

SVM - Support Vector Machines

Menu: QCExpert SVM

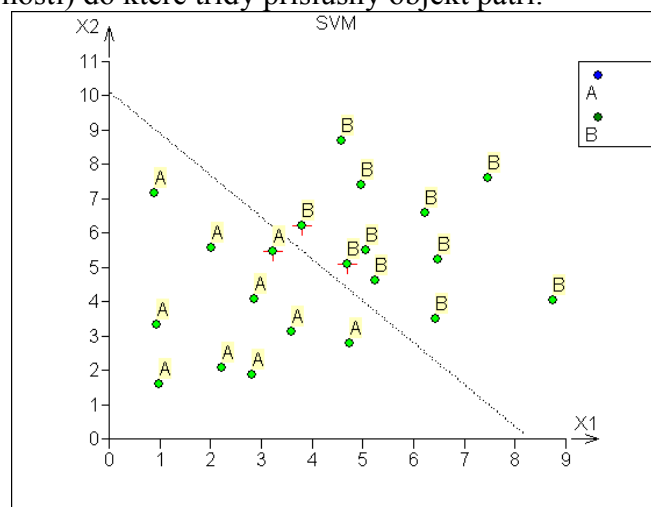
Modul Support Vector Machines (SVM, do češtiny se někdy překládá jako metoda podpůrných vektorů) nabízí velmi progresivní a novou metodu z oblasti strojového učení, kterou rozpracoval koncem 20. a začátkem 21. století Vladimir Naumovič Vapnik a Alexej Jakovlevič Červoněnkis (Akademie věd SSSR, Stanford University, Royal Holloway College London, AT&T Bell Labs New Jersey, NEC Labs Princeton, Columbia University New York). Těto metody lze využít především pro klasifikační úlohy, rovněž ale nachází uplatnění v regresním modelování a neparametrických odhadech hustoty. Modely SVM využívají teorii empirického risku R a Vapnik-Chervonenkisovy (VC) dimenze modelu h . S pravděpodobností $1 - \eta$ platí nerovnost

$$R(\alpha) \leq R_{emp}(\alpha) + \sqrt{\frac{h(\ln(2l/h) + 1) - \ln(\eta/4)}{l}}, \quad (0-1)$$

kde $R(\alpha) = \int \frac{1}{2} |y - f(\mathbf{x}, \alpha)| p(\mathbf{x}, y) d\mathbf{x} dy$ je risk (skutečná střední chyba modelu), l je počet dat, α parametry modelu, $R_{emp}(\alpha) = \frac{1}{2l} \sum_{i=1}^l |y_i - f(x, \alpha)|$ je empirický risk a h je nezáporná celočíselná VC-dimenze modelu. Poslední člen na pravé straně (celá odmocnina) se nazývá VC-confidence.

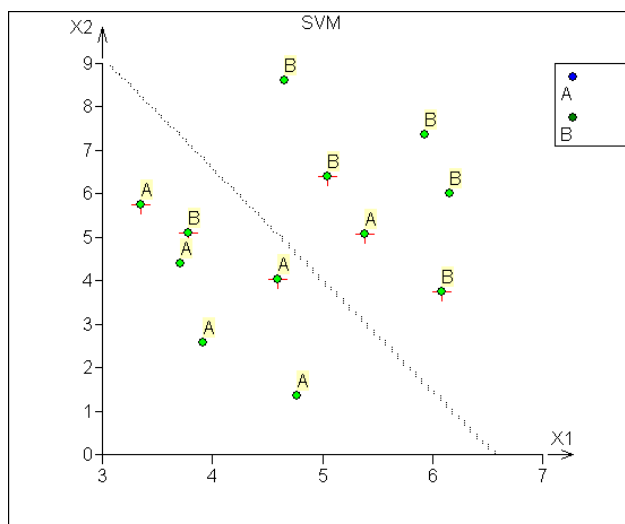
SVM-C – Klasifikační modely SVM

Vlastní modely SVM pak minimalizují vhodně definovanou chybu (např. misklasifikaci u klasifikačních modelů, chybu v nějaké metrice u regrese). Například v úloze lineární separabilní klasifikace ve dvou dimenzích (dvě číselné nezávisle proměnné) se hledá přímka, která rozděluje (diskriminuje) obě třídy tak, že zachová maximální vzdálenost od experimentálních dat a tím za poměrně obecných předpokladů minimalizuje riziko misklasifikace v predikci z nových dat, viz Obrázek 1. Model SVM pak může mimo jiné sloužit k tomu, abychom na základě nových hodnot nezávisle proměnné uměli odhadnout (včetně pravděpodobností) do které třídy příslušný objekt patří.



Obrázek 1 SVM, separabilní data, lineární model

V neseparabilním případě se hledá přímka, která minimalizuje „přešlapy“ nesprávně klasifikovaných dat. Příklad uvádí Obrázek 2, kde jsou minimalizovány kolmé vzdálenosti jednoho nesprávně klasifikovaného bodu B a jednoho nesprávně klasifikovaného bodu A od separační přímky při současném požadavku maximální separace ostatních dat. Tím je zajištěna vysoká účinnost klasifikačního modelu.



Obrázek 2 SVM, lineárně neseparabilní data, lineární model

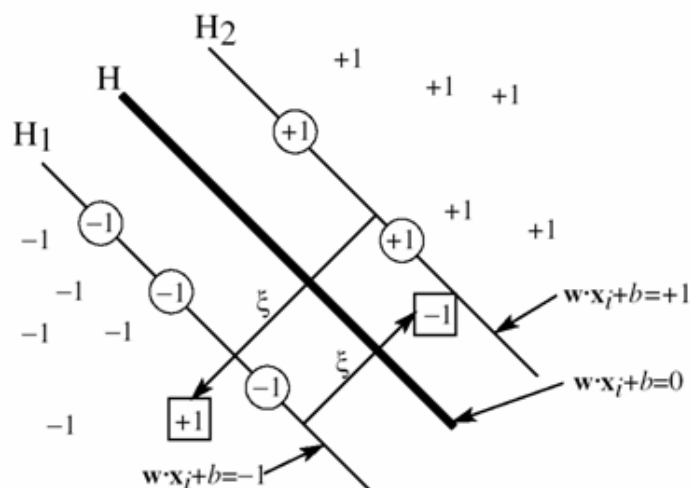
V případě separabilních dat s binární odezvou ($y_i = -1$ nebo 1) se minimalizuje délka normály \mathbf{w} separační přímky (obecně nadroviny)

$$f(x) = \frac{\|\mathbf{w}\|^2}{2} \text{ za omezující podmínky } y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geq 0, i = 1, \dots, l,$$

v případě neseparabilních dat se k optimalizovanému kritériu přidá ještě člen penalizace za „přešlap“ s uživatelem definovaným parametrem C , který lze chápat jako cenu (cost) za chybnou klasifikaci

$$f(x) = \frac{\|\mathbf{w}\|^2}{2} + C \sum_{i=1}^l \xi_i \text{ za podmínky } \begin{cases} y_i (\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, i = 1, \dots, l \\ \xi_i \geq 0, i = 1, \dots, l \end{cases} \quad (\mathbf{0-2})$$

Geometrickou interpretaci v případě neseparabilních dat ilustruje Obrázek 3, souřadnice bodů (odpovídající řádek nezávisle proměnných), o které se „opírají“ separační meze H_1 a H_2 (na obrázku označeny kroužkem), se nazývají *support vectors* (česky se překládají jako podpůrné vektory).



Obrázek 3 Schéma separace SVM-nadrovinou H pro neseparabilní data

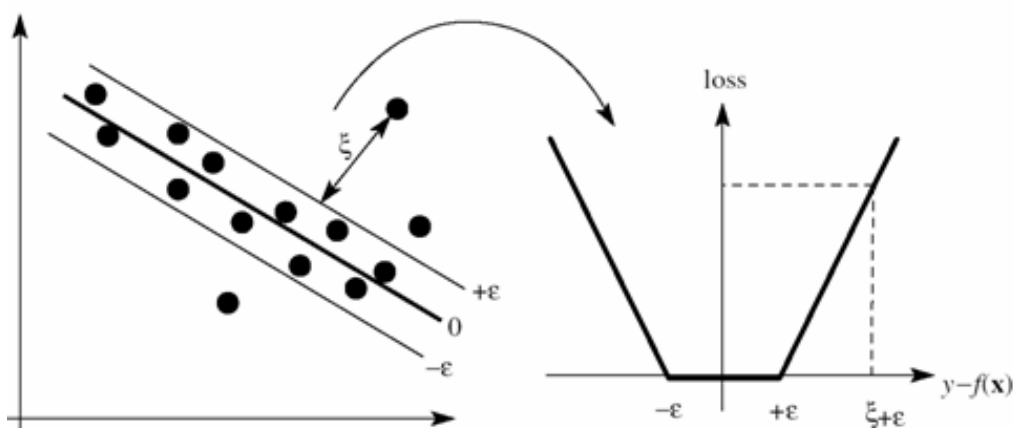
Alternativně lze použít kriteriální podmínku obsahující místo ztrátového koeficientu C zlomek ν ($0 < \nu < 1$), který přibližně odpovídá očekávanému maximálnímu podílu misklasifikací (chybných klasifikací).

$$f(x) = \frac{\|\mathbf{w}\|^2}{2} - \nu\rho + \frac{1}{l} \sum_{i=1}^l \xi_i \quad \text{za podmínky} \quad \begin{aligned} y_i(\mathbf{w} \cdot \mathbf{x}_i + b) &\geq \rho - \xi_i, i = 1, \dots, l \\ \xi_i &\geq 0, i = 1, \dots, l; \rho \geq 0 \end{aligned} \quad (0-3)$$

