

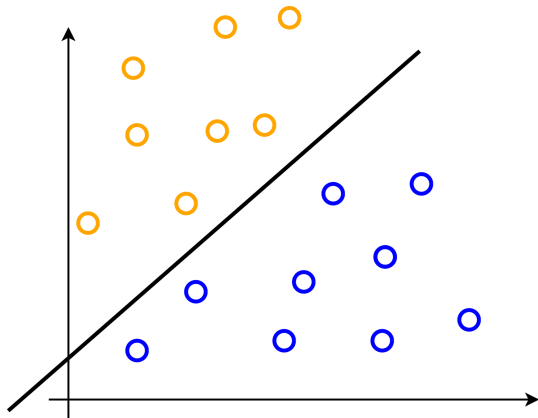
Support Vector Machines (úvod)

- formulace úlohy
- *Support Vector Classifier (SVC)*
 - matematické odvození 💡
- “jaderný trik” (*kernel trick*)
- *Support Vector Machine*
 - $SVM = SVC + kernel$
- hledání parametrů
- rošíření
 - klasifikace s “měkkou hranicí” (*soft-margin classification*)
 - klasifikace do více tříd
 - detekce nečekaných pozorování *novelty detection*
 - shlukování založené na SVM

Formulace úlohy

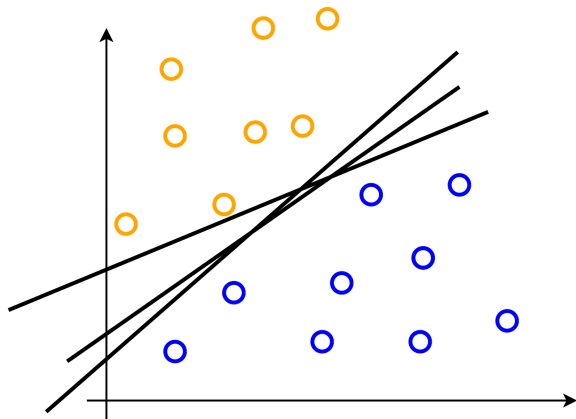
- dáno:
 - $\{\mathbf{x}_i, y_i\}_{i=1}^n$
 - $\mathbf{x}_i \in \mathbb{R}^p$ příznaky
 - $y_i \in \{-1, 1\}$ třídy
- úloha: klasifikovat nové $\mathbf{x} \in \mathbb{R}^p$ do třídy -1 nebo 1

Motivace



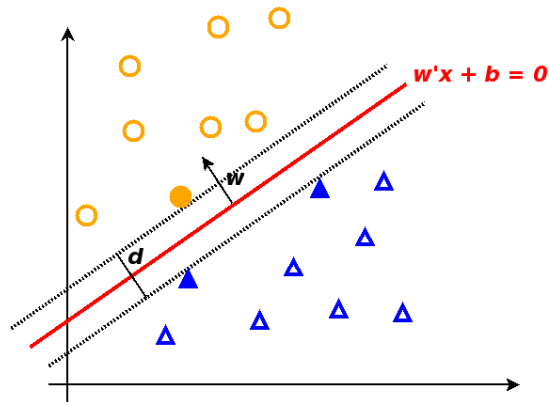
kteřá přímka nejlépe odděluje instance daných dvou tříd?

Motivace



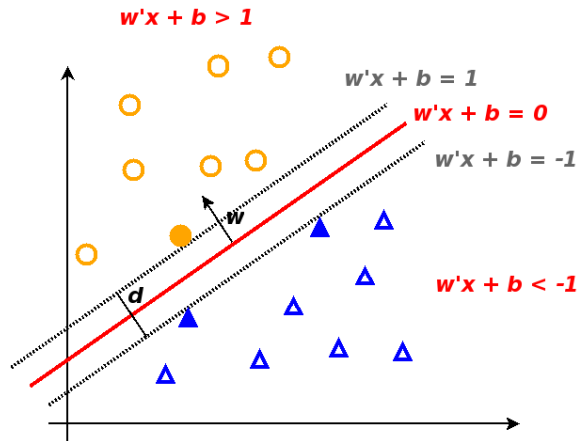
kteřá pŕímka nejlépe odděluje instance daných dvou tříd?

