Support Vector Machines
(jemný úvod)

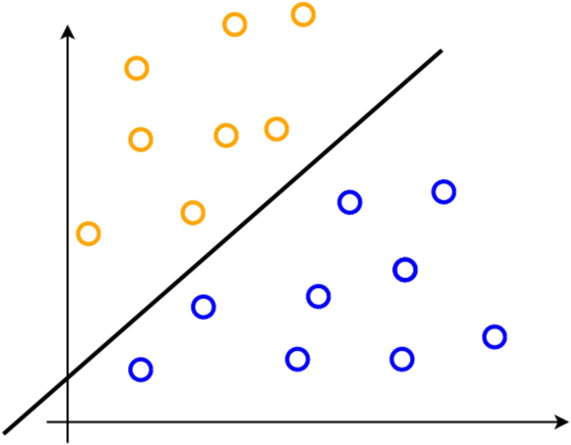
Osnova

- *Support Vector Classifier (SVC)*
- *Support Vector Machine (SVM)*
- “jádrový trik” (*kernel trick*)
- klasifikace s “měkkou hranicí” (*soft-margin classification*)
- hledání optimálních parametrů

Formulace úlohy

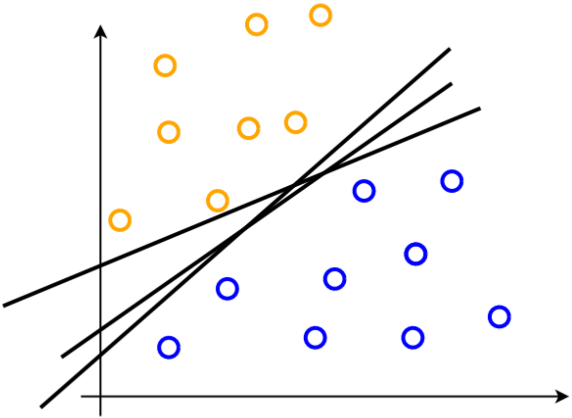
- dáno:
  - $\{x_i, y_i\}_{i=1}^{n}$
  - $x_i \in \mathbb{R}^p$ příznaky
  - $y_i \in \{-1,\ 1\}$ třídy
- úloha: klasifikovat nové $x \in \mathbb{R}^p$ do třídy -1 nebo 1

Motivace



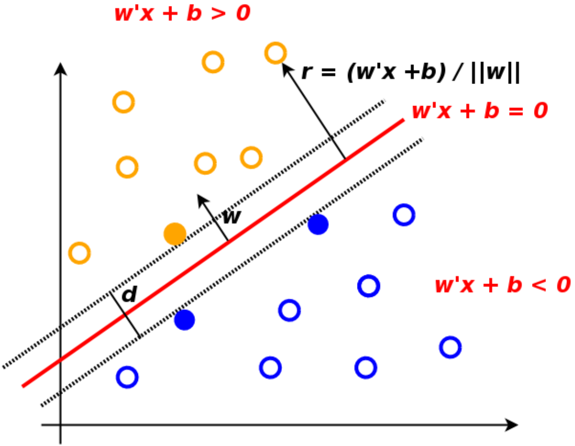
která přímka nejlépe odděluje instance daných dvou tříd?

Motivace



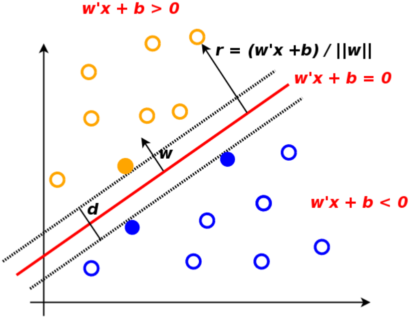
která přímka nejlépe odděluje instance daných dvou tříd?

