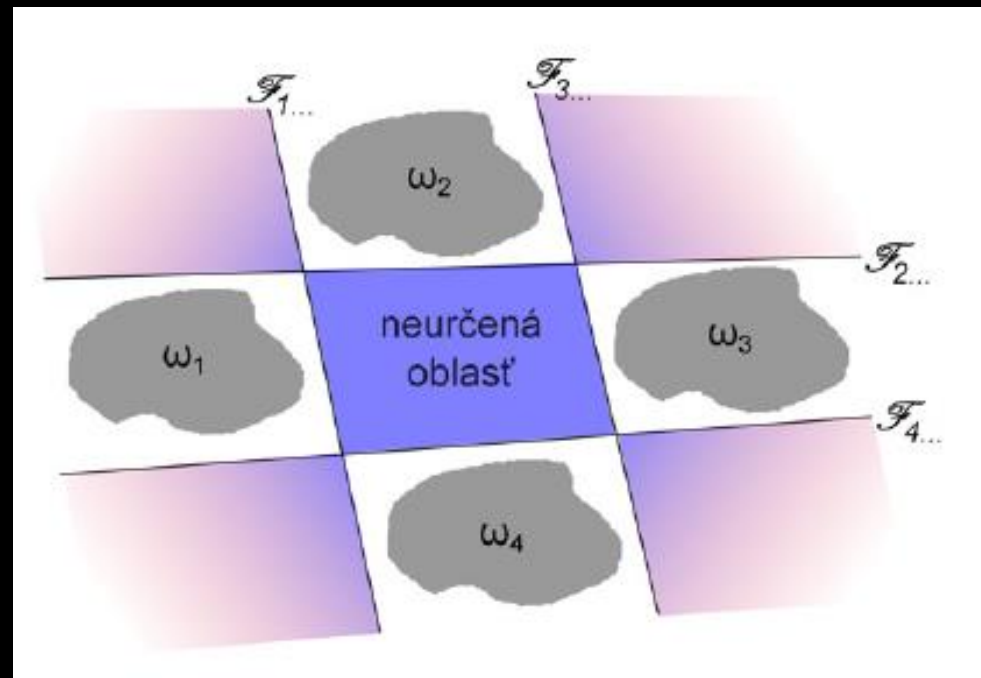
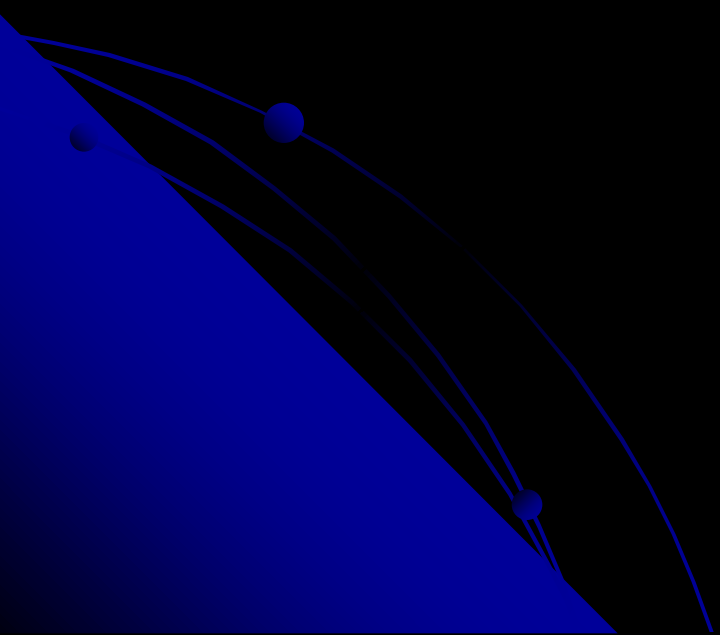
kým $\mathbf{u}^T \mathbf{z}_j > \epsilon$ pre všetky $\mathbf{z}_j$



Klasifikácia do K tried

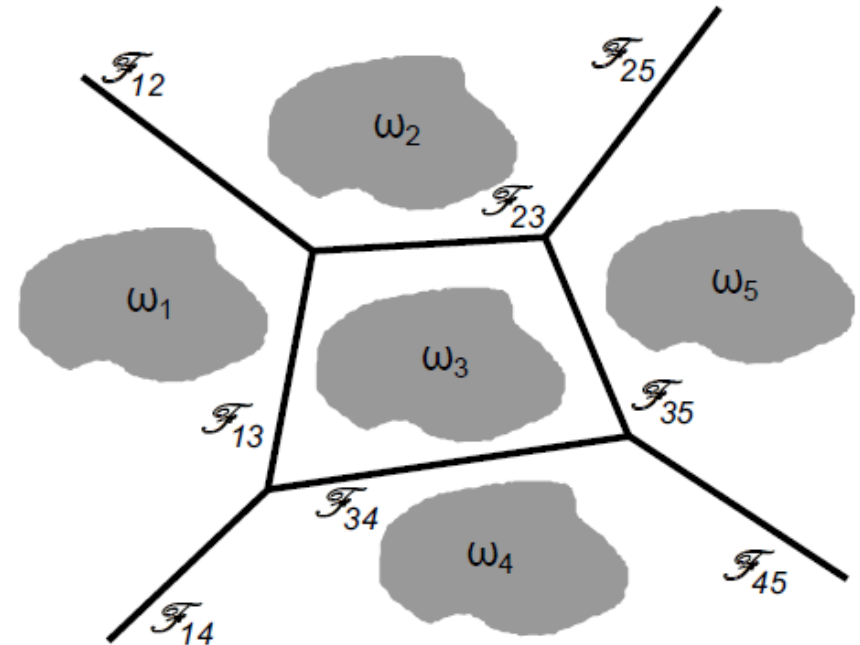
1. Pre každú klasifikačnú triedu vytvoríme rozhodovaciu funkciu, ktorá oddelí vzorky tejto triedy od vzoriek ostatných tried.