Přístup *Support Vector Classifier (SVC)*



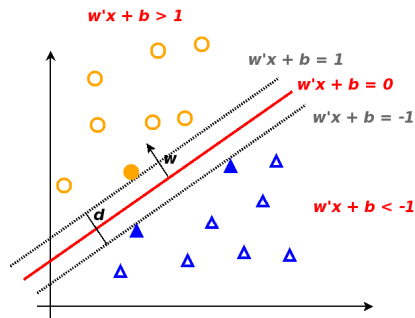
rozhodovací funkce $f(x) = \text{sign}(w'x + b)$

Přístup *Support Vector Classifier (SVC)*



rozhodovací funkce $f(x) = \text{sign}(w'x + b)$

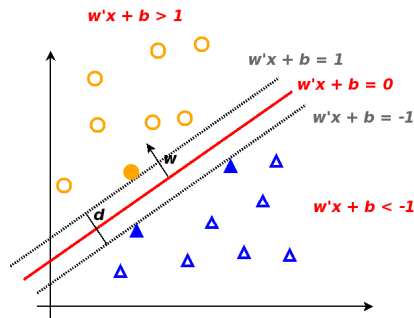
Přístup SVC (2)



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

- $\operatorname{argmax}_{\mathbf{w}, b} \left\{ \frac{2}{\|\mathbf{w}\|} \right\}$
- za podmínky správné klasifikace:
 - $\mathbf{w}'\mathbf{x}_i + b \geq +1$, když $y_i = +1$
 - $\mathbf{w}'\mathbf{x}_i + b \leq -1$, když $y_i = -1$
 - nebo ekvivalentně: $y_i(\mathbf{w}'\mathbf{x}_i + b) \geq 1$

Formulace optimalizační úlohy

- $\operatorname{argmax}_{\mathbf{w}, b} \left\{ \frac{2}{\|\mathbf{w}\|} \right\}$
- za podmínky správné klasifikace:
 - $\mathbf{w}'\mathbf{x}_i + b \geq +1$, když $y_i = +1$
 - $\mathbf{w}'\mathbf{x}_i + b \leq -1$, když $y_i = -1$
 - nebo ekvivalentně: $y_i(\mathbf{w}'\mathbf{x}_i + b) \geq 1$

ekvivalentně (formulace 1):

- $\operatorname{argmin}_{\mathbf{w}, b} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 \right\}$
- za podmínky správné klasifikace: $y_i(\mathbf{w}'\mathbf{x}_i + b) \geq 1$

Formulace optimalizační úlohy

- $\operatorname{argmax}_{\mathbf{w}, b} \left\{ \frac{2}{\|\mathbf{w}\|} \right\}$
- za podmínky správné klasifikace:
 - $\mathbf{w}'\mathbf{x}_i + b \geq +1$, když $y_i = +1$
 - $\mathbf{w}'\mathbf{x}_i + b \leq -1$, když $y_i = -1$
 - nebo ekvivalentně: $y_i(\mathbf{w}'\mathbf{x}_i + b) \geq 1$

ekvivalentně (formulace 1):

- $\operatorname{argmin}_{\mathbf{w}, b} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 \right\}$
- za podmínky správné klasifikace: $y_i(\mathbf{w}'\mathbf{x}_i + b) \geq 1$

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

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

Formulace optimalizační úlohy

Primární a duální úloha

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

*argmin*_{w,b} L_p (za podmínky $\alpha_j \geq 0$)

Formulace optimalizační úlohy

Primární a duální úloha

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

argmin _{\mathbf{w}, b} L_p (za podmínky $\alpha_j \geq 0$)

Požadujeme $\nabla L_p = 0$:

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

Formulace optimalizační úlohy

Primární a duální úloha

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

argmin _{\mathbf{w}, b} L_p (za podmínky $\alpha_j \geq 0$)

Požadujeme $\nabla L_p = 0$:

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

Formulace optimalizační úlohy

Primární a duální úloha

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

argmin _{\mathbf{w}, b} L_p (za podmínky $\alpha_j \geq 0$)

Požadujeme $\nabla L_p = 0$:

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

Formulace optimalizační úlohy

Primární a duální úloha

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

argmin _{\mathbf{w}, b} L_p (za podmínky $\alpha_j \geq 0$)

Požadujeme $\nabla L_p = 0$:

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

Substitucí do L_p získáme duální úlohu:

$$L_d = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j \mathbf{x}_i \mathbf{x}_j y_i y_j$$

argmax _{α_j} L_d (za podmínky $\alpha_j \geq 0$)

Optimalizační úlohy

Rekapitulace

- formulace 1:

- $\operatorname{argmin}_{\mathbf{w}, b} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 \right\}$ (za podm.: $y_i(\mathbf{w}'\mathbf{x}_i + b) \geq 1$)

Optimalizační úlohy

Rekapitulace

- formulace 1:
 - $\operatorname{argmin}_{\mathbf{w}, b} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 \right\}$ (za podm.: $y_i(\mathbf{w}'\mathbf{x}_i + b) \geq 1$)
 - $p + 1$ parametrů, n lin. omezení

Optimalizační úlohy

Rekapitulace

- formulace 1:
 - $\operatorname{argmin}_{\mathbf{w}, b} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 \right\}$ (za podm.: $y_i(\mathbf{w}'\mathbf{x}_i + b) \geq 1$)
 - $p + 1$ parametrů, n lin. omezení
- formulace 2:
 - $L_p = -\frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i [(\mathbf{w}'\mathbf{x}_i + b)y_i] + \sum_{i=1}^n \alpha_i$
 - $\operatorname{argmin}_{\mathbf{w}, b} L_p$ (za podmínky $\alpha_i \geq 0$)

Optimalizační úlohy

Rekapitulace

- formulace 1:
 - $\operatorname{argmin}_{\mathbf{w},b} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 \right\}$ (za podm.: $y_i(\mathbf{w}'\mathbf{x}_i + b) \geq 1$)
 - $p + 1$ parametrů, n lin. omezení
- formulace 2:
 - $L_p = -\frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i [(\mathbf{w}'\mathbf{x}_i + b)y_i] + \sum_{i=1}^n \alpha_i$
 - $\operatorname{argmin}_{\mathbf{w},b} L_p$ (za podmínky $\alpha_i \geq 0$)
 - $p + 1$ parametrů, n omezení

Optimalizační úlohy

Rekapitulace

- formulace 1:
 - $\operatorname{argmin}_{\mathbf{w}, b} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 \right\}$ (za podm.: $y_i(\mathbf{w}'\mathbf{x}_i + b) \geq 1$)
 - $p + 1$ parametrů, n lin. omezení
- formulace 2:
 - $L_p = -\frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i [(\mathbf{w}'\mathbf{x}_i + b)y_i] + \sum_{i=1}^n \alpha_i$
 - $\operatorname{argmin}_{\mathbf{w}, b} L_p$ (za podmínky $\alpha_i \geq 0$)
 - $p + 1$ parametrů, n omezení
- formulace 3:
 - $L_d = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j \mathbf{x}_i \mathbf{x}_j y_i y_j$
 - $\operatorname{argmax}_{\alpha_i} L_d$ (za podmínky $\alpha_i \geq 0$)

Optimalizační úlohy

Rekapitulace

- formulace 1:
 - $\operatorname{argmin}_{\mathbf{w}, b} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 \right\}$ (za podm.: $y_i(\mathbf{w}'\mathbf{x}_i + b) \geq 1$)
 - $p + 1$ parametrů, n lin. omezení
- formulace 2:
 - $L_p = -\frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i [(\mathbf{w}'\mathbf{x}_i + b)y_i] + \sum_{i=1}^n \alpha_i$
 - $\operatorname{argmin}_{\mathbf{w}, b} L_p$ (za podmínky $\alpha_i \geq 0$)
 - $p + 1$ parametrů, n omezení
- formulace 3:
 - $L_d = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j \mathbf{x}_i' \mathbf{x}_j y_i y_j$
 - $\operatorname{argmax}_{\alpha_i} L_d$ (za podmínky $\alpha_i \geq 0$)
 - n parametrů, n omezení