SVM-R – Regresní modely SVM

Podobně lze zapsat analogické kritérium pro SVM regresi, které definuje přijatelnou chybu modelu ε , mimo tuto chybu je model opět penalizován lineární funkcí se ztrátovým koeficientem $C > 0$. Odhad lineárního modelu se získá minimalizací

$$\frac{\|\mathbf{w}\|^2}{2} + C \sum_{i=1}^l (\xi_i^+ + \xi_i^-) \quad \text{za podmínky} \quad \begin{aligned} \mathbf{w} \cdot \mathbf{x}_i + b - y_i &\leq \varepsilon + \xi_i^+, i = 1, \dots, l \\ y_i - \mathbf{w} \cdot \mathbf{x}_i + b &\leq \varepsilon + \xi_i^-, i = 1, \dots, l \end{aligned} \quad (0-4)$$



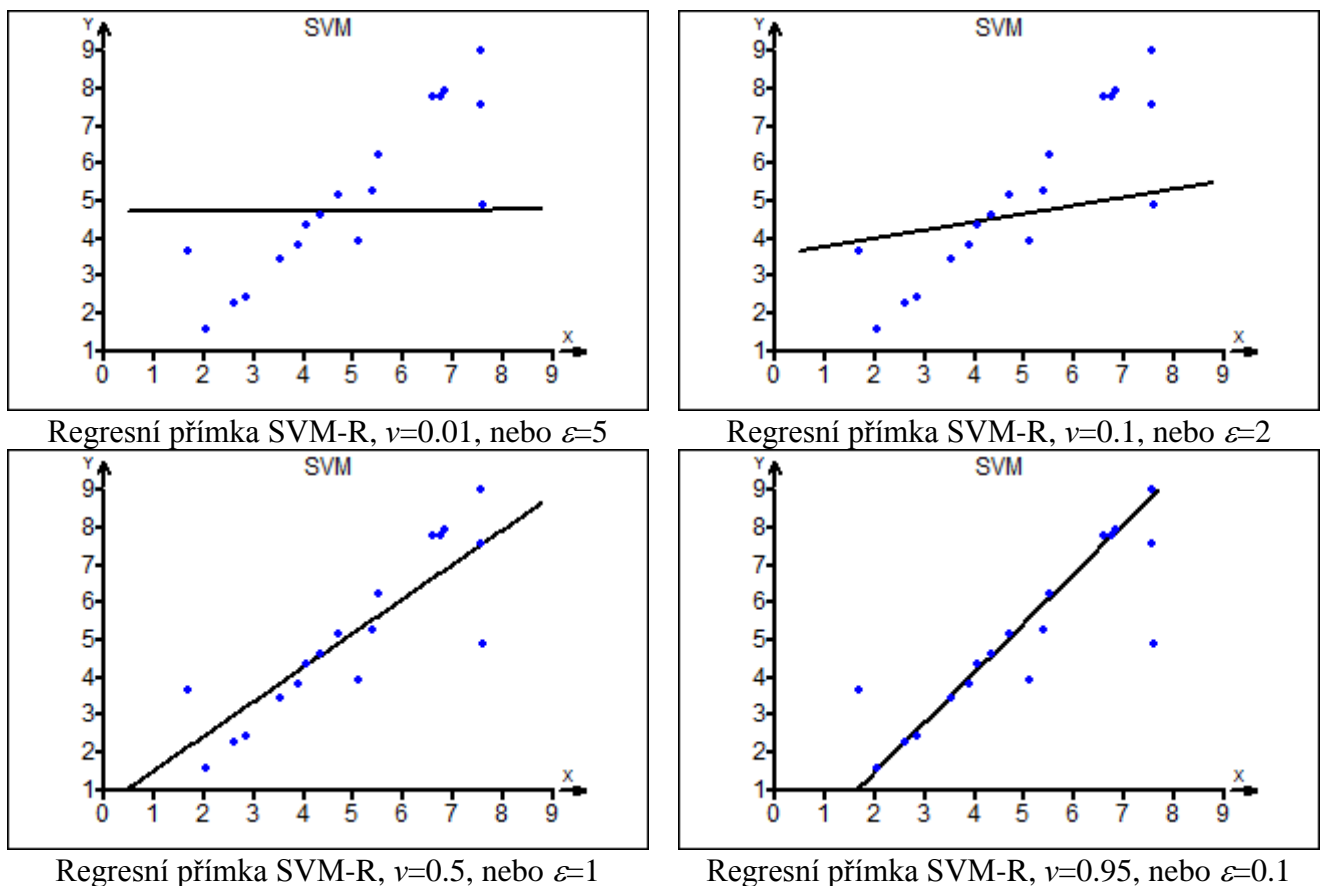
Obrázek 4 Schéma SVM regrese (SVM-R)

Koeficienty \mathbf{w} zde odpovídají regresním koeficientům lineárního modelu, b je absolutní člen. Parametr ε je pološířka pásu, v němž očekáváme správné hodnoty y_i , lze tedy ε

chápat jako nejvyšší přípustnou absolutní chybu. Takto definované regresní kritérium vede mimochodem i k nastavitelné robustnosti regresního modelu, který se snaží „dostat“ co nejvíce dat do pásu $f(x) \pm \varepsilon$, proto je SVM regrese vhodná také pro data znečištěná hrubými chybami měření, odlehlými hodnotami y a podobně. Čím je vyšší koeficient ztráty C , tím vyšší je penalizace za překročení absolutní odchylky ε , případně podílu ν , jak ilustruje Obrázek 5. Podobně jako v předchozím odstavci lze rovněž zavést alternativní vyjádření kritéria SVM-R s parametrem ν ($0 < \nu < 1$), který přibližně odpovídá očekávanému maximálnímu podílu korektních bodů, tedy dat, která odpovídají očekávanému modelu, který se získá vázanou minimalizací vzhledem k \mathbf{w} , b , ξ , ξ^* , ε výrazu

$$f(x) = \frac{\|\mathbf{w}\|^2}{2} + C \left(\nu \varepsilon + \frac{1}{l} \sum_{i=1}^l (\xi_i + \xi_i^*) \right) \text{ za podmínek } \begin{aligned} &(\mathbf{w} \cdot \mathbf{x}_i + b) - y_i \leq \varepsilon + \xi_i \\ &y_i - (\mathbf{w} \cdot \mathbf{x}_i + b) \leq \varepsilon + \xi_i^* \\ &\xi_i, \xi_i^* \geq 0, i = 1, \dots, l; \varepsilon \geq 0 \end{aligned} \quad (0-5)$$

Regresní model SVM lze pak využít také pro predikci očekávané hodnoty závisle proměnné pro nové známé hodnoty nezávisle proměnných. Na rozdíl od klasické regrese nemá tento odhad nutně charakter odhadu střední hodnoty ve smyslu statistického momentu.



Obrázek 5 Vliv ztrátového koeficientu na robustnost modelu SVM-R při $C = 1$

SVM-OneClass – Hustota rozdělení

Pro nalezení hranice rozdělení vícerozměrné náhodné veličiny s libovolným rozdělením lze minimalizovat kritérium (B. Schoelkopf et al., 2001)

$$\frac{\|w\|^2}{2} - \rho + \frac{1}{\nu l} C \sum_{i=1}^l (\xi_i) \quad \text{za podmínky} \quad \begin{cases} \mathbf{w} \cdot \mathbf{x}_i \geq \rho - \xi_i, i = 1, \dots, l \\ \xi_i \geq 0, i = 1, \dots, l \end{cases} \quad (0-6)$$

Takto získaná hranice je opět lineární hyperplocha v \mathbf{x} a odpovídá přibližně ν -kvantilu rozdělení. Koeficient ν , $0 < \nu \leq 1$ tedy odpovídá maximu podílu dat, které chceme oddělit od jádra rozdělení, což lze interpretovat jako $(1 - \nu)$ kvantil rozdělení, ale s výhodou také jako maximální očekávaný podíl vybočujících bodů, které do rozdělení nepatří. Řešením uvedených podmínek je hyperplocha, která ohraničuje „vnitřní“ $(1 - \nu) \cdot 100\%$ část rozdělení. Výsledná hranice pak identifikuje vybočující body a lze jí použít i k predikci, zda nově naměřená data patří do natrénovaného vícerozměrného rozdělení. Takové interpretace lze s výhodou využít například pro detekci změn ve složitých vícerozměrných procesech s libovolným rozdělením.

SVM-jádrové transformace

Uvedené modely a kritéria vedou jen na lineární funkce (diskriminační, regresní) typu $\mathbf{w} \cdot \mathbf{x}$ a nebyly by příliš praktické. Jedním ze zásadních přínosů SVM je transformace l -rozměrného prostoru \mathbf{x} na prostor definovaný systémem nelineárních funkcí $\varphi(\mathbf{x})$, jehož dimenze n nesouvisí s dimenzí \mathbf{x} , často bývá větší a může být až nekonečná. V tomto prostoru se teprve vytvoří příslušný lineární SVM-model. Protože se však model vytváří na nelineárně transformovaném prostoru $\varphi(\mathbf{x})$, je i samotný model nelineární v prostoru původních proměnných \mathbf{x} . Tím získávají SVM neobvyklou flexibilitu. Definují-li se transformace ve formě kvadratické formy $K(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}_j)$, (kde funkce K se nazývá jádrová), lze úlohu nalezení modelu formulovat jako kvadratickou vázanou optimalizaci, pro níž lze v kombinaci s metodou Lagrangeových multiplikátorů nalézt poměrně rychlé a stabilní derivační algoritmy.