Přístup *Support Vector Classifier (SVC)*



rozhodovací funkce $f(x) = sign(w'x + b)$

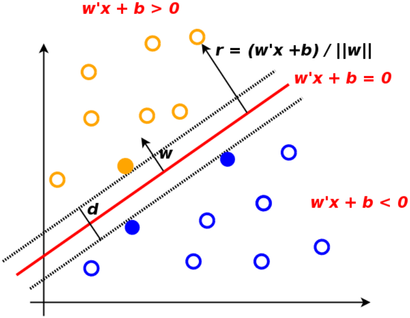
Přístup SVC (2)



myšlenka:

- maximalizace hranice o šířce $d = \frac{2}{\|w\|}$

Přístup SVC (2)



myšlenka:

- maximalizace hranice o šířce $d = \frac{2}{\|w\|}$
- za podmínky správné klasifikace:
  - $w' x_i + b \geq +1$, když $y_i = +1$
  - $w' x_i + b \leq -1$, když $y_i = -1$

Formulace optimalizační úlohy

- $argmax_{w,b} \left\{ \frac{2}{\|w\|} \right\}$
- za podmínky správné klasifikace:
  - $w'x_i + b \geq +1$, když $y_i = +1$
  - $w'x_i + b \leq -1$, když $y_i = -1$
  - nebo ekvivalentně: $y(w'x_i + b) \geq 1$

Formulace optimalizační úlohy

- $argmax_{w,b} \{\frac{2}{\|w\|}\}$
- za podmínky správné klasifikace:
  - $w'x_i + b \geq +1$, když $y_i = +1$
  - $w'x_i + b \leq -1$, když $y_i = -1$
  - nebo ekvivalentně: $y(w'x_i + b) \geq 1$

ekvivalentně (formulace 1):

- $argmin_{w,b} \{\frac{1}{2}\|w\|^2\}$
- za podmínky správné klasifikace: $y_i(w'x_i + b) \geq 1$

Formulace optimalizační úlohy

- $argmax_{w,b} \{\frac{2}{\|w\|}\}$
- za podmínky správné klasifikace:
  - $w'x_i + b \geq +1$, když $y_i = +1$
  - $w'x_i + b \leq -1$, když $y_i = -1$
  - nebo ekvivalentně: $y(w'x_i + b) \geq 1$

ekvivalentně (formulace 1):

- $argmin_{w,b} \{\frac{1}{2}\|w\|^2\}$
- za podmínky správné klasifikace: $y_i(w'x_i + b) \geq 1$

ekvivalentně (chce se zbavit omezení) (formulace 2):

- $L_p = -\frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i[(w'x_i + b)y_i - 1]$
- $L_p = -\frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i[(w'x_i + b)y_i] + \sum_{i=1}^{n} \alpha_i$
- $argmin_{w,b}\ L_p$
- za podmínky: $\alpha_i \geq 0$
- α_i - Lagrangeovy multiplikátory

Formulace optimalizační úlohy

Primární a duální úloha

$L_p = -\frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i[(w'x_i + b)y_i] + \sum_{i=1}^{n} \alpha_i$

$argmin_{w,b}\ L_p$ (za podmínky $\alpha_i \geq 0$)

Formulace optimalizační úlohy
Primární a duální úloha

$L_p = -\frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i[(w'x_i + b)y_i] + \sum_{i=1}^{n} \alpha_i$

$argmin_{w,b}\ L_p$ (za podmínky $\alpha_i \geq 0$)

Lze ukázat:

- $argmin_{w,b}\ L_p$ za podm. $C_1 \equiv argmax_{w,b}\ L_p$ za podm. C_2
- $C_1 : \frac{\partial L_p}{\alpha_i} = 0,\ \alpha_i \geq 0$ (odpovídá podm. správné klasifikace)
- $C_2 : \frac{\partial L_p}{w} = 0,\ \frac{\partial L_p}{b} = 0,\ \alpha_i \geq 0$

Formulace optimalizační úlohy

Primární a duální úloha

$L_p = -\frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i[(w'x_i + b)y_i] + \sum_{i=1}^{n} \alpha_i$

$argmin_{w,b}\, L_p$ (za podmínky $\alpha_i \geq 0$)

Lze ukázat:

- $argmin_{w,b}\, L_p$ za podm. $C_1 \equiv argmax_{w,b}\, L_p$ za podm. C_2
- $C_1 : \frac{\partial L_p}{\alpha_i} = 0,\ \alpha_i \geq 0$ (odpovídá podm. správné klasifikace)
- $C_2 : \frac{\partial L_p}{w} = 0,\ \frac{\partial L_p}{b} = 0,\ \alpha_i \geq 0$

Hledám extrém L_p:

- $\frac{\partial L_p}{\partial w} = w - \sum_{i=1}^{n} \alpha_i x_i y_i = 0$

Formulace optimalizační úlohy

Primární a duální úloha

$L_p = -\frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i[(w'x_i + b)y_i] + \sum_{i=1}^{n} \alpha_i$

$argmin_{w,b}\ L_p$ (za podmínky $\alpha_i \geq 0$)

Lze ukázat:

- $argmin_{w,b}\ L_p$ za podm. $C_1 \equiv argmax_{w,b}\ L_p$ za podm. C_2
- $C_1 : \frac{\partial L_p}{\alpha_i} = 0,\ \alpha_i \geq 0$ (odpovídá podm. správné klasifikace)
- $C_2 : \frac{\partial L_p}{w} = 0,\ \frac{\partial L_p}{b} = 0,\ \alpha_i \geq 0$

Hledám extrém L_p:

- $\frac{\partial L_p}{\partial w} = w - \sum_{i=1}^{n} \alpha_i x_i y_i = 0 \rightarrow w = \sum_{i=1}^{n} \alpha_i x_i y_i$

$L_p = -\frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i[(w'x_i + b)y_i] + \sum_{i=1}^{n} \alpha_i$

$argmin_{w,b}\, L_p$ (za podmínky $\alpha_i \geq 0$)

Lze ukázat:

- $argmin_{w,b}\, L_p$ za podm. $C_1 \equiv argmax_{w,b}\, L_p$ za podm. C_2
- $C_1 : \frac{\partial L_p}{\alpha_i} = 0, \ \alpha_i \geq 0$ (odpovídá podm. správné klasifikace)
- $C_2 : \frac{\partial L_p}{w} = 0, \ \frac{\partial L_p}{b} = 0, \ \alpha_i \geq 0$

Hledám extrém L_p:

- $\frac{\partial L_p}{\partial w} = w - \sum_{i=1}^{n} \alpha_i x_i y_i = 0 \rightarrow w = \sum_{i=1}^{n} \alpha_i x_i y_i$
- $\frac{\partial L_p}{\partial b} = \sum_{i=1}^{n} \alpha_i y_i = 0$

Formulace optimalizační úlohy
Primární a duální úloha

$L_p = -\frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i[(w'x_i + b)y_i] + \sum_{i=1}^{n} \alpha_i$
$argmin_{w,b}\ L_p$ (za podmínky $\alpha_i \geq 0$)
Lze ukázat:

- $argmin_{w,b}\ L_p$ za podm. $C_1 \equiv argmax_{w,b}\ L_p$ za podm. C_2
- $C_1 : \frac{\partial L_p}{\alpha_i} = 0,\ \alpha_i \geq 0$ (odpovídá podm. správné klasifikace)
- $C_2 : \frac{\partial L_p}{w} = 0,\ \frac{\partial L_p}{b} = 0,\ \alpha_i \geq 0$

Hledám extrém L_p:

- $\frac{\partial L_p}{\partial w} = w - \sum_{i=1}^{n} \alpha_i x_i y_i = 0 \rightarrow w = \sum_{i=1}^{n} \alpha_i x_i y_i$
- $\frac{\partial L_p}{\partial b} = \sum_{i=1}^{n} \alpha_i y_i = 0$

Substitucí do L_p získáme duální úlohu:
$L_d = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i\alpha_j x_i x_j y_i y_j$
$argmax_{\alpha_i} L_d$ (za podmínky $\alpha_i \geq 0$)

Optimalizační úlohy
Rekapitulace

- formulace 1:
  - $argmin_{w,b} \left\{\frac{1}{2}\|w\|^2\right\}$ (za podm.: $y_i(w'x_i + b) \geq 1$)