Optimalizační úlohy

Rekapitulace

- formulace 1:
 - $\operatorname{argmin}_{\mathbf{w}, b} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 \right\}$ (za podm.: $y_i(\mathbf{w}'\mathbf{x}_i + b) \geq 1$)
 - $p + 1$ parametrů, n lin. omezení
- formulace 2:
 - $L_p = -\frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i [(\mathbf{w}'\mathbf{x}_i + b)y_i] + \sum_{i=1}^n \alpha_i$
 - $\operatorname{argmin}_{\mathbf{w}, b} L_p$ (za podmínky $\alpha_i \geq 0$)
 - $p + 1$ parametrů, n omezení
- formulace 3:
 - $L_d = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j \mathbf{x}_i \mathbf{x}_j y_i y_j$
 - $\operatorname{argmax}_{\alpha_i} L_d$ (za podmínky $\alpha_i \geq 0$)
 - n parametrů, n omezení
 - **proč to všechno snažení?**

Optimalizační úlohy

Rekapitulace

- formulace 1:
 - $\operatorname{argmin}_{\mathbf{w}, b} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 \right\}$ (za podm.: $y_i(\mathbf{w}'\mathbf{x}_i + b) \geq 1$)
 - $p + 1$ parametrů, n lin. omezení
- formulace 2:
 - $L_p = -\frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i [(\mathbf{w}'\mathbf{x}_i + b)y_i] + \sum_{i=1}^n \alpha_i$
 - $\operatorname{argmin}_{\mathbf{w}, b} L_p$ (za podmínky $\alpha_i \geq 0$)
 - $p + 1$ parametrů, n omezení
- formulace 3:
 - $L_d = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j \mathbf{x}_i \mathbf{x}_j y_i y_j$
 - $\operatorname{argmax}_{\alpha_i} L_d$ (za podmínky $\alpha_i \geq 0$)
 - n parametrů, n omezení
 - **proč to všechno snažení?**
 - **data \mathbf{x}_i vystupují pouze ve formě součinů $\mathbf{x}_i \mathbf{x}_j$**

Optimalizační úlohy

Rekapitulace

- formulace 1:
 - $\operatorname{argmin}_{\mathbf{w}, b} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 \right\}$ (za podm.: $y_i(\mathbf{w}'\mathbf{x}_i + b) \geq 1$)
 - $p + 1$ parametrů, n lin. omezení
- formulace 2:
 - $L_p = -\frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i [(\mathbf{w}'\mathbf{x}_i + b)y_i] + \sum_{i=1}^n \alpha_i$
 - $\operatorname{argmin}_{\mathbf{w}, b} L_p$ (za podmínky $\alpha_i \geq 0$)
 - $p + 1$ parametrů, n omezení
- formulace 3:
 - $L_d = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j \mathbf{x}_i \mathbf{x}_j y_i y_j$
 - $\operatorname{argmax}_{\alpha_i} L_d$ (za podmínky $\alpha_i \geq 0$)
 - n parametrů, n omezení
 - **proč to všechno snažení?**
 - **data \mathbf{x}_i vystupují pouze ve formě součinů $\mathbf{x}_i \mathbf{x}_j$**
 - většina α_i nulových, $\alpha_i > 0$ právě pro *support vectors*

Řešení optimalizační úlohy

Rozhodovací funkce $f(\mathbf{x}) = \text{sign}(\mathbf{w}'\mathbf{x} + b)$

$$\mathbf{w} = \sum_{i=1}^n \alpha_i \mathbf{x}_i y_i$$

Řešení optimalizační úlohy

Rozhodovací funkce $f(\mathbf{x}) = \text{sign}(\mathbf{w}'\mathbf{x} + b)$

$$\mathbf{w} = \sum_{i=1}^n \alpha_i \mathbf{x}_i y_i$$

jak získat b ?

Řešení optimalizační úlohy

Rozhodovací funkce $f(\mathbf{x}) = \text{sign}(\mathbf{w}'\mathbf{x} + b)$

$$\mathbf{w} = \sum_{i=1}^n \alpha_i \mathbf{x}_i y_i$$

jak získat b ?

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

Řešení optimalizační úlohy

Rozhodovací funkce $f(\mathbf{x}) = \text{sign}(\mathbf{w}'\mathbf{x} + b)$

$$\mathbf{w} = \sum_{i=1}^n \alpha_i \mathbf{x}_i y_i$$

jak získat b ?

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

Řešení optimalizační úlohy

Rozhodovací funkce $f(\mathbf{x}) = \text{sign}(\mathbf{w}'\mathbf{x} + b)$

$$\mathbf{w} = \sum_{i=1}^n \alpha_i \mathbf{x}_i y_i$$

jak získat b ?