Nejčastěji používaná jádra jsou typu RBF (Radial Base Functions) definovaná jako

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|) \quad (0-7)$$

Mezi další, méně používané jádrové transformace patří:

Polynomické jádro $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(\gamma \mathbf{x}_i^T \mathbf{x}_j + r)^d$; $\gamma > 0$

Sigmoida $K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma \mathbf{x}_i^T \mathbf{x}_j + r)$

Lineární (bez transformace) $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$

Parametry γ a d zadává uživatel, r se počítá. Parametr γ má význam strmosti jádra, vyšší hodnoty γ vedou k podrobnějším, někdy i méně stabilním modelům. Zavedením takové nelineární transformace je pak možné v klasifikaci separovat libovolné nelineární útvary v l -rozměrném prostoru, v regresi vytvářet libovolné nelineární regresní funkce a v modelování rozdělení nalézt nejlepší „hranici rozdělení“, přesněji nalézt nelineární hyperplochu, která ohraničuje minimální objem v prostoru dat \mathbb{R}^l , který obsahuje alespoň $100(1 - \nu)\%$ dat.

Příklady

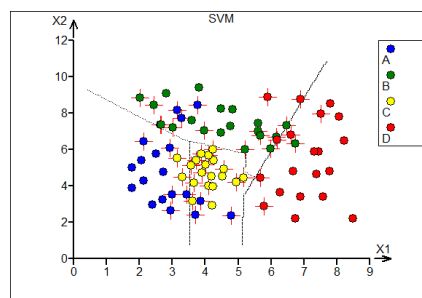
Zde uvádíme několik jednoduchých ilustračních příkladů pro snadnější orientaci v metodách SVM a rychlou referenci pro nastavení a smysl parametrů metod SVM. Přestože metody SVM jsou určeny pro větší množství vysvětlujících proměnných x (prediktorů), pro účely lepší ilustrace se omezíme převážně na dvourozměrné, snáze pochopitelné příklady (jeden nebo dva sloupce x). Pro vícerozměrné modely se nicméně SVM chovají analogicky.

Příklad 1 – Klasifikace

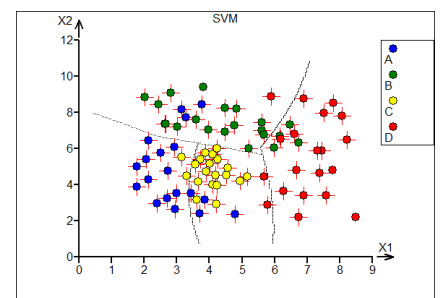
Pro hodnoty spojitých prediktorů X a Y jsou známy kategorie A, B, C, D. Tyto kategorie mohou být například výsledky experimentů při podmínkách x_i, y_i , nebo předem známé kategorie, u nichž jsou známy hodnoty X a Y . Výskyty různých úrovní kategorie se překrývají nejsou zřejmě zcela separovatelné v rovině X, Y . Pro tato (trénovací) data se vytvoří model SVM-C. Následující grafy ukazují chování modelu SVM nejprve bez transformace (lineární jádro) a pak s RBF transformací s různými hodnotami strmosti (nebo nonlinearity) γ od $\gamma=0.01$ do $\gamma=10$. U výsledných modelů jsou uvedeny i počty misklasifikací (misclass), tedy chybných klasifikací. Volbou vhodného γ chceme získat takový model, který bude nejlépe předpovídat kategorii pro nové hodnoty x a y bez ohledu na náhodné umístění trénovacích dat. Proto poslední („přetrénovaný“) model s nejmenším počtem misklasifikací nebude zřejmě nejlepší. Kvalitu modelu lze posoudit cross-validací, kdy odstraníme část z trénovacích dat (obvykle 10-30%) a tuto část pak použijeme pro predikci kategorie. Podíl správně predikovaných kategorií pak může být mírou úspěšnosti zvoleného modelu.

X	Y	Kategorie
4.57	4.93	C
5.21	6.01	B
4.24	3.95	C
3.91	4.73	C
3.29	7.71	A
7.32	5.89	D
5.17	4.43	C
2.05	5.41	A
4.22	2.91	C
7.77	4.78	D
3.15	8.15	A
3.87	5.77	C
3.15	5.53	C
2.63	7.38	A
6.47	7.32	B
4.18	4.52	C

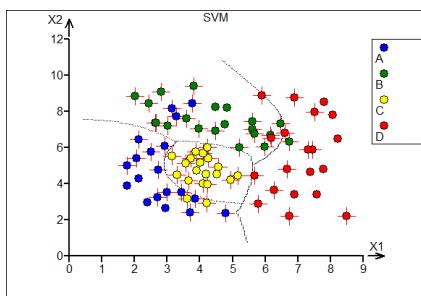
Ukázka tabulky dat



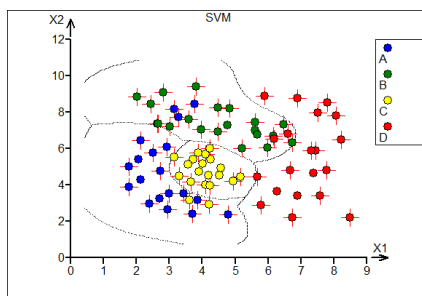
Lineární, Misclass=15



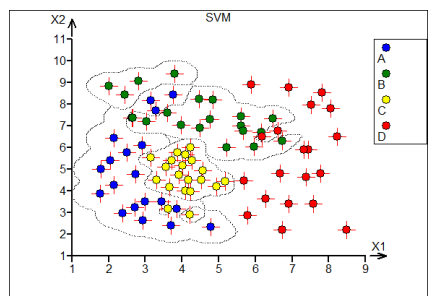
RBF, $\gamma=0.01$, Misclass=14



RBF, $\gamma=0.1$, Misclass=8



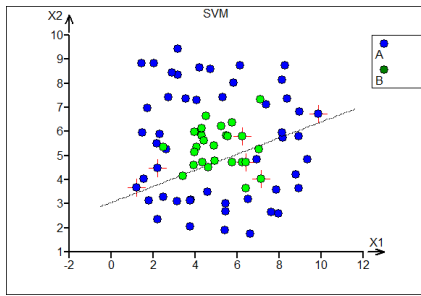
RBF, $\gamma=1$, Misclass=5



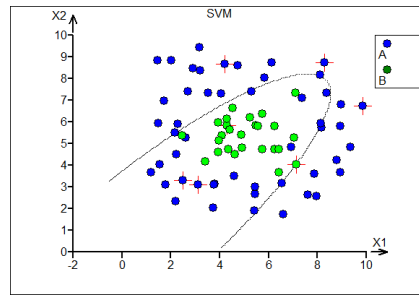
RBF, $\gamma=10$, Misclass=1

Příklad 2 – Klasifikace

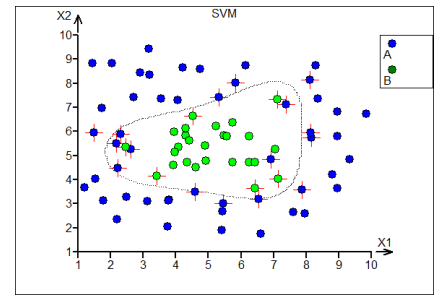
Následující tři grafy ukazují lineárně neseparovatelná data s binární odezvou (A,B), kdy lineární model SVM-C, a částečně i polynomické jádro 2. stupně je geometricky nevhodné a selhává. Transformace s RBF jádrem naopak velmi dobře separuje obě kategorie.



Lineární jádro



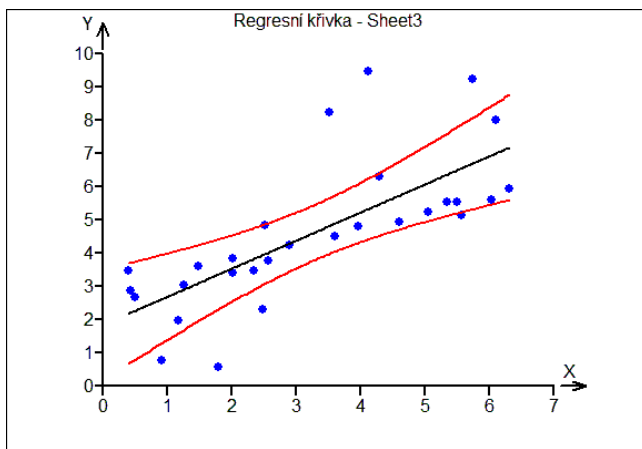
Kvadratické jádro (stupeň=2)