K rozhodovacích funkcií



2. Pre každú dvojicu rôznych klasifikačných tried vytvoríme rozhodovaciu funkciu oddeľujúcu vzorky týchto dvoch tried bez ohľadu na vzorky ostatných tried.

$K(K - 1)/2$ rozhodovacích funkcií



Lineárny stroj

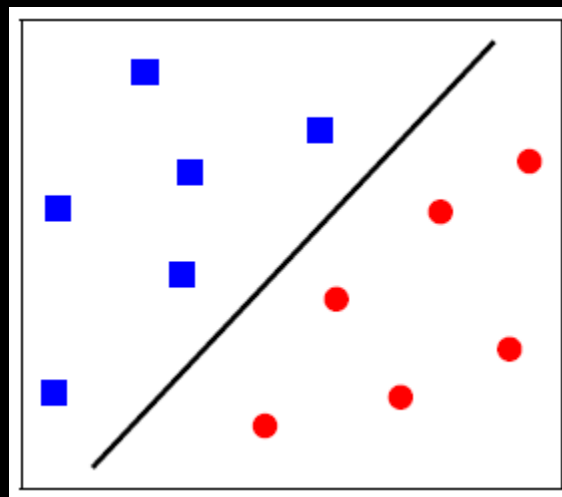
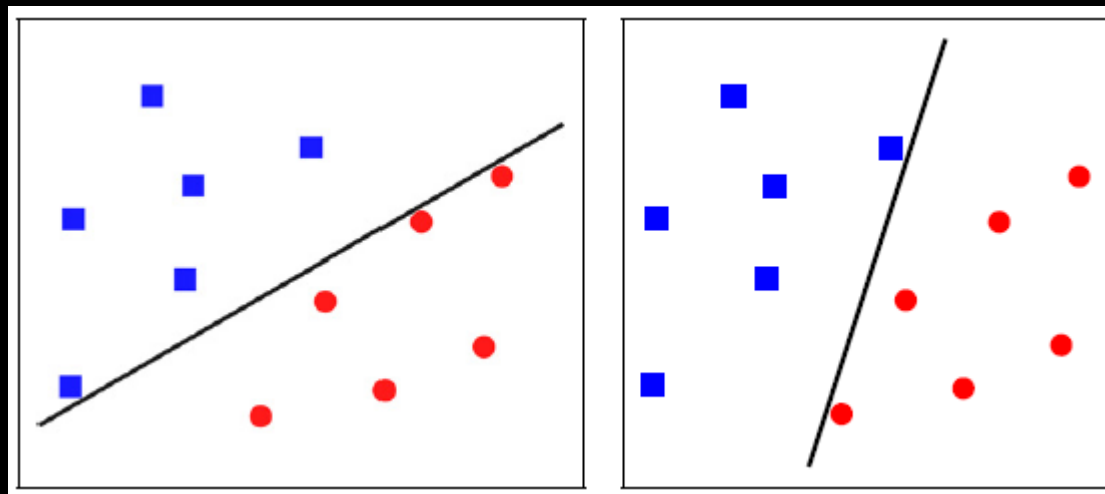
K rozhodovacích funkcií

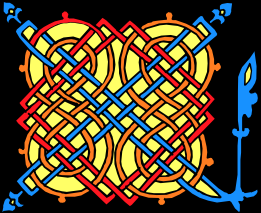
$$\{f_k(\mathbf{x})\}_{k=1}^K$$

$\mathbf{x}$ patrí do tej klasifikačnej triedy, ktorej hodnota po dosadení je najväčšia

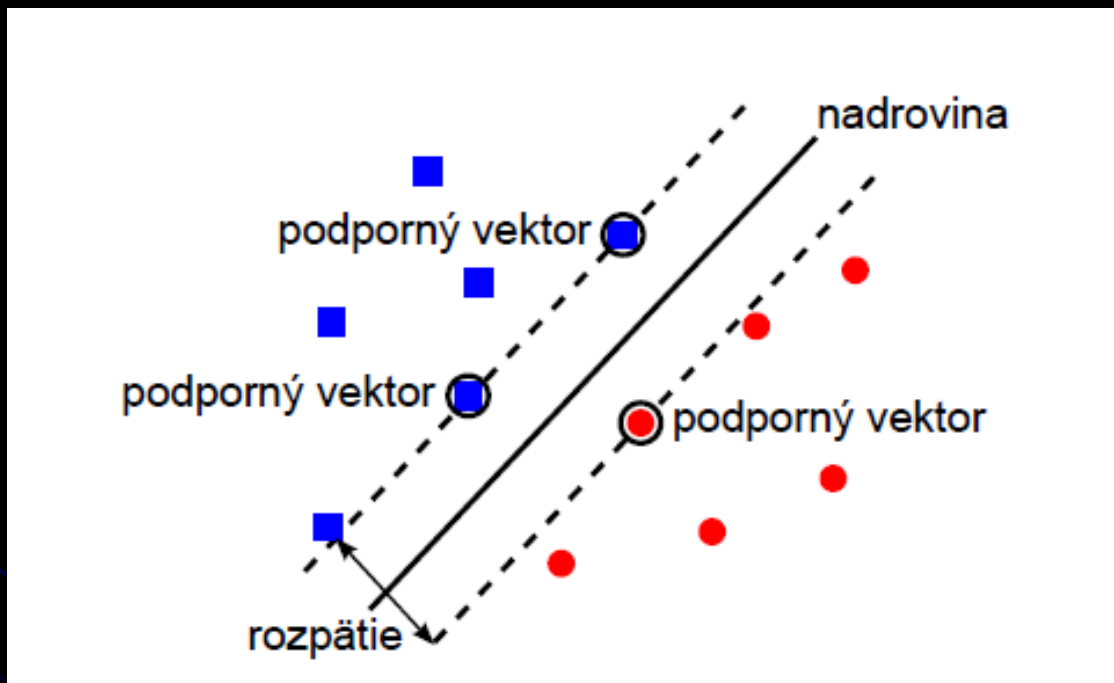
$$\mathbf{x} \in \omega_i \Leftrightarrow \forall_{j \neq i} f_i(\mathbf{x}) > f_j(\mathbf{x})$$