- formulace 1:
  - $argmin_{w,b} \{\frac{1}{2}\|w\|^2\}$ (za podm.: $y_i(w'x_i + b) \geq 1$)
  - $p + 1$ parametrů, n lin. omezení

Optimalizační úlohy
Rekapitulace

- formulace 1:
  - $argmin_{w,b} \{\frac{1}{2}\|w\|^2\}$ (za podm.: $y_i(w'x_i + b) \geq 1$)
  - $p + 1$ parametrů, n lin. omezení
- formulace 2:
  - $L_p = -\frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i[(w'x_i + b)y_i] + \sum_{i=1}^{n} \alpha_i$
  - $argmin_{w,b}\ L_p$ (za podmínky $\alpha_i \geq 0$)

- formulace 1:
  - $argmin_{w,b}\, \{\frac{1}{2}\|w\|^2\}$ (za podm.: $y_i(w'x_i + b) \geq 1$)
  - $p + 1$ parametrů, n lin. omezení
- formulace 2:
  - $L_p = -\frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i[(w'x_i + b)y_i] + \sum_{i=1}^{n} \alpha_i$
  - $argmin_{w,b}\, L_p$ (za podmínky $\alpha_i \geq 0$)
  - $p + 1$ parametrů, n omezení

Optimalizační úlohy

Rekapitulace

- formulace 1:
  - $argmin_{w,b} \{\frac{1}{2}\|w\|^2\}$ (za podm.: $y_i(w'x_i + b) \geq 1$)
  - $p + 1$ parametrů, n lin. omezení
- formulace 2:
  - $L_p = -\frac{1}{2}\|w\|^2 - \sum_{i=1}^n \alpha_i[(w'x_i + b)y_i] + \sum_{i=1}^n \alpha_i$
  - $argmin_{w,b} L_p$ (za podmínky $\alpha_i \geq 0$)
  - $p + 1$ parametrů, n omezení
- formulace 3:
  - $L_d = \sum_{i=1}^n \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i\alpha_j x_i x_j y_i y_j$
  - $argmax_{\alpha_i} L_d$ (za podmínky $\alpha_i \geq 0$)

Optimalizační úlohy
Rekapitulace

- formulace 1:
  - $argmin_{w,b} \{\frac{1}{2}\|w\|^2\}$ (za podm.: $y_i(w'x_i + b) \geq 1$)
  - $p + 1$ parametrů, n lin. omezení
- formulace 2:
  - $L_p = -\frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i[(w'x_i + b)y_i] + \sum_{i=1}^{n} \alpha_i$
  - $argmin_{w,b} L_p$ (za podmínky $\alpha_i \geq 0$)
  - $p + 1$ parametrů, n omezení
- formulace 3:
  - $L_d = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j x_i x_j y_i y_j$
  - $argmax_{\alpha_i} L_d$ (za podmínky $\alpha_i \geq 0$)
  - n parametrů, n omezení

Optimalizační úlohy

Rekapitulace

- formulace 1:
  - $argmin_{w,b} \{\frac{1}{2}\|w\|^2\}$ (za podm.: $y_i(w'x_i + b) \geq 1$)
  - $p+1$ parametrů, n lin. omezení
- formulace 2:
  - $L_p = -\frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i[(w'x_i + b)y_i] + \sum_{i=1}^{n} \alpha_i$
  - $argmin_{w,b}\, L_p$ (za podmínky $\alpha_i \geq 0$)
  - $p+1$ parametrů, n omezení
- formulace 3:
  - $L_d = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j x_i x_j y_i y_j$
  - $argmax_{\alpha_i} L_d$ (za podmínky $\alpha_i \geq 0$)
  - n parametrů, n omezení
  - data x_i vystupují pouze ve formě součinů $x_i x_j$