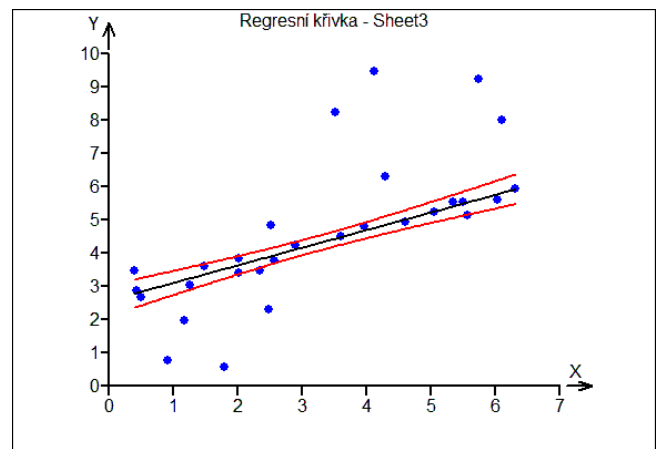
RBF jádro

Příklad 3 – Robustní regrese

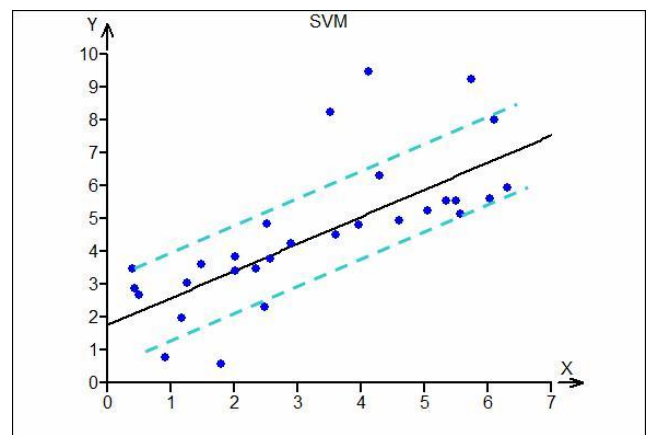
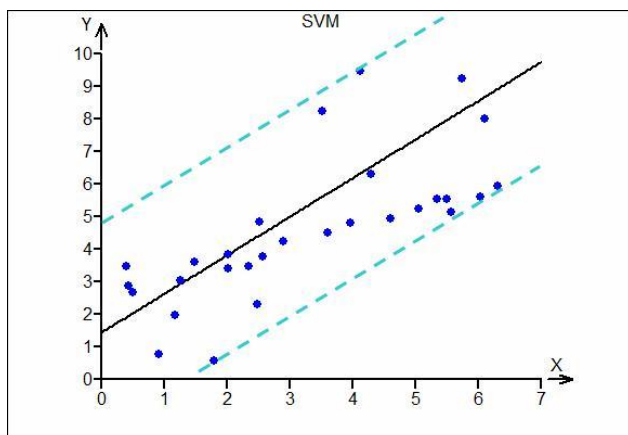
Nastavením parametru ε v rovnici (0-4) na str. 3 se určuje šířka pásu kolem modelu, v němž má ležet co nejvíce dat, viz Obrázek 4, str. 3. Zmenšováním hodnoty tohoto parametru lze dosáhnout obdoby robustnosti u klasických regresních modelů. Data mimo interval $\langle f(x) - \varepsilon; f(x) + \varepsilon \rangle$ se v SVM regresi považují za data, která jsou mimo interval přípustných chyb a tedy neodpovídají modelu. Na následujících grafech je ilustrováno chování lineárního SVM-regresního modelu v porovnání s klasickou lineární regresí metodou nejmenších čtverců (A) a iterativní robustní metodou bounded influence regression (B). Na obrázcích (C) až (F) jsou přímky pro různé široké přípustné meze $\pm\varepsilon$. Z ilustrace je zřejmé, že se model snaží „nacpat“ do zadaného pásu co nejvíce dat. Tím lze dosáhnout plynulho zvyšování robustnosti modelu vůči odlehlým datům. Povaha této robustnosti je však značně odlišná od robustních metod v klasické regresi.



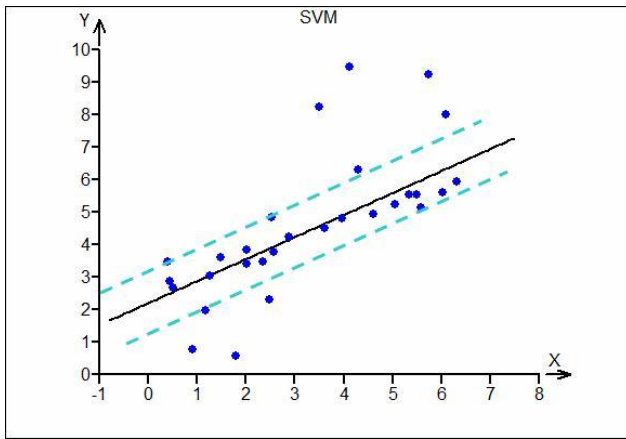
(A) Klasická regrese, metoda nejmenších čtverců



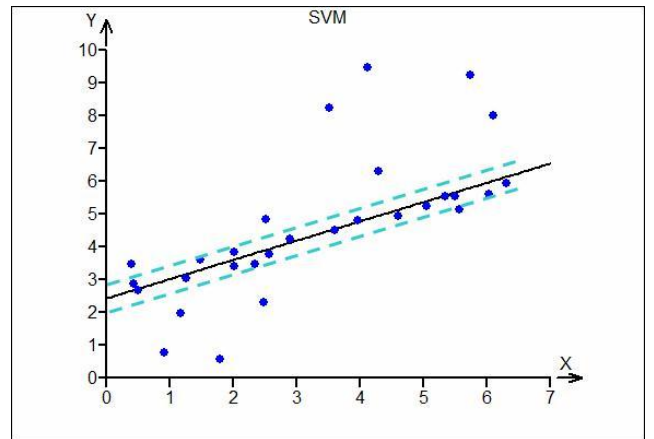
(B) Klasická regrese, robustní metoda BIR



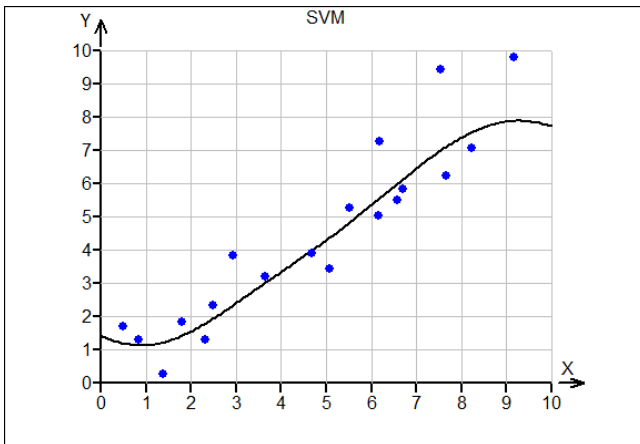
(C) SVM-R, $\epsilon = 3$



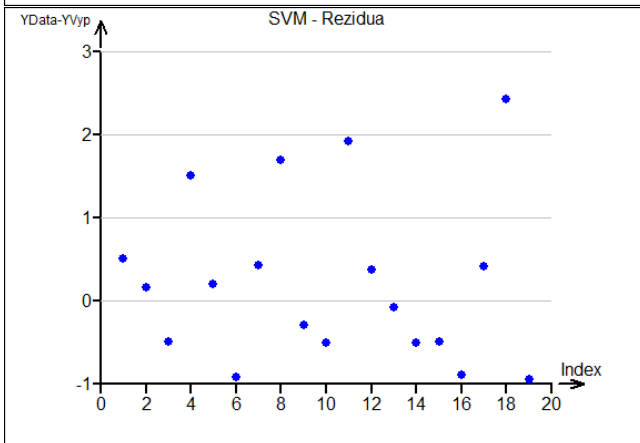
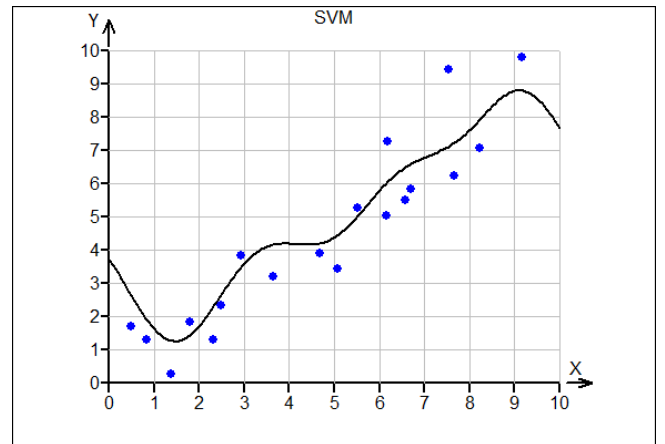
(D) SVM-R, $\epsilon = 1.5$