rozdelenie priestoru na K konvexných regiónov





Mechanizmus podporných vektorov



$$\mathbf{w}^T \mathbf{x} + b = \pm 1$$

$$\frac{2}{\|\mathbf{w}\|}$$

$$\mathbf{w}^* = \operatorname{argmax} \frac{2}{\|\mathbf{w}\|}$$

$$(\mathbf{w}^*, b^*) = \operatorname{argmin} \frac{1}{2} \|\mathbf{w}\|^2,$$

aby

$$\mathbf{w}^T \mathbf{x}_i + b \geq 1 \text{ pre } \mathbf{x} \in \omega_1,$$

$$\mathbf{w}^T \mathbf{x}_i + b \leq -1 \text{ pre } \mathbf{x} \in \omega_2.$$

$$(\mathbf{w}^T \mathbf{x}_i + b) k_i \geq 1$$

$$k_i = \begin{cases} 1 & \text{pre } \mathbf{x} \in \omega_1 \\ -1 & \text{pre } \mathbf{x} \in \omega_2 \end{cases}$$

$$\begin{aligned} L(\mathbf{w}, b, \alpha_i) &= \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^N \alpha_i ((\mathbf{w}^T \mathbf{x}_i + b) k_i - 1) \\ &= \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^N \alpha_i (\mathbf{w}^T \mathbf{x}_i + b) k_i + \sum_{i=1}^N \alpha_i, \end{aligned}$$

aby

$$\alpha_i \geq 0.$$

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^N \alpha_i k_i \mathbf{x}_i \equiv 0$$

$$\frac{\partial L}{\partial b} = - \sum_{i=1}^N \alpha_i k_i \equiv 0$$

$$(\mathbf{w}^T \mathbf{x}_i + b) k_i - 1 \geq 0$$

$$\alpha_i \geq 0$$

$$\alpha_i ((\mathbf{w}^T \mathbf{x}_i + b) k_i - 1) = 0$$

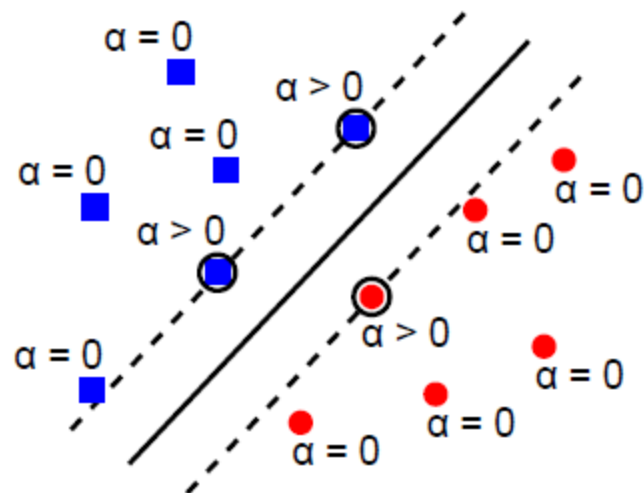
$$\mathbf{w} = \sum_{i=1}^N \alpha_i k_i \mathbf{x}_i$$

$$\sum_{i=1}^N \alpha_i k_i = 0$$

$$\alpha_i = 0 \Leftrightarrow (\mathbf{w}^T \mathbf{x}_i + b) k_i > 1$$

$$\alpha_i > 0 \Leftrightarrow (\mathbf{w}^T \mathbf{x}_i + b) k_i = 1$$

Kuhn Tuckerove podmienky



$$\mathbf{w} = \sum_{i=1}^N \alpha_i k_i \mathbf{x}_i$$

$$\sum_{i=1}^N \alpha_i k_i = 0$$

$$L(\mathbf{w}, b, \alpha_i) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^N \alpha_i (\mathbf{w}^T \mathbf{x}_i + b) k_i + \sum_{i=1}^N \alpha_i$$

Aby sme našli hodnoty α , spočítame duálny problém

$$\tilde{\alpha}_i = \arg \max \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j k_i k_j \mathbf{x}_i^T \mathbf{x}_j$$