Optimalizační úlohy
Rekapitulace

- formulace 1:
  - $argmin_{w,b} \{\frac{1}{2}\|w\|^2\}$ (za podm.: $y_i(w'x_i + b) \geq 1$)
  - $p + 1$ parametrů, n lin. omezení
- formulace 2:
  - $L_p = -\frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i[(w'x_i + b)y_i] + \sum_{i=1}^{n} \alpha_i$
  - $argmin_{w,b}\, L_p$ (za podmínky $\alpha_i \geq 0$)
  - $p + 1$ parametrů, n omezení
- formulace 3:
  - $L_d = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i\alpha_j x_i x_j y_i y_j$
  - $argmax_{\alpha_i} L_d$ (za podmínky $\alpha_i \geq 0$)
  - n parametrů, n omezení
  - data x_i vystupují pouze ve formě součinů $x_i x_j$
  - většina α_i nulových, $\alpha_i = 1 > 0$ právě pro *support vectors*

Řešení optimalizační úlohy

Rozhodovací funkce $f(x) = sign(w'x + b)$
$w = \sum_{i=1}^{n} \alpha_i x_i y_i$

Řešení optimalizační úlohy

Rozhodovací funkce $f(x) = sign(w'x + b)$

$w = \sum_{i=1}^{n} \alpha_i x_i y_i$

jak získat b?

Řešení optimalizační úlohy

Rozhodovací funkce $f(x) = sign(w'x + b)$
$w = \sum_{i=1}^{n} \alpha_i x_i y_i$
jak získat b?

- pro lib. *support vector*: $y_i(w'x_i + b) = 1$
- $\rightarrow b = \frac{1}{y_i} - w'x_i$

Řešení optimalizační úlohy

Rozhodovací funkce $f(x) = sign(w'x + b)$
$w = \sum_{i=1}^{n} \alpha_i x_i y_i$
jak získat b?

- pro lib. *support vector*: $y_i(w'x_i + b) = 1$
- $\rightarrow b = \frac{1}{y_i} - w'x_i$
- prakticky: $b = \frac{\sum_{i,\alpha_i>0}(\frac{1}{y_i} - w'x_i)}{\sum \alpha_i}$

Řešení optimalizační úlohy

Rozhodovací funkce $f(x) = sign(w'x + b)$
$w = \sum_{i=1}^{n} \alpha_i x_i y_i$
jak získat b?

- pro lib. *support vector*: $y_i(w'x_i + b) = 1$
- $\rightarrow b = \frac{1}{y_i} - w'x_i$
- prakticky: $b = \frac{\sum_{i,\alpha_i>0}(\frac{1}{y_i} - w'x_i)}{\sum \alpha_i}$

tedy konečně dostáváme:

- $f(x) = sign(\sum_{i=1}^{n} \alpha_i y_i x_i' x + b)$

Trik s jádrem *(Kernel trick)*

Rozhodovací funkce:

- $f(x) = sign(\sum_{i=1}^{n} \alpha_i y_i x_i' x + b)$

Trik s jádrem *(Kernel trick)*

Rozhodovací funkce:

- $f(x) = sign(\sum_{i=1}^{n} \alpha_i y_i x_i' x + b)$

Pozorování:

- příznaky x se vyskytují pouze ve formě skalárním součinu: $x_i' x$

Trik s jádrem *(Kernel trick)*

Rozhodovací funkce:

- $f(x) = sign(\sum_{i=1}^{n} \alpha_i y_i x_i' x + b)$

Pozorování:

- příznaky x se vyskytují pouze ve formě skalárním součinu: $x_i' x$
- skalární součin $x_1' x_2$ lze nahradit jádrem $K(x_1, x_2)$

Trik s jádrem *(Kernel trick)*

Rozhodovací funkce:

- $f(x) = sign(\sum_{i=1}^{n} \alpha_i y_i x_i' x + b)$

Pozorování:

- příznaky x se vyskytují pouze ve formě skalárním součinu: $x_i' x$
- skalární součin $x_1' x_2$ lze nahradit jádrem $K(x_1, x_2)$
- $K(x_1, x_2) = \phi(x_1)' \phi(x_2)$

Trik s jádrem *(Kernel trick)*

Rozhodovací funkce:

- $f(x) = sign(\sum_{i=1}^{n} \alpha_i y_i x_i' x + b)$

Pozorování:

- příznaky x se vyskytují pouze ve formě skalárním součinu: $x_i' x$
- skalární součin $x_1' x_2$ lze nahradit jádrem $K(x_1, x_2)$
- $K(x_1, x_2) = \phi(x_1)' \phi(x_2)$
- jádro může realizovat operaci odpovídající skalárnímu součinu ve vysokorozměrném prostoru

Trik s jádrem *(Kernel trick)*

Rozhodovací funkce:

- $f(x) = sign(\sum_{i=1}^{n} \alpha_i y_i x_i' x + b)$

Pozorování:

- příznaky x se vyskytují pouze ve formě skalárním součinu: $x_i' x$
- skalární součin $x_1' x_2$ lze nahradit jádrem $K(x_1, x_2)$
- $K(x_1, x_2) = \phi(x_1)' \phi(x_2)$
- jádro může realizovat operaci odpovídající skalárnímu součinu ve vysokorozměrném prostoru
- použitím jádra se z SVC stávají SVM

Trik s jádrem *(Kernel trick)*

Rozhodovací funkce:

- $f(x) = sign(\sum_{i=1}^{n} \alpha_i y_i x_i' x + b)$

Pozorování:

- příznaky x se vyskytují pouze ve formě skalárním součinu: $x_i' x$
- skalární součin $x_1' x_2$ lze nahradit jádrem $K(x_1, x_2)$
- $K(x_1, x_2) = \phi(x_1)' \phi(x_2)$
- jádro může realizovat operaci odpovídající skalárnímu součinu ve vysokorozměrném prostoru
- použitím jádra se z SVC stávají SVM

Používaná jádra:

- polynomiální: $K(x_1, x_2) = (x_1' x_2 + 1)^q$
- *Gaussian radial-basis function (RBF)*: $K(x_1, x_2) = exp\{\frac{\|x_1 - x_2\|^2}{2\sigma^2}\}$
- hyperbolický tangens: $K(x_1, x_2) = tanh\{\beta_1 x_1' x_2 + \beta_2\}$

Polynomiální jádro stupně $q = 2$:
$K(x, y) = (x'y + 1)^2, x, y \in \mathbb{R}^2$ operuje v prostoru $\mathbb{R}^6$:

Trik s jádrem *(Kernel trick)*
Příklad

Polynomiální jádro stupně $q = 2$:
$K(x, y) = (x'y + 1)^2, x, y \in \mathbb{R}^2$ operuje v prostoru $\mathbb{R}^6$:
$x \rightarrow \phi(x) = \{x_1^2,\ x_2^2,\ \sqrt{2}x_1x_2,\ \sqrt{2}x_1,\ \sqrt{2}x_2,\ 1\}$

Trik s jádrem *(Kernel trick)*
Příklad

Polynomiální jádro stupně $q = 2$:
$K(x, y) = (x'y + 1)^2, x, y \in \mathbb{R}^2$ operuje v prostoru $\mathbb{R}^6$:
$x \rightarrow \phi(x) = \{x_1^2,\ x_2^2,\ \sqrt{2}x_1 x_2,\ \sqrt{2}x_1,\ \sqrt{2}x_2,\ 1\}$
$y \rightarrow \phi(y) = \{y_1^2,\ y_2^2,\ \sqrt{2}y_1 y_2,\ \sqrt{2}y_1,\ \sqrt{2}y_2,\ 1\}$

Polynomiální jádro stupně $q = 2$:
$K(x, y) = (x'y + 1)^2, x, y \in \mathbb{R}^2$ operuje v prostoru $\mathbb{R}^6$:
$x \rightarrow \phi(x) = \{x_1^2,\ x_2^2,\ \sqrt{2}x_1x_2,\ \sqrt{2}x_1,\ \sqrt{2}x_2,\ 1\}$
$y \rightarrow \phi(y) = \{y_1^2,\ y_2^2,\ \sqrt{2}y_1y_2,\ \sqrt{2}y_1,\ \sqrt{2}y_2,\ 1\}$