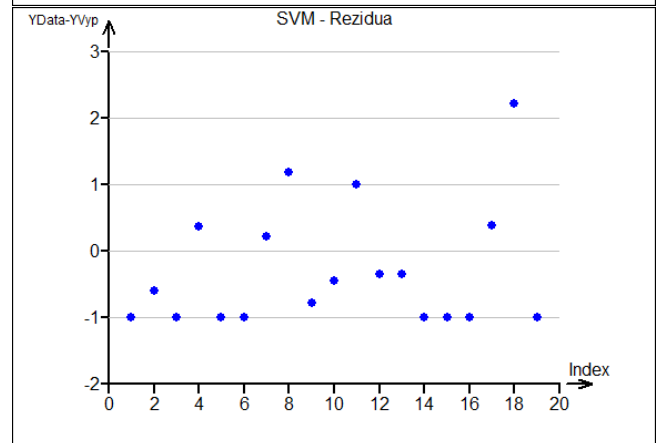
(E) SVM-R, $\epsilon = 1$



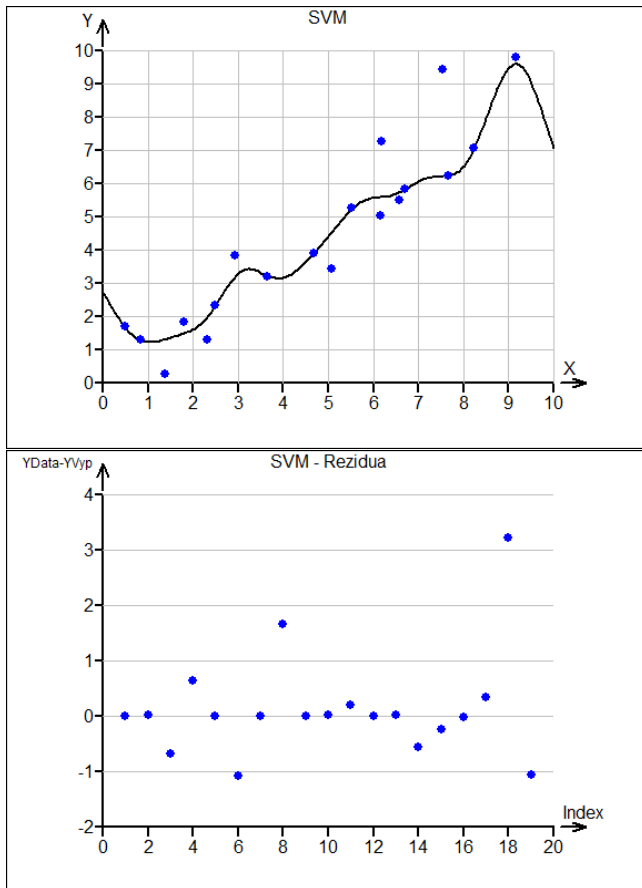
(F) SVM-R, $\epsilon = 0.5$



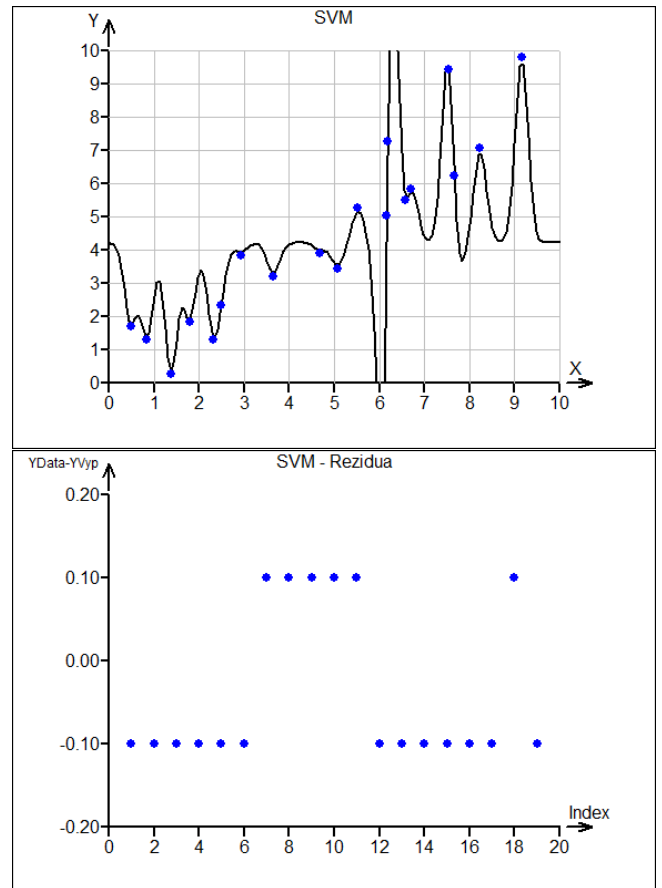
(A) RBF kernel, $\epsilon = 2$; $\gamma = 0.5$



(B) RBF kernel, $\epsilon = 1$; $\gamma = 0.5$



(C) RBF kernel, $\varepsilon=0.5$; $\gamma=5$



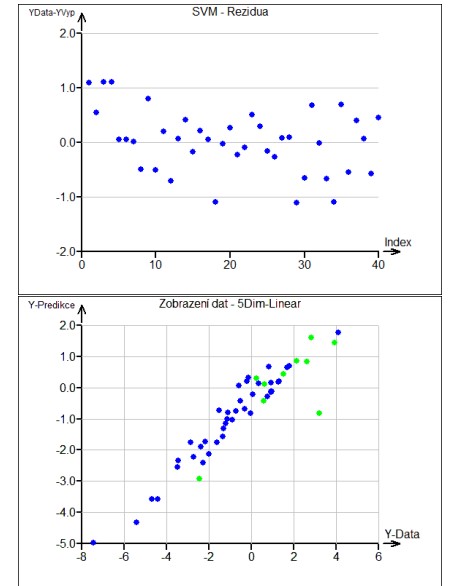
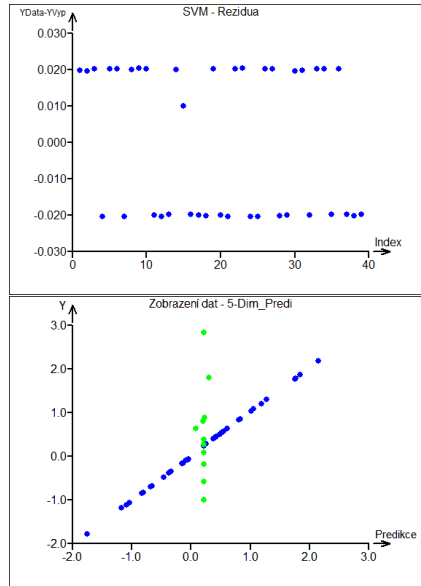
(D) RBF kernel, $\varepsilon=0.1$; $\gamma=25$

Příklad 4 – Adekvátnost regresního modelu

Data odpovídající regresnímu modelu se řídí vztahem $y_i = f(x_i) + e_i$, úkolem regrese je odhad $f(x)$ a tím i velikosti chyb e_i a jejich rozptylu. Díky značné flexibilitě modelů SVM je snadné model přeúčít. Takový model pak nereprezentuje střední hodnotu, ale spíše šum v datech a není použitelný jako popis sledované závislosti. Vhodná diagnostika, která pomůže indikovat přeúčení modelu je (mimo standardních cross-validačních technik) koncentrace výskytu residuí na hranici $\pm \varepsilon$. Grafy na předcházející straně ilustrují čtyři modely na stejných datech, z nichž pouze první (A) zřejmě správně modeluje hladkou křivku skutečné závislosti (zda je modelem křivka (RBF, polynom, sigmoida), či přímka, je volbou uživatele). Křivost ostatních modelů je ovlivněna náhodnými chybami v datech. Koncentrace residuí na nastavených hranicích indikující příliš velkou hodnotu γ v rov. (0-7), případně i C v rov. (0-5) je zvláště patrná na grafech (B) a (D). Model (D) ilustruje mezní možnost, kdy model kopíruje pouze náhodné chyby a je zcela nepoužitelný.

Následující příklad ilustruje postup cross-validace, kdy se model vytvoří z podmnožiny původních dat (obvykle náhodně vybraných 70-90% trénovacích dat) avšak predikce se počítá ze všech (100%) dat. Zobrazení vynechaných testovacích dat v grafu predikce pomůže posoudit predikční schopnost modelu. Modely SVM mohou být tak flexibilní, že „dokonale“ proloží i zcela náhodná data (A), takový model je však nepoužitelný, má nulovou predikční schopnost (světlejší predikovaná data v grafu predikce jsou zcela mimo dokonalý fit). V grafu (B) má sice model 50x větší chyby, než (A), avšak tento model má velmi dobrou predikční schopnost, světlejší testovací data prakticky splývají s trénovacími.

A	B	C	D	E	Y
2.113	0.874	-0.192	-0.764	-0.339	0.270
1.26	-0.318	1.477	-1.732	1.123	0.564
0.082	1.921	0.726	-0.684	1.342	0.832
-1.121	-0.107	-0.755	-1.235	0.351	-0.358
2.41	2.32	-0.119	-0.186	0.339	2.175
0.743	0.931	0.548	0.209	-1.563	0.436
0.524	-1.891	-1.037	0.407	0.416	-1.06
-0.653	0.838	0.961	0.534	-0.158	1.04
-0.699	-0.074	0.833	-0.525	0.445	0.852
-1.067	0.339	-1.595	-0.081	-1.313	-0.994
0.459	0.156	-0.779	-1.235	0.052	0.628
0.748	0.461	0.561	-0.148	-0.605	1.961
0.118	-2.445	-0.485	0.444	1.642	-0.849
-0.451	1.604	-0.077	-1.565	-0.694	-0.076
-1.849	0.276	2.436	-0.496	1.382	-1.778
-0.641	-1.074	0.816	-0.44	-0.515	2.83
-0.598	-0.331	-0.115	-0.869	0.179	0.377
-0.518	1.63	-1.062	-0.902	0.518	1.294
-0.415	-0.364	0.761	-0.331	-1.688	0.23
-0.484	0.281	-0.345	0.175	2.582	-0.186
0.961	-0.139	-0.853	-1.374	0.149	-0.706
-0.982	-0.95	1.277	-0.295	-0.254	0.261
-0.014	-1.972	0.313	0.101	-0.343	-0.108
-0.007	0.269	0.371	0.791	0.241	-0.475
0.091	-0.209	-0.637	1.049	-0.582	0.807
0.649	-1.652	0.739	-1.674	-0.456	1.205
0.315	0.101	-0.639	0.227	-0.197	-0.829
-1.518	0.834	0.353	-0.792	0.292	-0.693
0.756	1.438	-0.064	0.95	1.238	0.394
-1.468	-0.337	-1.246	-0.995	0.912	0.505
-0.127	1.203	-0.341	0.877	1.206	-0.581
-1.574	-3.03	-0.702	-0.118	1.592	-0.144
-0.523	-1.782	-1.036	-0.174	0.501	-1.112
-0.574	-0.795	0.615	0.016	0.388	0.865
2.498	1.585	-1.863	-0.989	0.471	0.45
-0.211	0.646	0.29	-0.229	-0.746	0.629
0.137	1.686	-0.303	-2.563	-0.72	-0.389
0.203	-1.582	0.818	-0.395	-0.346	-0.067
-0.025	0.686	-0.437	-1.871	-1.5	0.535
0.675	-2.284	1.016	-1.491	-1.163	1.711
1.243	-0.574	-1.327	-1.118	-1.23	-0.16
1.302	0.248	1.606	-0.922	0.23	1.785
-0.294	-1.754	1.017	1.2	0.152	1.638
-0.943	1.066	0.15	0.538	-1.711	-0.16
0.199	0.784	0.673	-0.634	-0.775	1.804
1.387	2.29	-0.099	-0.718	1.431	1.076
-0.539	1.192	-2.116	-1.65	0.174	-0.064
0.033	-0.257	1.721	-0.953	-1.276	-1.189
-0.184	-0.144	0.227	0.193	-0.041	-0.173
-1.572	-1.49	1.415	1.447	2.484	0.062