aby

$$\alpha_i \geq 0$$

$$\sum_{j=1}^N \alpha_i k_i = 0$$

1. Duálnu úlohu vyriešime metódami kvadratického programovania.

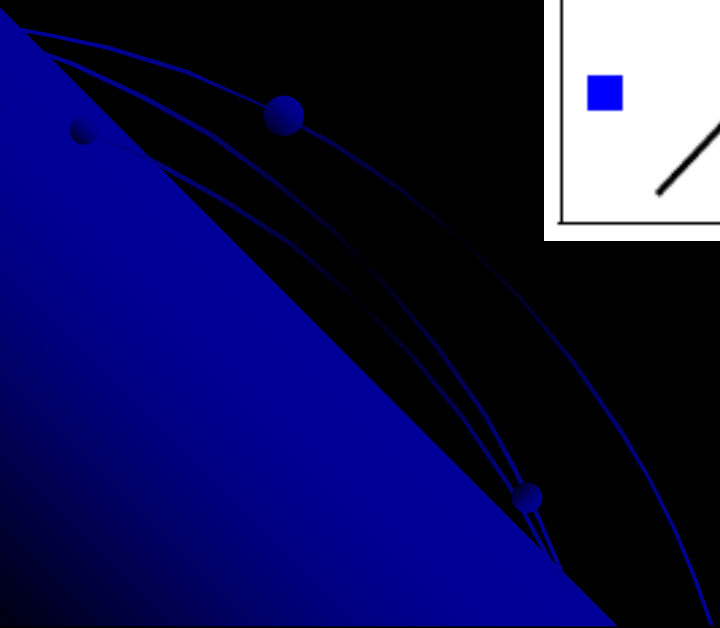
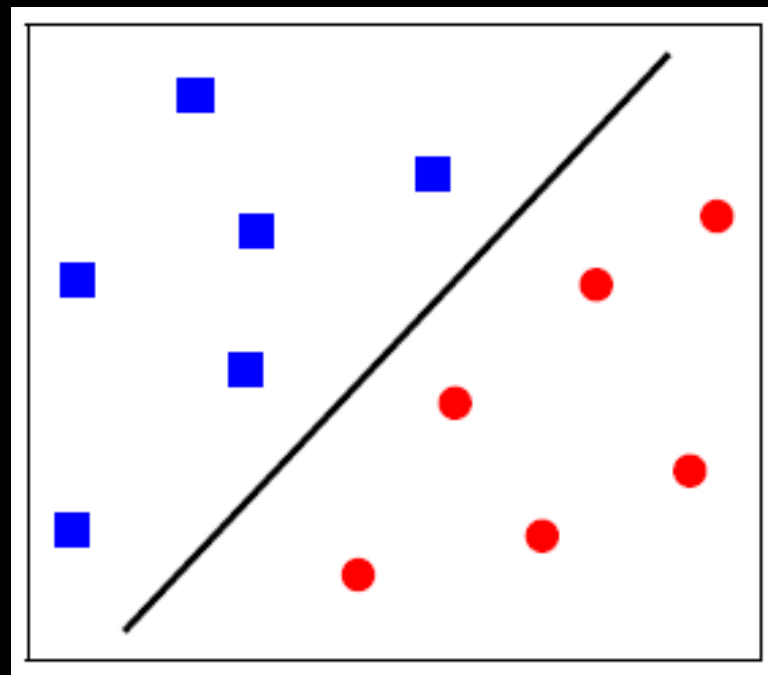
2. Určíme množinu podporných vektorov $\mathbf{x}_i \in SV$, pre ktoré $\alpha_i > 0$.

3. Určíme hodnotu

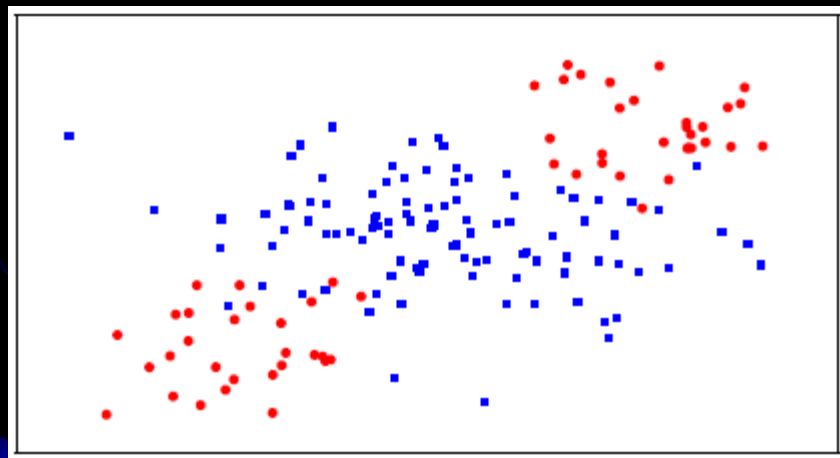
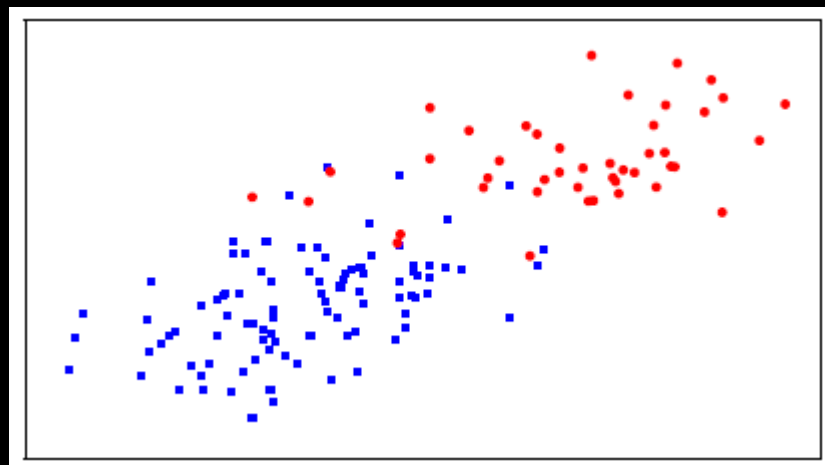
$$b = \frac{1}{|SV|} \sum_{\mathbf{x}_j \in SV} \left(k_j - \sum_{\mathbf{x}_i \in SV} \alpha_i k_i \mathbf{x}_i^T \mathbf{x}_j \right)$$

