

Algoritmy a struktury neuropočítačů ASN – P9
SVM – Support vector machines
Support vector networks
(Algoritmus podpůrných vektorů)

Autor: Vladimir Vapnik

Vapnik, V. *The Nature of Statistical Learning Theory*. Springer Verlag, New York, NY, USA. 1995

- metoda strojového učení (**učení s učitelem**)
- zavedena pro řešení úloh rozpoznání vzorů
- používá se pro klasifikaci , nově i pro regresi do 2 tříd (binární klasifikace)
- jedná se o lineární klasifikátor
- support vectors jsou nejbližší body k nadrovině

Princip: - nalezení nadroviny, která optimálně rozdělí trénovací data,
- základem je statistické učení využívající transformace nelineárního prostoru s nižší dimenzí do lineárního prostoru s vyšší dimenzí

Nástroj: jádrová funkce (*kernel function*)

Reálná data nejsou obvykle lineárně separabilní, zvýšením dimenze prostoru lze lineární separability dosáhnout.

Formální definice: SVM vytváří nadrovinu nebo množinu nadrovin v prostoru s vysokou nebo nekonečnou dimenzí

Dobrá separace se dosahuje pomocí nadroviny, od které mají datové body klasifikované do nějaké třídy největší vzdálenost - *(na obě strany od nadroviny je co nejširší pruh bez bodů)*.



maximal margin
pásmo necitlivosti, hraniční pásmo

Obecně větší vzdálenost vede na menší chybu při generalizaci.

Reálná data nejsou obvykle lineárně separabilní, zvýšením dimenze prostoru lze lineární separability dosáhnout.

Jak najít optimální nadrovinu ?