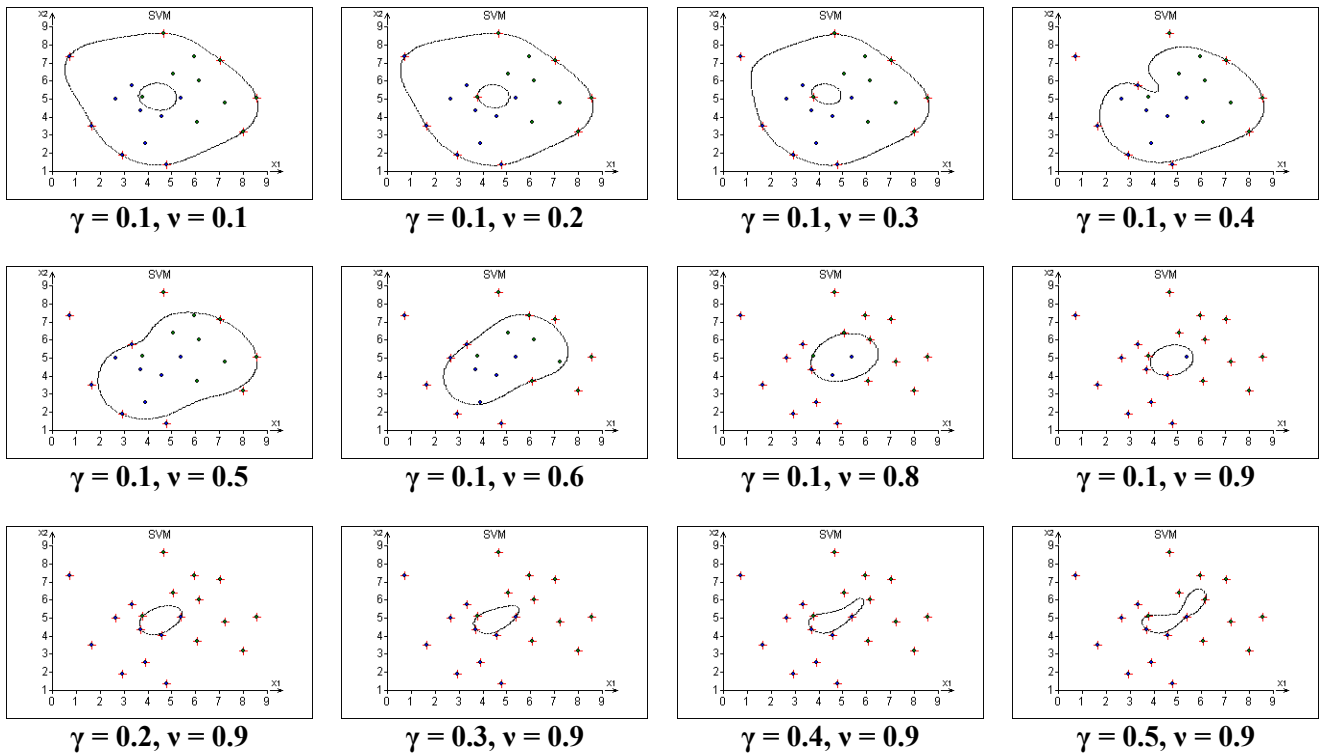


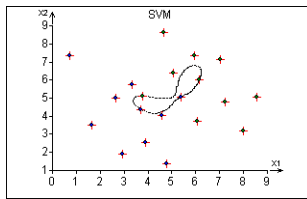
(A) Neadekvátní model s nereálným požadavkem na chyby ($\epsilon = 0.02$), a příliš velkým γ

(B) Adekvátní model pro daná data

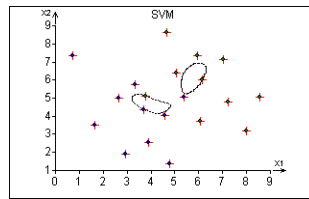
Příklad 5 – Vliv γ a ν na model rozdělení

Následující tabulka grafů ilustruje vliv volitelných parametrů na tvar hranice hustoty rozdělení na příkladu jádra RBF. Parametr γ ovlivňuje „tuhost“ rozdělení, parametr ν odpovídá podílu rozdělení mimo hranici. Uvedených 16 grafů znázorňuje tvar vypočítané hranice pro rostoucí podíl ν při konstantním γ a pro rostoucí γ při konstantním ν .

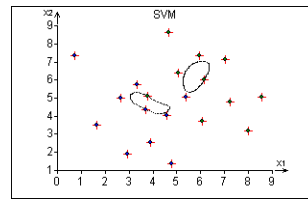




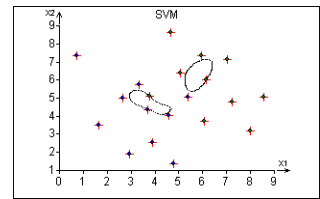
$\gamma = 0.6, \nu = 0.9$



$\gamma = 0.7, \nu = 0.9$



$\gamma = 0.8, \nu = 0.9$



$\gamma = 0.9, \nu = 0.9$

Implementace Support Vector Machines je založena na kódu LIB-SVM (c) Chih-Chung Chang and Chih-Jen Lin vyvíjeného na National Taiwan University, viz např. Chih-Chung Chang and Chih-Jen Lin: *Library for Support Vector Machines*

Doporučená literatura

Vapnik, V.: *The Nature of Statistical Learning Theory*. New York, Springer-Verlag (1995).

Vapnik, V.: *Statistical learning theory*. New York, John Wiley (1998).

Vapnik, V., & Chervonenkis, A.: *Theory of pattern recognition*. Nauka, Moscow (1974). [In Russian]

B. Schoelkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson: Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471, (2001).

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

SVM – Klasifikace

Menu:	QCExpert	SVM	SVM-Klasifikace
-------	----------	-----	-----------------

Data a parametry

Data jsou uspořádána do jednoho nebo více sloupců nezávisle proměnných a jednoho sloupců s úrovněmi faktoru. Nezávisle proměnné jsou číselné hodnoty, faktor musí mít alespoň 2 rozdílné úrovně. Úrovně faktoru mohou být čísla (např. 1, 2, 3), nebo text (například ANO, NE). Každý řádek musí obsahovat platné hodnoty ve všech sloupcích. Neúplné řádky nejsou použity. V dialogovém panelu (Obrázek 6 A) se označí jeden nebo více sloupců nezávisle proměnné a sloupec faktoru. Chceme-li počítat predikci, označíme ještě políčko *Predikce* a vybereme sloupec, pro něj chceme predikovat hodnotu (přesněji: úroveň) faktoru. Stiskem tlačítka „Doplň X“ lze označit sloupec odpovídající vybraným nezávisle proměnným a predikce se pak počítá pro taž data, z nichž se vytvořil model. Ve skupině *Data* se zvolí chceme-li pro výpočet použít jen nějakou podmnožinu dat, či všechna data. Ve skupině *Typ SVM* zvolíme, zda se má použít varianta *Cost* se ztrátovým koeficientem *C*, nebo *Nu* se zadáním koeficientu *ν*. Zaškrtneme-li políčko *Zobrazit Support Vectors*, vypíše se do protokolu všechny support vektory modelu a rovněž se označí v grafu křížkem, pokud se graf konstruuje. To má obvykle orientační význam hlavně u lineárních modelů. V políčku *Popis* lze vybrat sloupec, který se pak použije k identifikaci jednotlivých bodů (řádků) v grafech. Dále se zvolí jádro pro transformaci z možností

Lineární: $(x_i * x_j)$

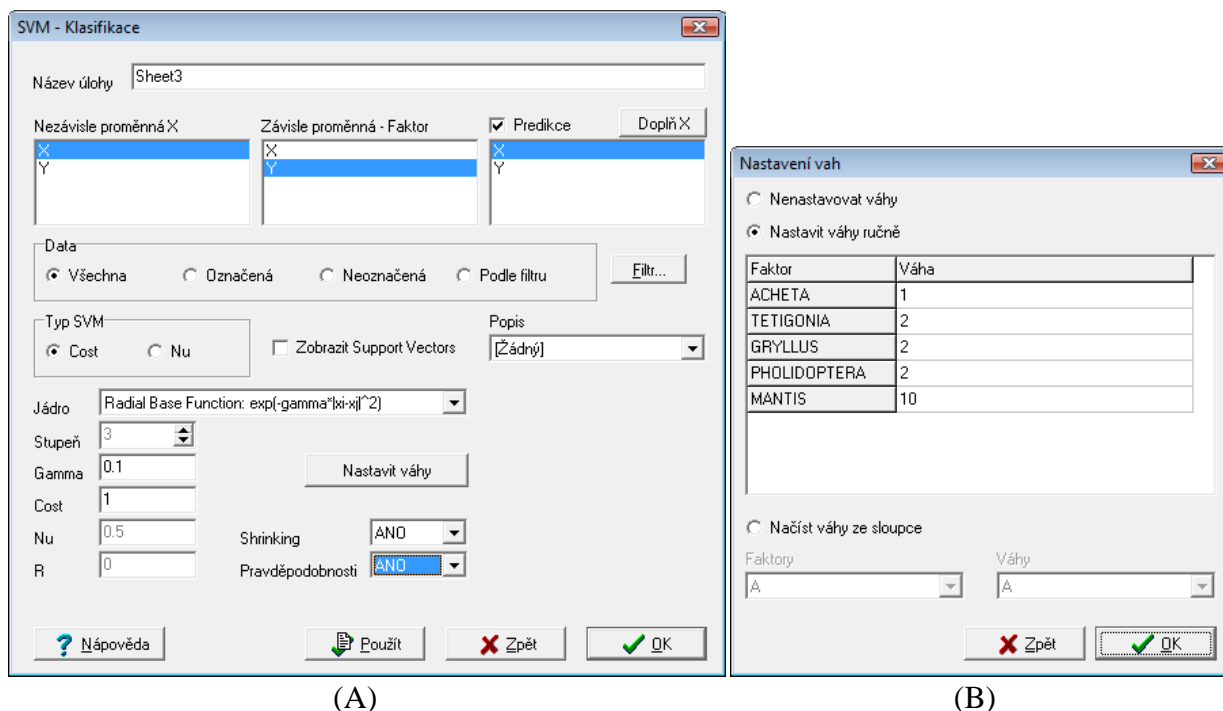
Polynomické: $(-\gamma * (x_i * x_j) + r)^d$