4. Zistíme hodnotu klasifikačnej funkcie

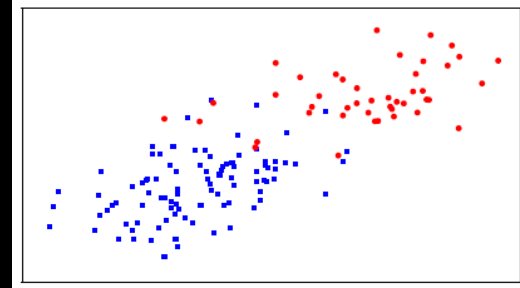
$$f(\mathbf{x}) = \sum_{\mathbf{x}_i \in SV} \alpha_i k_i \mathbf{x}_i^T \mathbf{x} + b$$



Neseparovatelné dáta



SVM s voľným rozpätím



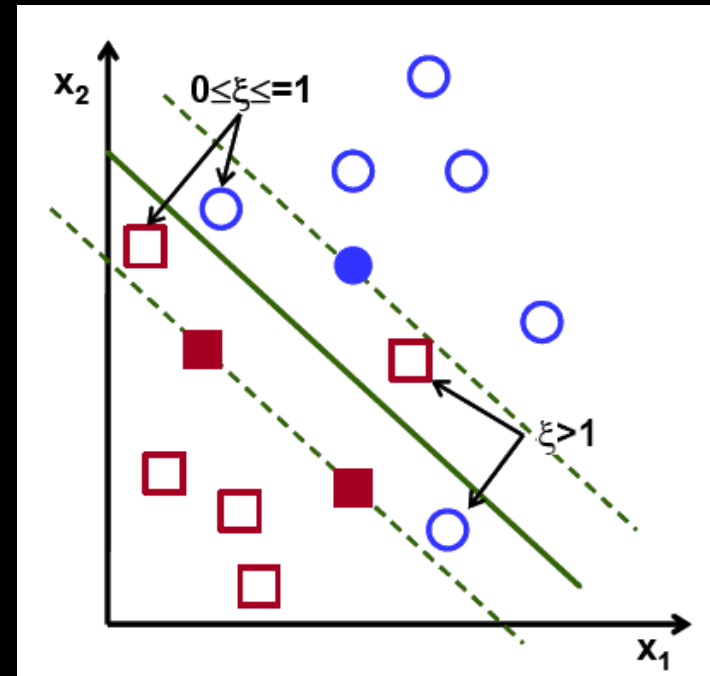
zavedenie druhotných premenných ξ_i

$$(\mathbf{w}^*, b^*, \xi_i^*) = \arg \min \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i,$$

aby

$$(\mathbf{w}^T \mathbf{x}_i + b) k_i \geq 1 - \xi_i,$$

C určuje penalizáciu za chybnú klasifikáciu a ovplyvňuje veľkosť rozpätia



Lagrange:

$$(\tilde{\mathbf{w}}, \tilde{b}, \tilde{\xi}) = \arg \min \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i$$

$$(\mathbf{w}^T \mathbf{x}_i + b)k_i \geq 1 - \xi_i$$

duál:

$$\tilde{\alpha}_i = \arg \max \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j k_i k_j \mathbf{x}_i^T \mathbf{x}_j$$

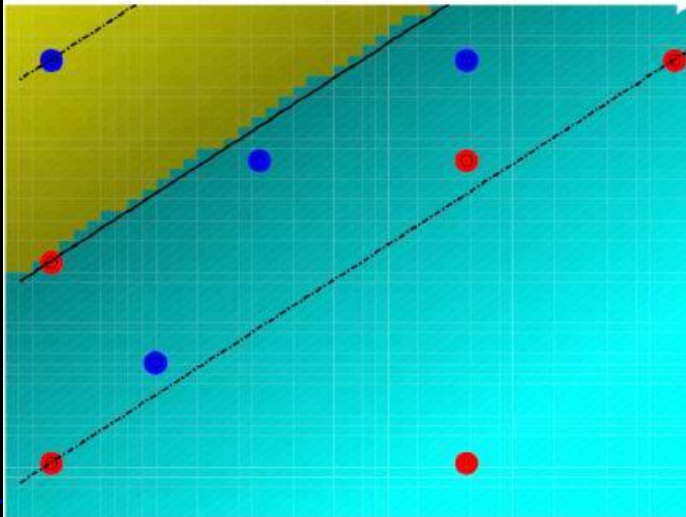
aby

$$0 \leq \alpha_i \leq C$$

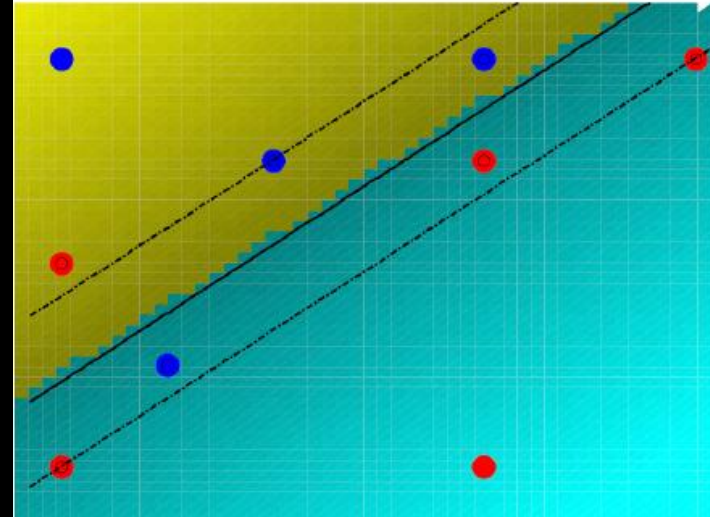
$$\sum_{j=1}^N \alpha_j k_j = 0$$

$$\mathbf{w} = \sum_{i=1}^N \alpha_i k_i \mathbf{x}_i = \sum_{i \in SV} \alpha_i k_i \mathbf{x}_i$$