- pro lib. *support vector*: $y_i(\mathbf{w}'\mathbf{x}_i + b) = 1$

- $\rightarrow b = \frac{1}{y_i} - \mathbf{w}'\mathbf{x}_i$

- prakticky: $b = \frac{\sum_{i, \alpha_i > 0} (\frac{1}{y_i} - \mathbf{w}'\mathbf{x}_i)}{\sum_{i, \alpha_i > 0} 1}$

tedy konečně dostáváme:

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

Trik s jádrem (*Kernel trick*)

Rozhodovací funkce:

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

Trik s jádrem (*Kernel trick*)

Rozhodovací funkce:

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

Pozorování:

- příznaky \mathbf{x} pouze ve formě skalárním součinu $\mathbf{x}_i' \mathbf{x}$

Trik s jádrem (*Kernel trick*)

Rozhodovací funkce:

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

Pozorování:

- příznaky \mathbf{x} pouze ve formě skalárním součinu $\mathbf{x}_i' \mathbf{x}$
- skalární součin $\mathbf{r}' \mathbf{s}$ lze nahradit jádrem $K(\mathbf{r}, \mathbf{s})$

Trik s jádrem (*Kernel trick*)

Rozhodovací funkce:

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

Pozorování:

- příznaky \mathbf{x} pouze ve formě skalárním součinu $\mathbf{x}_i' \mathbf{x}$
- skalární součin $\mathbf{r}' \mathbf{s}$ lze nahradit jádrem $K(\mathbf{r}, \mathbf{s})$
- $K(\mathbf{r}, \mathbf{s}) = \phi(\mathbf{r})' \phi(\mathbf{s})$

Trik s jádrem (*Kernel trick*)

Rozhodovací funkce:

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

Pozorování:

- příznaky \mathbf{x} pouze ve formě skalárním součinu $\mathbf{x}_i' \mathbf{x}$
- skalární součin $\mathbf{r}' \mathbf{s}$ lze nahradit jádrem $K(\mathbf{r}, \mathbf{s})$
- $K(\mathbf{r}, \mathbf{s}) = \phi(\mathbf{r})' \phi(\mathbf{s})$
- jádro může realizovat operaci odpovídající skalárnímu součinu ve vysokorozměrném prostoru

Trik s jádrem (*Kernel trick*)

Rozhodovací funkce:

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

Pozorování:

- příznaky \mathbf{x} pouze ve formě skalárním součinu $\mathbf{x}_i' \mathbf{x}$
- skalární součin $\mathbf{r}' \mathbf{s}$ lze nahradit jádrem $K(\mathbf{r}, \mathbf{s})$
- $K(\mathbf{r}, \mathbf{s}) = \phi(\mathbf{r})' \phi(\mathbf{s})$
- jádro může realizovat operaci odpovídající skalárnímu součinu ve vysokorozměrném prostoru
- použitím jádra se z SVC stávají SVM

Trik s jádrem (*Kernel trick*)

Rozhodovací funkce:

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

Pozorování:

- příznaky \mathbf{x} pouze ve formě skalárním součinu $\mathbf{x}_i' \mathbf{x}$
- skalární součin $\mathbf{r}' \mathbf{s}$ lze nahradit jádrem $K(\mathbf{r}, \mathbf{s})$
- $K(\mathbf{r}, \mathbf{s}) = \phi(\mathbf{r})' \phi(\mathbf{s})$
- jádro může realizovat operaci odpovídající skalárnímu součinu ve vysokorozměrném prostoru
- použitím jádra se z SVC stávají SVM

Používaná jádra:

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

Trik s jádrem (*Kernel trick*)

Příklad

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

Trik s jádrem (*Kernel trick*)

Příklad

Polynomiální jádro stupně $q = 2$: $K(\mathbf{r}, \mathbf{s}) = (\mathbf{r}'\mathbf{s} + 1)^2$,

$\mathbf{r}, \mathbf{s} \in \mathbb{R}^2$, operuje v prostoru \mathbb{R}^6 :

$\mathbf{r} \rightarrow \phi(\mathbf{r}) = \{r_1^2, r_2^2, \sqrt{2}r_1r_2, \sqrt{2}r_1, \sqrt{2}r_2, 1\}$

Trik s jádrem (*Kernel trick*)

Příklad

Polynomiální jádro stupně $q = 2$: $K(\mathbf{r}, \mathbf{s}) = (\mathbf{r}'\mathbf{s} + 1)^2$,