kernel

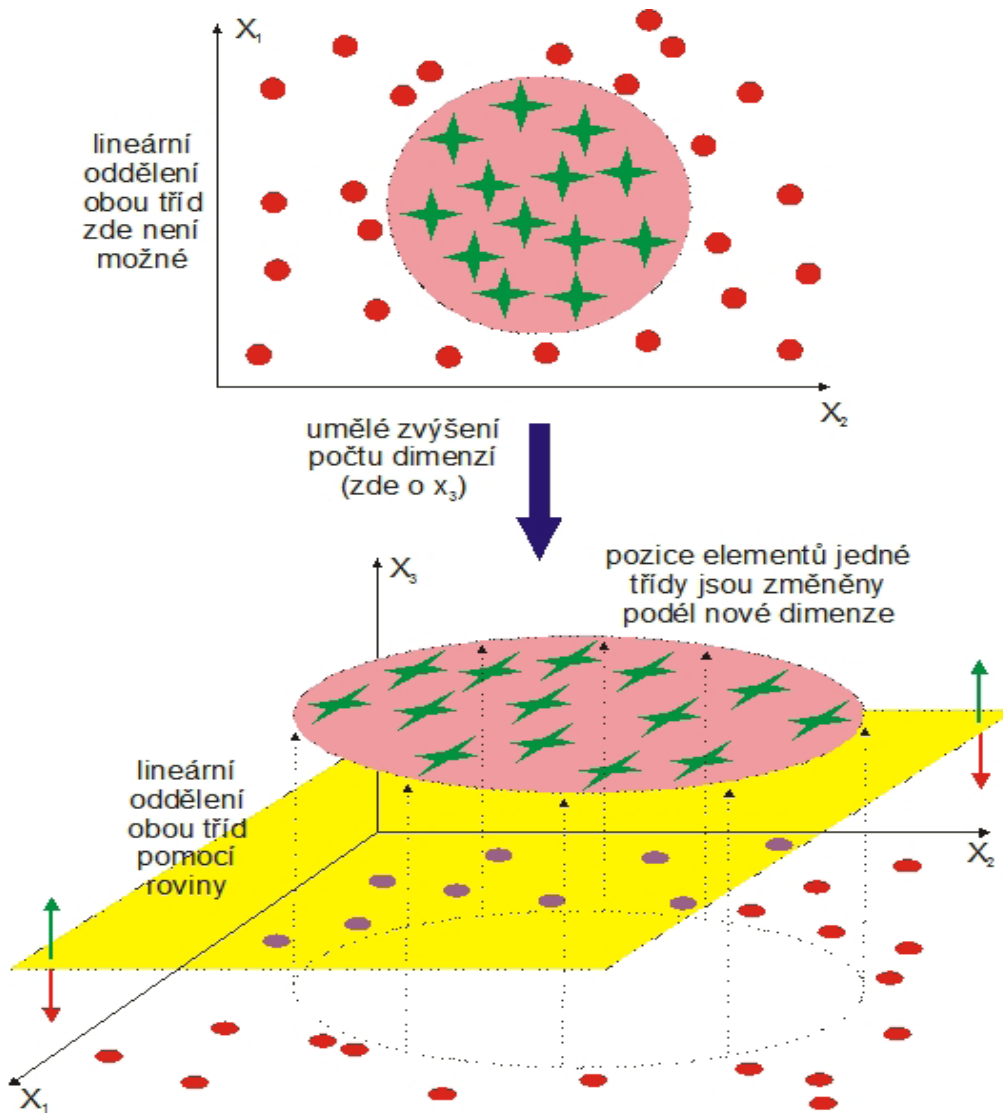


pomocí jádrové funkce jako skalární součin dvojic datových bodů

Optimální nadrovina se nazývá *maximum-margin hyperplane*

oddělovací hranice

$$x^2_1 + x^2_2 \leq 1$$

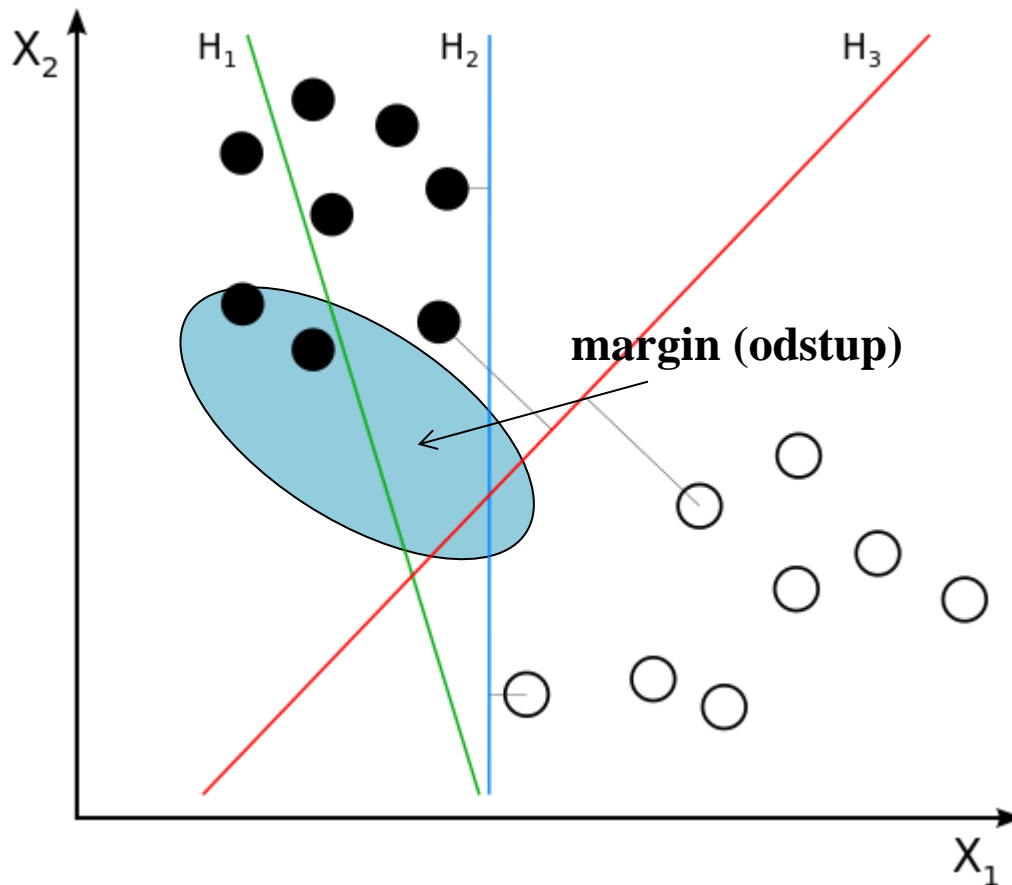


jádrová (kernel) transformace převádí nelineárně separabilní data na lineárně separabilní