Effect of C: $C = 1$

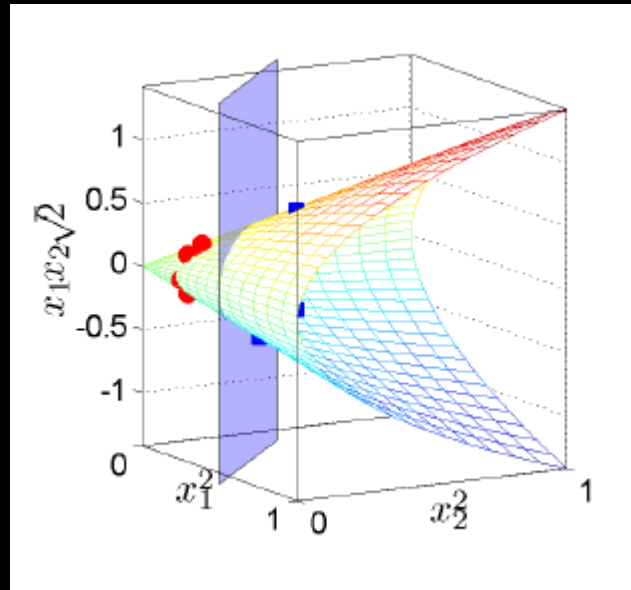
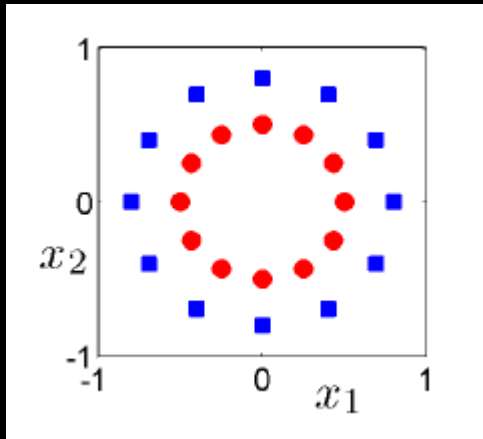
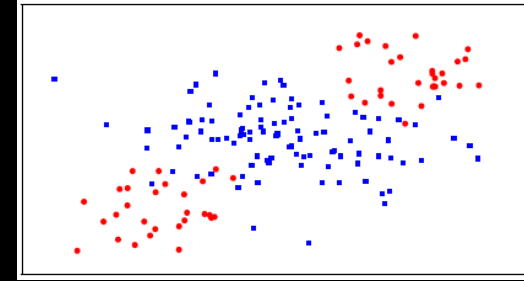


Effect of C: $C = 1000$



Vhodnú hodnotu C nájdeme vzájomnou validáciou

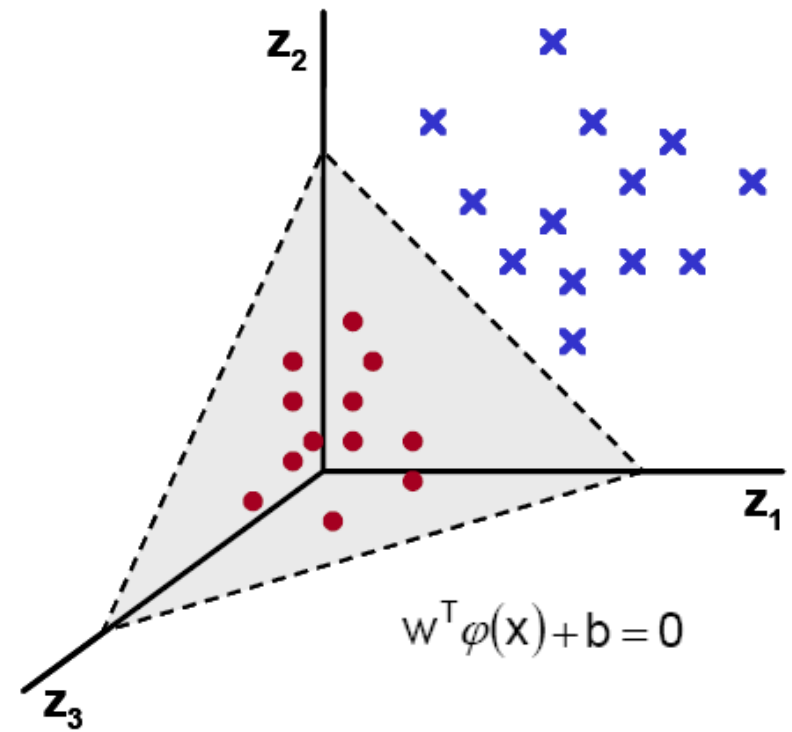
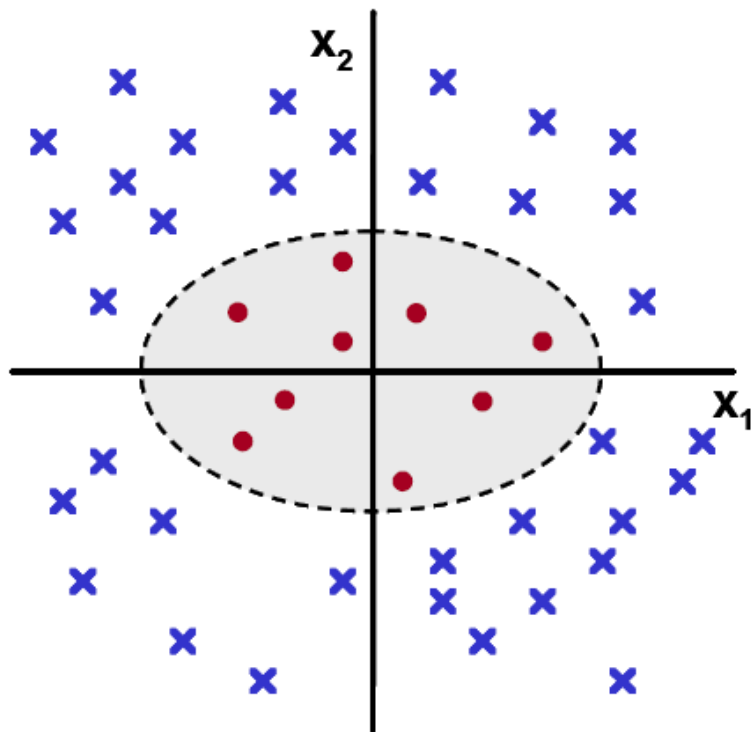
Nelineárny SVM