$\mathbf{r}, \mathbf{s} \in \mathbb{R}^2$, operuje v prostoru \mathbb{R}^6 :

$$\mathbf{r} \rightarrow \phi(\mathbf{r}) = \{r_1^2, r_2^2, \sqrt{2}r_1r_2, \sqrt{2}r_1, \sqrt{2}r_2, 1\}$$

$$\mathbf{s} \rightarrow \phi(\mathbf{s}) = \{s_1^2, s_2^2, \sqrt{2}s_1s_2, \sqrt{2}s_1, \sqrt{2}s_2, 1\}$$

Trik s jádrem (*Kernel trick*)

Příklad

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

$$\mathbf{r} \rightarrow \phi(\mathbf{r}) = \{r_1^2, r_2^2, \sqrt{2}r_1r_2, \sqrt{2}r_1, \sqrt{2}r_2, 1\}$$

$$\mathbf{s} \rightarrow \phi(\mathbf{s}) = \{s_1^2, s_2^2, \sqrt{2}s_1s_2, \sqrt{2}s_1, \sqrt{2}s_2, 1\}$$

$$\phi(\mathbf{r})'\phi(\mathbf{s}) = r_1^2s_1^2 + r_2^2s_2^2 + 2r_1r_2s_1s_2 + 2r_1s_1 + 2r_2s_2 + 1$$

Trik s jádrem (*Kernel trick*)

Příklad

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

$$\mathbf{r} \rightarrow \phi(\mathbf{r}) = \{r_1^2, r_2^2, \sqrt{2}r_1r_2, \sqrt{2}r_1, \sqrt{2}r_2, 1\}$$

$$\mathbf{s} \rightarrow \phi(\mathbf{s}) = \{s_1^2, s_2^2, \sqrt{2}s_1s_2, \sqrt{2}s_1, \sqrt{2}s_2, 1\}$$

$$\phi(\mathbf{r})'\phi(\mathbf{s}) = r_1^2s_1^2 + r_2^2s_2^2 + 2r_1r_2s_1s_2 + 2r_1s_1 + 2r_2s_2 + 1$$

$$\begin{aligned}(\mathbf{r}'\mathbf{s} + 1)^2 &= (r_1s_1 + r_2s_2 + 1)^2 = \\ &r_1^2s_1^2 + r_2^2s_2^2 + 1 + 2r_1r_2s_1s_2 + 2r_1s_1 + 2r_2s_2\end{aligned}$$

demonstrační data:

Bayes Optimal Classifier

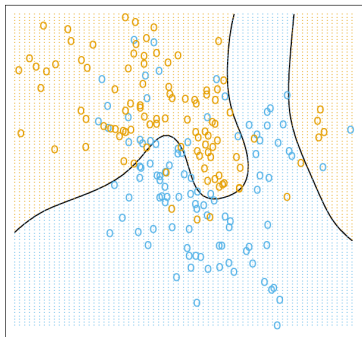


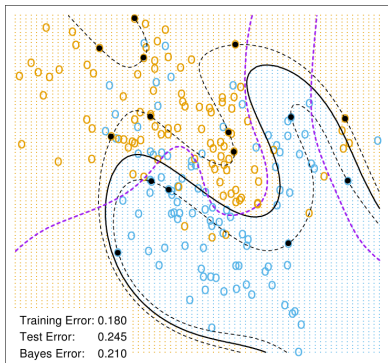
FIGURE 2.5. The optimal Bayes decision boundary for the simulation example of Figures 2.1, 2.2 and 2.3. Since the generating density is known for each class, this boundary can be calculated exactly (Exercise 2.2).

(převzato z <http://statweb.stanford.edu/~tibs/ElemStatLearn/>)

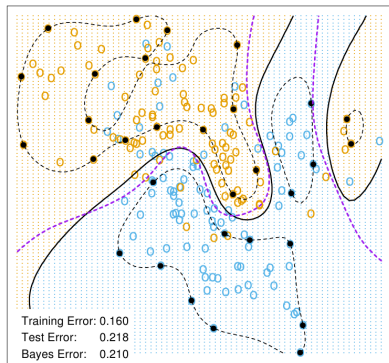
Ukázka jader

jádra:

SVM - Degree-4 Polynomial in Feature Space



SVM - Radial Kernel in Feature Space



(převzato z <http://statweb.stanford.edu/~tibs/ElemStatLearn/>)

Klasifikace s “měkkou hranicí”

Soft-margin classification

Motivace:

- třídy nemusejí být oddělitelné
- přesto chceme SVM použít

Řešení:

- dovolit SVM udělat “malou” chybu

Jak:

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

- 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*

Porovnání metod

TABLE 10.1. Some characteristics of different learning methods. Key: ▲ = good, ◆ = fair, and ▼ = poor.