H1 neoddělí třídy

H2 oddělí třídy, ale s malou vzdáleností od nadroviny

H3 oddělí třídy s maximální vzdáleností od nadroviny



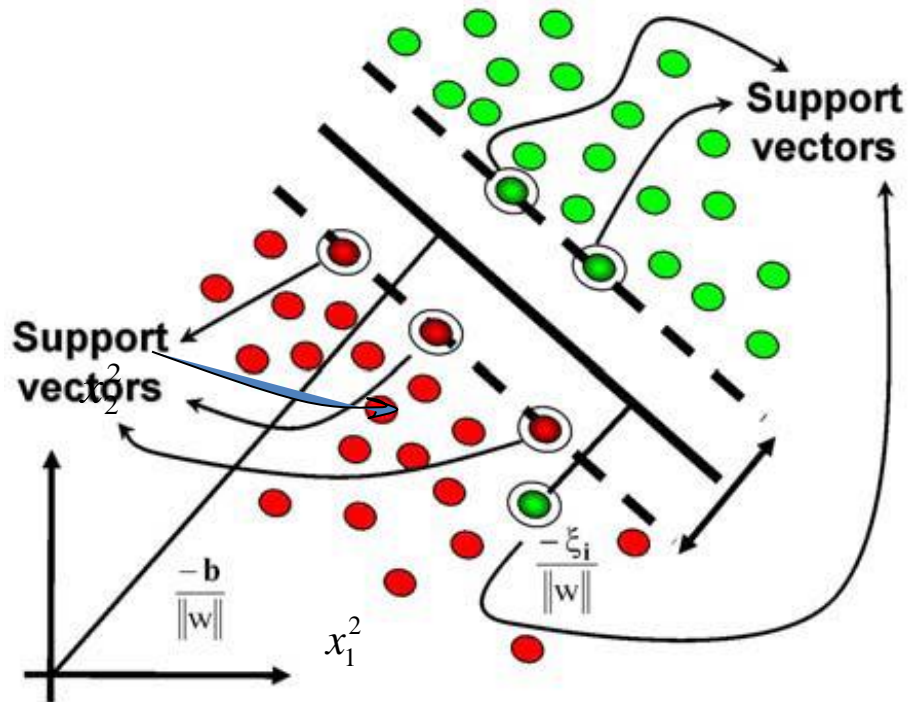
Obrázek převzat a upraven z

http://en.wikipedia.org/wiki/Support_vector_machine

optimální nadrovina

Obrázek převzat z:

http://dspace.upce.cz/bitstream/10195/39021/1/diplomka_Flegl.pdf



nejbližší body k nadrovině = podpůrné vektory
jádrová funkce (*kernel function*)

$$K(\mathbf{x}_i \bullet \mathbf{x}_j) = (\mathbf{x}_i \bullet \mathbf{x}_j)^2 = F(\mathbf{x}_i) \bullet F(\mathbf{x}_j)$$

skalární součin

Klasifikátor zkonstruovaný pomocí SVM:

maximum margin classifier nebo
perceptron of optimal stability

Vhodný i pro práci se zašuměnými daty.

SVM v MATLABu

Matlab R2015b, R2016b

Statistics and Machine Learning Toolbox

Funkce:

fitcsvm	trénink binárního SVM klasifikátoru
predict	predikce labelů pro SVM klasifikátor
templateSVM	SVM šablona
fitcecoc	vhodné modely pro SVM nebo ostatní klasifikátory
templateECOC	kódovací šablona pro korekci chyb
predict	predikce labelů pro korekci chyb výstupního kódu klasifikátoru
loss	klasifikace chyb u SVM klasifikátoru
margin	klasifikace odstupů u SVM klasifikátoru
ClassificationSVM	SVM pro binární klasifikaci
ClassificationECOC	SVM pro klasifikaci do více tříd

**transformace se dá definovat pro různé typy objektů, nejen body v \mathbb{R}^n .
Např. pro grafy, stromy, posloupnosti DNA ...**

Podpůrné vektory jsou nejdůležitějšími body trénovací množiny.

**Při odstranění všech ostatních bodů se novým natrénováním získá tatáž
rozdělující nadrovina.**

**V testovací fázi klasifikujeme neznámé body, u nichž nevíme, do které třídy
náležejí.**

**Spočítáme $\text{sgn}(w \cdot x + b)$, výsledné znaménko určí, na které straně rozdělující
nadroviny daný bod leží  určení třídy.**

Příklady nelineárních SVM:

Polynom stupně, radiální bázová funkce, dvojvrstvá neuronová síť.

Problém optimalizace pomocí podpůrných vektorů lze řešit analyticky jen při malém množství trénovacích dat nebo jsou-li data separovatelná nebo je předem známo, která data se stanou podpůrnými vektory.

Support Vector Machines mají široké uplatnění (zvláště pro klasifikaci).

Reálné aplikace:

- oblast zpracování textu,
- počítačové vidění (např. rozpoznávání obličejů,
rozpoznávání ručně psaného textu)
- bioinformatika (např. klasifikace proteinů)

Výhody:

- trénování vždy najde globální řešení
- neexistují žádná lokální minima, která by nebyla zároveň globálním minimem
- reprodukovatelné výsledky
- výsledky nezávislé na volbě konkrétního algoritmu
- výsledky nezávislé na volbě počátečního bodu
- malý počet parametrů modelu, které je třeba určit po výběru typu jádrové funkce

Problémy a omezení:

- správná funkce SVM záleží na vhodné volbě jádrové funkce (z velké části)
- po stanovení jádrové funkce jen jeden parametr - chyba
- obtížné určování jádrové funkce
- u velmi rozsáhlých úloh (milióny podpůrných vektorů) je problém v rychlosti a s paměťovými nároky

Publikace:

- [SUY99] Suykens, J.A.K., Vandewalle, J.: Least Squares Support Vector Machine Classifiers. Neural Processing Letters, 1999.**
- [TON01] Tong, S., Koller, D.: Support Vector Machine Active Learning with Applications to Text Classification. Journal of Machine learning Research, 2001, pp: 45-66.**
- [STE08] Steinwart, B., Smola, A.: Support Vector Machine. Springer, New York, 2008, ISBN 978-0-387-77241-7, 602pp.**