$$\phi: \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \rightarrow \begin{pmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2}x_1 x_2 \end{pmatrix}$$

$$\varphi: \mathbb{R}^2 \rightarrow \mathbb{R}^3$$

$$\tilde{\alpha}_i = \arg \max \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j k_i k_j \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}_j)$$



1. Duálnu úlohu vyriešime metódami kvadratického programovania.

2. Určíme množinu podporných vektorov $\mathbf{x}_i \in SV$, pre ktoré $\alpha_i > 0$.

3. Určíme hodnotu

$$b = \frac{1}{|SV|} \sum_{\mathbf{x}_j \in SV} \left(k_j - \sum_{\mathbf{x}_i \in SV} \alpha_i k_i \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) \right)$$

4. Zistíme hodnotu klasifikačnej funkcie

$$f(\mathbf{x}) = \sum_{\mathbf{x}_i \in SV} \alpha_i k_i \phi(\mathbf{x}_i)^T \phi(\mathbf{x}) + b$$

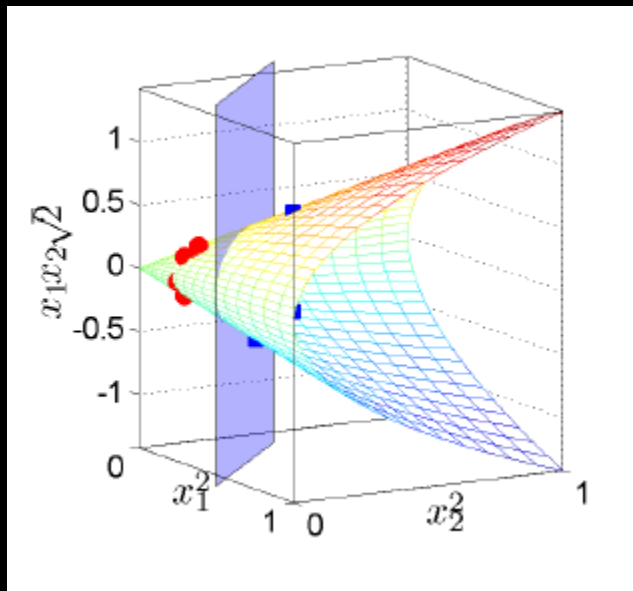
kernelový trik

$$K(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{x})^T \phi(\mathbf{y})$$

$$\alpha^* = \operatorname{argmax} \left(\sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j k_i k_j K(\mathbf{x}_i, \mathbf{x}_j) \right)$$

$$b = \frac{1}{|SV|} \sum_{\mathbf{x}_j \in SV} \left(k_j - \sum_{\mathbf{x}_i \in SV} \alpha_i k_i K(\mathbf{x}_i, \mathbf{x}_j) \right)$$

$$f(\mathbf{x}) = \sum_{\mathbf{x}_i \in SV} \alpha_i k_i K(\mathbf{x}_i, \mathbf{x}) + b$$



$$\phi: \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \rightarrow \begin{pmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2}x_1 x_2 \end{pmatrix}$$

$$K(\mathbf{x}, \mathbf{y}) = (x_1^2, x_2^2, \sqrt{2}x_1 x_2) \cdot \begin{pmatrix} y_1^2 \\ y_2^2 \\ \sqrt{2}y_1 y_2 \end{pmatrix} =$$

$$\begin{aligned} &= x_1^2 y_1^2 + x_2^2 y_2^2 + 2x_1 y_1 x_2 y_2 = \\ &= (x_1 y_1 + x_2 y_2)^2 = \\ &= (\mathbf{x}^T \mathbf{y})^2 \end{aligned}$$

príklad

$$\begin{aligned} K(\mathbf{x}_i, \mathbf{x}_j) &= (\mathbf{x}_i^T \mathbf{x}_j)^2 = \left((x_{i1} \quad x_{i2}) \begin{pmatrix} x_{j1} \\ x_{j2} \end{pmatrix} \right)^2 = \\ &= (x_{i1}x_{j1} + x_{i2}x_{j2})^2 = \\ &= x_{i1}^2x_{j1}^2 + 2x_{i1}x_{j1}x_{i2}x_{j2} + x_{i2}^2x_{j2}^2 \end{aligned}$$

$$\varphi(\mathbf{x}) = \begin{pmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{pmatrix}$$

$$\varphi(\mathbf{x}) = \begin{pmatrix} x_1^2 \\ x_1x_2 \\ x_1x_2 \\ x_2^2 \end{pmatrix}$$

$$\varphi(\mathbf{x}) = \frac{1}{\sqrt{2}} \begin{pmatrix} x_1^2 - x_2^2 \\ 2x_1x_2 \\ x_1^2 + x_2^2 \end{pmatrix}$$

Pre ľubovoľnú funkciu K , ako zistíme, či zodpovedá skalárnemu súčinu $\varphi(\mathbf{x}_1)^T \varphi(\mathbf{x}_2)$ v nejakom priestore?