Radial Base Function: $\exp(-\gamma * |x_i - x_j|^2)$

Sigmoida: $\tanh(\gamma * (x_i * x_j) + r)$

a parametry výpočtu: *Stupeň* (pouze u polynomického jádra), hodnotu parametru γ *Gamma*, který souvisí se strmostí jádra a tím s nelinearitou modelu, tento parametr není nutno zadávat

implicitně se použije hodnota $1/n$, ztrátový koeficient C Cost, podíl v Nu a parametr jádra R . Zvolíme, zda chceme provést závěrečnou korekci modelu po výpočtu (*Shrinking*, většinou nemá na výsledek vliv) a zda chceme pro predikované úrovně vypočítat pravděpodobnosti (pole *Pravděpodobnosti*). Tlačítkem *Nastavit váhy* vyvoláme dialogový panel pro zadání vah (Obrázek 6 B). Zde můžeme přiřadit každé úrovni faktoru váhu, kterou můžeme chápat také jako poměr škod při chybné predikci jednotlivých úrovní. Implicitně se váhy nepoužívají, resp. jsou rovny jedné. Po stisku *OK* se zahájí výpočet.



Obrázek 6 Dialogový panel SVM – Klasifikace

Délka	Tloušťka	Typ
6.55	6.01	A
6.35	6.22	A
6.66	4.9	B
7.13	6.3	A
6.49	5.77	C
5.58	5.68	C
4.45	5.26	B
3.7	5.32	B
3.27	4.9	C
3.02	4.55	B

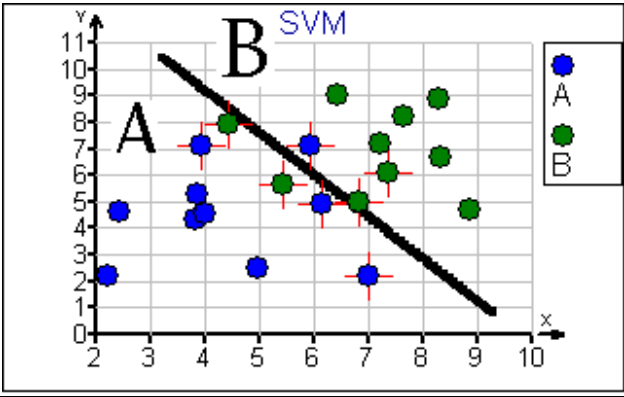
Neznámé vzorky	
Délka 1	Tloušťka 1
4.88	3.89
7.09	5.14
1.25	5.2
1.19	6.48
2.51	5.53

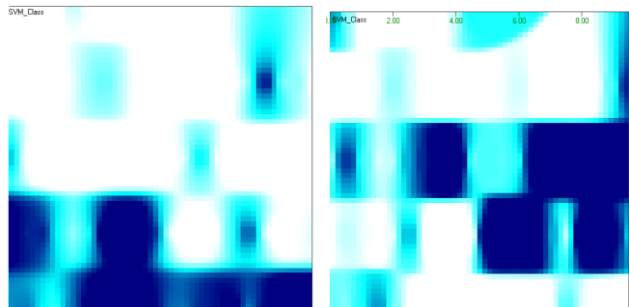
Na	K	Zn	Fe	Ni	Cr	Al	Původ
1.164	0.064	2.975	4.218	3.306	2.155	3.006	Tur
1.098	0.052	4.117	2.38	3.031	1.904	3.325	Tur
0.571	0.107	3.886	0.4	4.66	4.815	3.817	Tur
0.613	0.127	2.986	1.002	3.771	1.812	4.102	Col
0.165	0.06	2.034	2.148	1.597	3.679	1.89	Tur
0.872	0.046	2.543	3.077	2.625	1.699	5.447	Tur
0.578	0.077	2.439	2.698	1.29	1.858	2.954	Afg
0.32	0.061	1.793	4.572	2.912	2.881	4.919	Afg
1.023	0.121	0.746	4.019	1.985	4.339	2.722	Afg
0.583	0.09	2.265	3.503	4.695	2.817	1.207	Afg
1.046	0.084	3.938	0.784	2.565	1.978	1.111	Afg
1.122	0.101	3.009	3.586	1.668	3.519	3.306	Mor

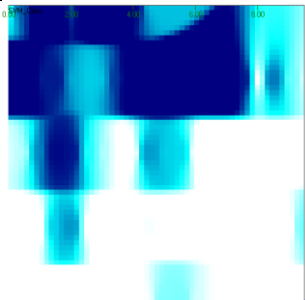
Obrázek 7 Příklad typických dat pro SVM-klasifikaci (zkrácená tabulka)

Protokol

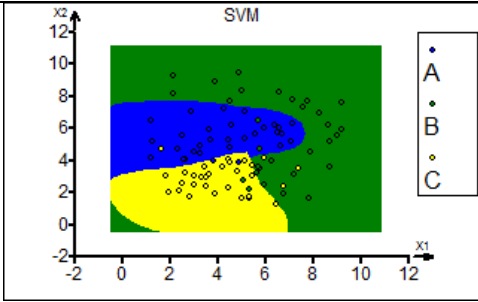
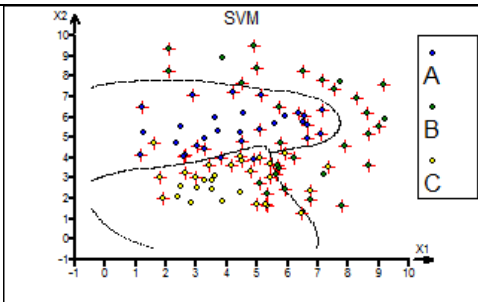
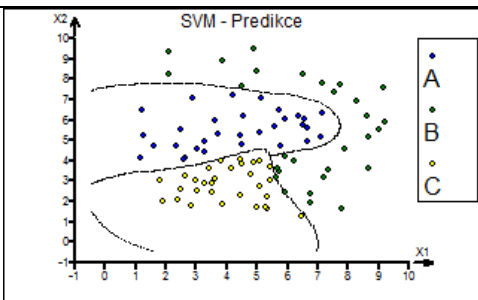
Název úlohy	Zadaný název úlohy
Nezávisle proměnná	Seznam použitých nezávisle proměnných
Závisle proměnná	Název sloupce s faktorem
Predikce	Sloupce použité pro predikci
Typ SVM	Zvolený typ SVM klasifikace, pomocí zadaného ztrátového

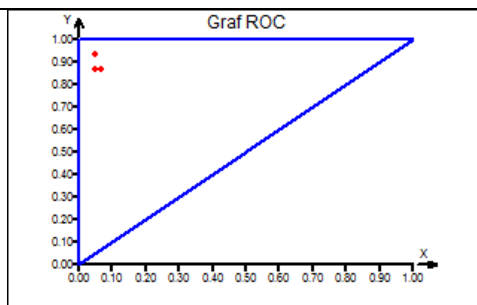
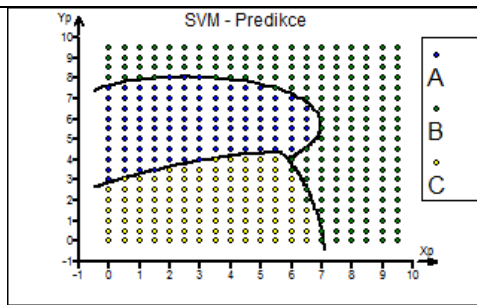
	koeficientu C <i>Klasifikace - Cost</i> , nebo podle předpokládaného podílu chybných klasifikací <i>Klasifikace - Nu</i>												
Jádro	Zvolený typ jádra: <i>Lineární: $(x_i * x_j)$</i> <i>Polynomické: $(-\gamma * (x_i * x_j) + r)^d$</i> <i>Radial Base Function: $\exp(-\gamma * x_i - x_j ^2)$</i> <i>Sigmoida: $\tanh(\gamma * (x_i * x_j) + r)$</i>												
Stupeň	Požadovaný stupeň polynomického jádra (je-li použito)												
Gamma	Zadaný parametr jádra γ												
Cost	Zadaný ztrátový koeficient pro optimalizaci C (pouze pro zvolený typ C)												
R	Zadaný parametr jádra R												
Nu	Zadaný parametr ν (pouze pro zvolený typ Nu)												
Shrinking	Zvolená možnost dodatečné stabilizace modelu pomocí techniky Shrinking												
Pravděpodobnosti	Volba, zda se mají počítat predikované pravděpodobnosti klasifikace (A_{no} / N_e)												
Primální parametry	<p>Počítají se pouze v případě lineární klasifikace. Hodnoty parametrů lineární hranice $\mathbf{w}^T \mathbf{x} + \beta = w_1 x_1 + w_2 x_2 + \dots + w_m x_m + \beta = 0$ mezi všemi dvojicemi úrovní faktoru v prostoru prediktorů X. Ve sloupci Beta je absolutní člen β, v dalších sloupcích jsou koeficienty k jednotlivým sloupcům prediktoru. Řádky tabulky jsou označeny dvojicemi hodnot (úrovní) faktoru, například „A – B“. Vyjde-li po dosazení nějakých hodnot prediktorů kladné číslo, leží tento bod na straně první z dvojice úrovní (v tomto příkladu „A“), vyjde-li záporné číslo, leží tento bod na straně druhé z dvojice úrovní. V níže uvedeném příkladu bude bod $(x=5; y=4)$ hodnota $\mathbf{w}^T \mathbf{x} = 5 * (-0.7061) + 4 * (-0.445) + 6.9123 = 1.602$, tedy pro bod $(5,4)$ predikujeme úroveň „A“.</p> <table border="1" data-bbox="502 1339 1133 1444"> <thead> <tr> <th>Primální parametry</th> <th>W</th> <th></th> <th></th> </tr> <tr> <td></td> <td>Beta</td> <td>X</td> <td>Y</td> </tr> </thead> <tbody> <tr> <td>A - B</td> <td>6.9123</td> <td>-0.7061</td> <td>-0.445</td> </tr> </tbody> </table> 	Primální parametry	W				Beta	X	Y	A - B	6.9123	-0.7061	-0.445
Primální parametry	W												
	Beta	X	Y										
A - B	6.9123	-0.7061	-0.445										
Tabulka	Souhrnný počet správně a chybně klasifikovaných dat (<i>Správně</i> ,												