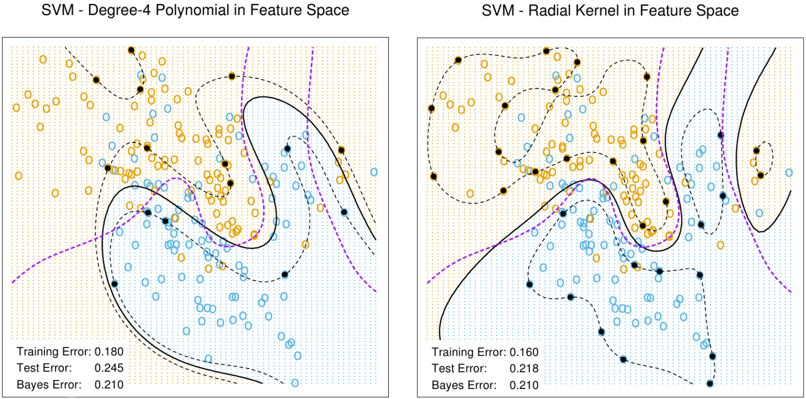
$$\phi(x)'\phi(y) = x_1^2y_1^2 + x_2^2y_2^2 + 2x_1x_2y_1y_2 + 2x_1y_1 + 2x_2y_2 + 1$$

Polynomiální jádro stupně $q = 2$:
$K(x, y) = (x'y + 1)^2, x, y \in \mathbb{R}^2$ operuje v prostoru $\mathbb{R}^6$:
$x \rightarrow \phi(x) = \{x_1^2,\ x_2^2,\ \sqrt{2}x_1x_2,\ \sqrt{2}x_1,\ \sqrt{2}x_2,\ 1\}$
$y \rightarrow \phi(y) = \{y_1^2,\ y_2^2,\ \sqrt{2}y_1y_2,\ \sqrt{2}y_1,\ \sqrt{2}y_2,\ 1\}$

$$\phi(x)'\phi(y) = x_1^2y_1^2 + x_2^2y_2^2 + 2x_1x_2y_1y_2 + 2x_1y_1 + 2x_2y_2 + 1$$

$(x'y + 1)^2 = (x_1y_1 + x_2y_2 + 1)^2 =$
$x_1^2y_1^2 + x_2^2y_2^2 + 1 + 2x_1x_2y_1y_2 + 2x_1y_1 + 2x_2y_2$

Ukázka jader



Klasifikace s “měkkou hranicí”
Soft-margin classification

Motivace:

- třídy nemusejí být oddělitelné
- přesto chceme SVM použít

Řešení:

- dovolit SVM udělat “malou” chybu

Jak:

- $argmin_{w,b} \{\frac{1}{2}\|w\|^2\} + C\sum_{i=1}^{n} \xi_i$
- za podm. téměř správné klasifikace: $y_i(w'x_i + b) \geq 1 - \xi_i$
- C představuje regularizační konstantu
- $C = \infty$ odpovídá původní formulaci separabilní úlohy
- C se hledá nejčastěji pomocí křížové validace

SVM v praxi

- fungují velmi dobře
- časová složitost trénování: $O(n^2)$
- uživatel volí typ jádra a parametry
- parametry se hledají typicky křížovou validací
- po natrénování si stačí pamatovat *support vectors*

Klasifikace do více tříd

SVM umí rozlišovat jen do dvou tříd

možná řešení klasifikace do K tříd:

- "jeden proti všem": K úloh: klasifikace třídy k proti zbytku, "vítěz bere vše"
- "jeden na jednoho": $\frac{K(K-1)}{2}$ úloh: klasifikace třídy k_1 proti k_2, hlasování

Rozšíření SVM

- *SVM regression*
- detekce nečekaných pozorování *novelty detection*

Literatura

Christopher J.C. Burges: A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery (1998), volume 2, p.121-167.

Hastie T., Tibshirani R., Friedman J.: The Elements of Statistical Learning, Springer New York Inc., 2001, New York, NY, USA, *http://statweb.stanford.edu/ tibs/ElemStatLearn/*