Konštruujeme špeciálne funkcie – Mercerove jadrá:

Ak K je

Symetrická

$$K(\mathbf{x}_i, \mathbf{x}_j) = K(\mathbf{x}_j, \mathbf{x}_i)$$

Pozitívne semidefinitná Gramova matica

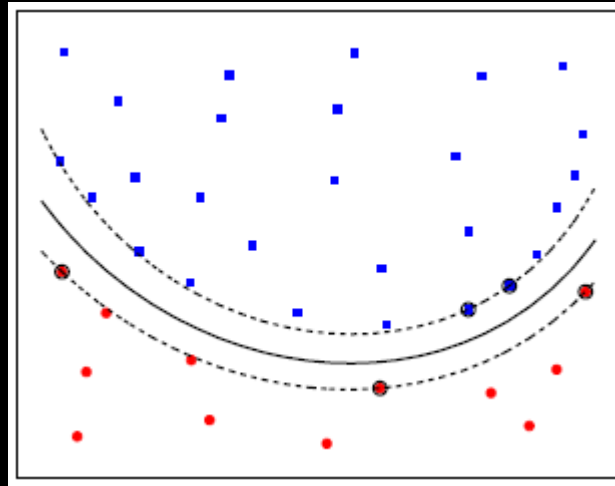
$$\forall \mathbf{a} \in \mathbb{R}^N, \mathbf{a}^T \mathbf{K} \mathbf{a} \geq 0,$$

$$k_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$$

Potom K je platné jadro

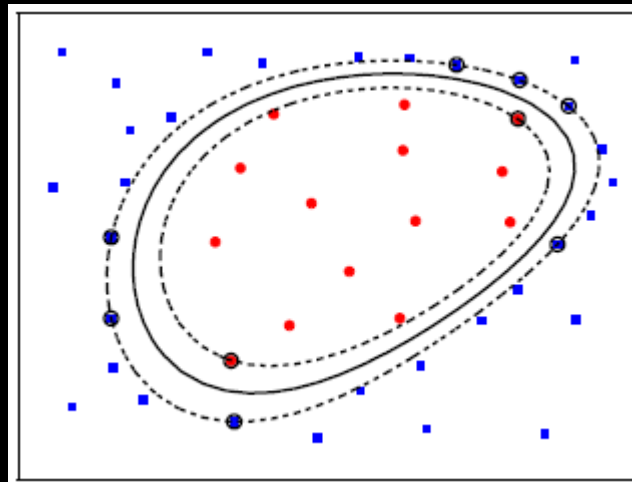
Polynomiálne jadrá

$$K(\mathbf{x}, \mathbf{y}) = (a + \mathbf{x}^T \mathbf{y})^p$$



Gaussovské jadrá

$$K(\mathbf{x}, \mathbf{y}) = e^{-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}}$$



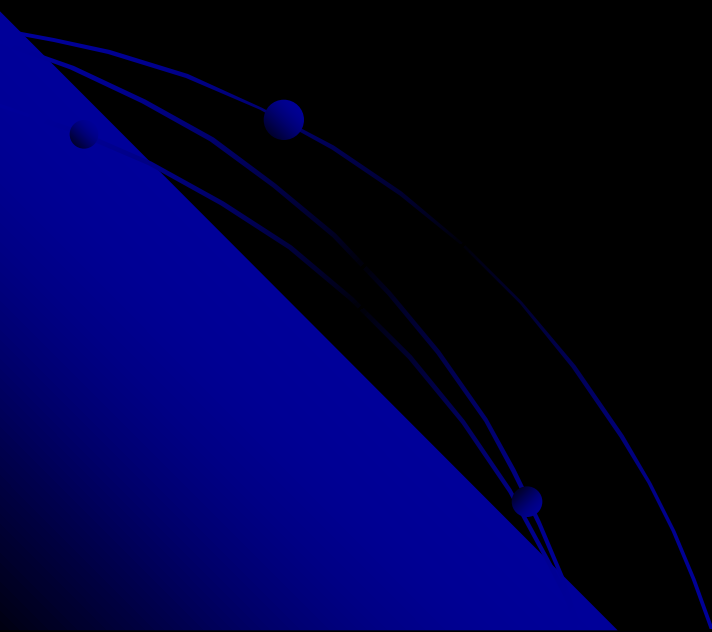
SVM aplikácie

Kategorizácia textu a hypertextu

Klasifikácia obrazov

Bioinformatika (klasifikácia proteínov, detekcia nádorov)

Rozpoznávanie písaného textu



A Tutorial on Support Vector Machines for Pattern Recognition

CHRISTOPHER J.C. BURGES

Bell Laboratories, Lucent Technologies

burges@lucent.com

IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 12, NO. 2, MARCH 2001

181

An Introduction to Kernel-Based Learning Algorithms

Klaus-Robert Müller, Sebastian Mika, Gunnar Rätsch, Koji Tsuda, and Bernhard Schölkopf