Characteristic	Neural Nets	SVM	Trees	MARS	k-NN, Kernels
Natural handling of data of “mixed” type	▼	▼	▲	▲	▼
Handling of missing values	▼	▼	▲	▲	▲
Robustness to outliers in input space	▼	▼	▲	▼	▲
Insensitive to monotone transformations of inputs	▼	▼	▲	▼	▼
Computational scalability (large N)	▼	▼	▲	▲	▼
Ability to deal with irrelevant inputs	▼	▼	▲	▲	▼
Ability to extract linear combinations of features	▲	▲	▼	▼	◆
Interpretability	▼	▼	◆	▲	▼
Predictive power	▲	▲	▼	◆	▲

(převzato z <http://statweb.stanford.edu/~tibs/ElemStatLearn/>)

Klasifikace do více tříd

SVM umí rozlišovat jen do dvou tříd

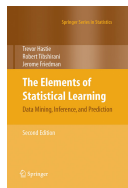
možná řešení klasifikace do K tříd:

- “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*)
- shlukování založené na SVM

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** (2009). Springer-Verlag, <http://statweb.stanford.edu/~tibs/ElemStatLearn/>

Shlukování založené na SVM

myšlenka: obklopit body co nejmenší koule v mnohorozměrném prostoru, koule promítnout zpět do příznakového prostoru jako kontury a shluky definovat jako pozorování ve stejných konturách

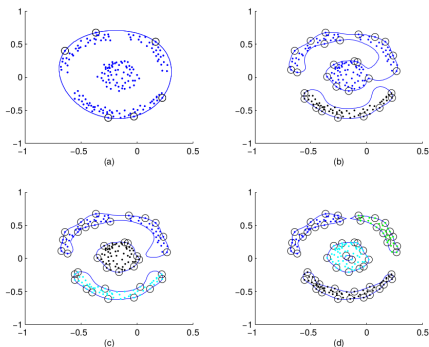
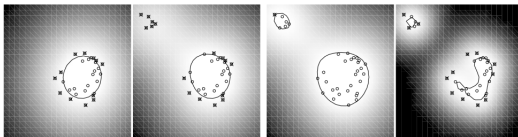


Figure 1: Clustering of a data set containing 183 points using SVC with $C = 1$. Support vectors are designated by small circles, and cluster assignments are represented by different grey scales of the data points. (a): $q = 1$ (b): $q = 20$ (c): $q = 24$ (d): $q = 48$.

Detekce nečekaných pozorování *novelty detection*

myšlenka: v mnohorozměrném prostoru oddělit pozorování od počátku nadrovinou (maximalizace hranice),
nečekaná pozorování pak jsou ta vně oné nadroviny



ν , width c	0.5, 0.5	0.5, 0.5	0.1, 0.5	0.5, 0.1
frac. SVs/OLs	0.54, 0.43	0.59, 0.47	0.24, 0.03	0.65, 0.38
margin $\rho/\ w\ $	0.84	0.70	0.62	0.48

Figure 1: *First two pictures:* A single-class SVM applied to two toy problems; $\nu = c = 0.5$, domain: $[-1, 1]^2$. Note how in both cases, at least a fraction of ν of all examples is in the estimated region (cf. table). The large value of ν causes the additional data points in the upper left corner to have almost no influence on the decision function. For smaller values of ν , such as 0.1 (*third picture*), the points cannot be ignored anymore. Alternatively, one can force the algorithm to take these ‘outliers’ into account by changing the kernel width (2): in the *fourth picture*, using $c = 0.1$, $\nu = 0.5$, the data is effectively analyzed on a different length scale which leads the algorithm to consider the outliers as meaningful points.

zdroj: *Support Vector Method for Novelty Detection*, B Schölkopf et al., 2000