misklasifikací	<i>Chybně</i>) a kontingenční tabulka misklasifikací, která uvádí počty správně a chybně klasifikovaných dat pro všechny úrovně faktoru. Ve spodním řádku tabulky jsou skutečné počty výskytů jednotlivých úrovní, v pravém sloupci jsou predikované počty výskytů jednotlivých úrovní. Jednotlivé hodnoty v tabulce pak vyjadřují počty predikovaných úrovní. Například hodnota „9“ v následující ilustraci znamená, že SVM-model predikuje 9-krát úroveň B pro řádek, který má ve skutečnosti úroveň A.																									
	<table border="1"> <thead> <tr> <th></th> <th>A</th> <th>B</th> <th>C</th> <th>Celkem</th> </tr> </thead> <tbody> <tr> <th>A</th> <td>19</td> <td>6</td> <td>1</td> <td>26</td> </tr> <tr> <th>B</th> <td>9</td> <td>17</td> <td>2</td> <td>28</td> </tr> <tr> <th>C</th> <td>2</td> <td>7</td> <td>27</td> <td>36</td> </tr> <tr> <th>Celkem</th> <td>30</td> <td>30</td> <td>30</td> <td>90</td> </tr> </tbody> </table>		A	B	C	Celkem	A	19	6	1	26	B	9	17	2	28	C	2	7	27	36	Celkem	30	30	30	90
	A	B	C	Celkem																						
A	19	6	1	26																						
B	9	17	2	28																						
C	2	7	27	36																						
Celkem	30	30	30	90																						
Efektivita klasifikace	Tato tabulka uvádí pro každou úroveň faktoru podíl chybných klasifikací do dané úrovně (false positives, FP), podíl správných klasifikací do dané úrovně (true positives, TP) a plochu pod bodem příslušné úrovně v grafu ROC (Receiver Operating Characteristic area under curve, ROC AUC), tedy $(TP*(1 - FP)+1/2*(TP*FP+(1 - TP)*(1 - FP)))$																									
Support Vectors	Pokud byla možnost vybrána v dialogovém panelu, vypíše se všechny použité support vektory.																									
Predikce	Tabulka predikovaných hodnot pro trénovací (původní zadaná) data.																									
Index	Pořadové číslo																									
Predikce	Predikovaná úroveň z modelu																									
Data	Skutečná pozorovaná úroveň z dat																									
Rezidua	0 v případě správné klasifikace, 1 v případě chybné klasifikace																									
Pravd. úrovně	Pravděpodobnost každé z úrovní faktoru pro dané hodnoty nezávisle proměnných. Má-li faktor například tři úrovně, lze snadno zkonstruovat (pomocí modulu <i>Grafy / 3D Spline</i>) rozložení pravděpodobnosti výskytu dané úrovně pro zvolenou dvojici nezávisle proměnné, jak ukazuje ilustrace: <div style="display: flex; justify-content: space-around; margin-top: 10px;">  </div>																									

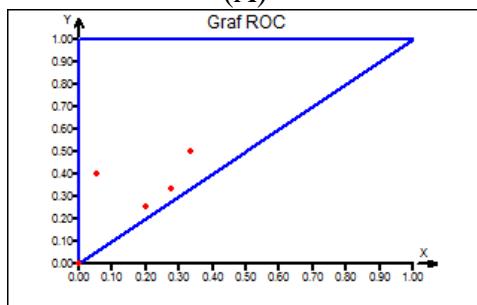
	
	pravděpodobnost A — pravděpodobnost B — pravděpodobnost C
Predikce	Pokud byla zvolena <i>Predikce</i> v dialogovém panelu, obsahuje tato tabulka predikované úrovně faktoru pro zadané hodnoty nezávisle proměnné včetně vypočítaných pravděpodobností. Jako predikce je vždy uvedena úroveň, která má nejvyšší pravděpodobnost (při více úrovních faktoru to může být i pravděpodobnost mnohem menší než 0.5)
Index	Pořadové číslo predikce
(nezávisle proměnné)	Hodnoty zadaných nezávisle proměnných
Predikce	Predikovaná (nejpravděpodobnější) úroveň faktoru
Pravd. úrovně	Predikované pravděpodobnosti jednotlivých úrovní faktoru

Grafy

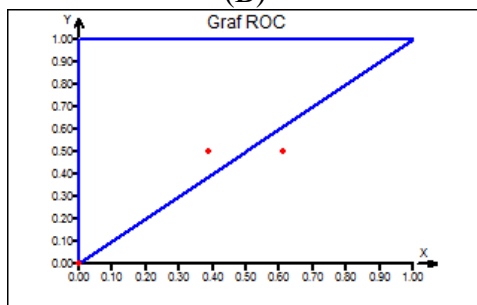
	Pokud je nezávisle proměnná dvourozměrná (má dva sloupce), konstruuje se graf predikčních oblastí pro jednotlivé úrovně. Každá barva reprezentuje oblast, v níž očekáváme výskyt příslušné úrovně faktoru častěji, než výskyt kterékoliv jiné úrovně. Je-li například počet úrovní faktoru $Y = \{„A“, „B“, „C“, „D“\}$ $m = 4$, je v oblasti predikované hodnoty „A“ pravděpodobnost výsledku „A“ alespoň $1/m$, tedy 0.25.
	Graf hranice predikčních oblastí je tvarově totožná s předchozím grafem, vyznačuje v případě dvourozměrné nezávisle proměnné hranice, na nichž jsou pravděpodobnosti sousedících úrovní stejné.
	Graf predikce vyznačuje v případě dvourozměrné nezávisle proměnné nové zadané hodnoty nezávisle proměnné pro predikci v grafu, pokud byla zvolena <i>Predikce</i> v dialogovém panelu.



(A)



(B)



(C)

Body ROC (Receiver Operating Characteristic) pro posouzení efektivity klasifikačního modelu pro každou úroveň faktoru. Na ose Y je podíl správných klasifikací do dané úrovně (true positives, TP), na ose X je podíl chybných klasifikací do dané úrovně (false positives, FP). Čím jsou body blíže levému hornímu rohu trojúhelníka (bodu [0, 1]), tím lepší je rozlišení této úrovně daným modelem, obr. (A). jsou-li body blízko úhlopříčce, je rozlišovací schopnost modelu špatná, obr. (B), body mimo trojúhelník, viz obr. (C), představují predikci horší, než náhodné hádání, model je zcela nepoužitelný. Je dobré mít na paměti, že špatné rozlišení nemusí vždy znamenat chybu v modelu, někdy úroveň faktoru na zvolených nezávisle proměnných prostě nezávisí, nebo ji rozlišit nelze.

SVM – Regrese

Menu:

QCExpert	SVM	SVM-Regrese
----------	-----	-------------

Data a parametry

Data jsou uspořádána do sloupců nezávisle proměnných a jednoho sloupce s hodnotami závisle proměnné. Nezávisle proměnné i závisle proměnná musí být číselné hodnoty. Každý řádek musí obsahovat platné hodnoty ve všech sloupcích. Neúplné řádky nejsou ve výpočtu použity. V dialogovém panelu (Obrázek 8) se označí jeden nebo více sloupců nezávisle proměnné a jeden sloupec závisle proměnné. Chceme-li počítat predikci, označíme ještě políčko *Predikce* a vybereme sloupce, pro něž chceme predikovat hodnotu závisle proměnné. Stiskem tlačítka „Doplň X“ lze označit sloupce odpovídající vybraným

nezávisle proměnným a predikce se pak počítá pro táž data, z nichž se vytvořil model. Ve skupině *Data* se zvolí chceme-li pro výpočet použít jen nějakou podmnožinu dat, či všechna data. Ve skupině *Typ SVM* zvolíme, zda se má použít varianta *Epsilon* se zadáním velikosti očekávané maximální chyby ϵ , nebo *Nu* se zadáním koeficientu ν . Zaškrtneme-li políčko *Zobrazit Support Vectors*, vypíší se do protokolu všechny support vektory modelu a rovněž se označí v grafu křížkem, pokud se graf konstruuje. To má obvykle orientační význam hlavně u lineárních modelů. V políčku *Popis* lze vybrat sloupec, který se pak použije k identifikaci jednotlivých bodů (řádků) v grafech. Dále se zvolí jádro pro transformaci z možností

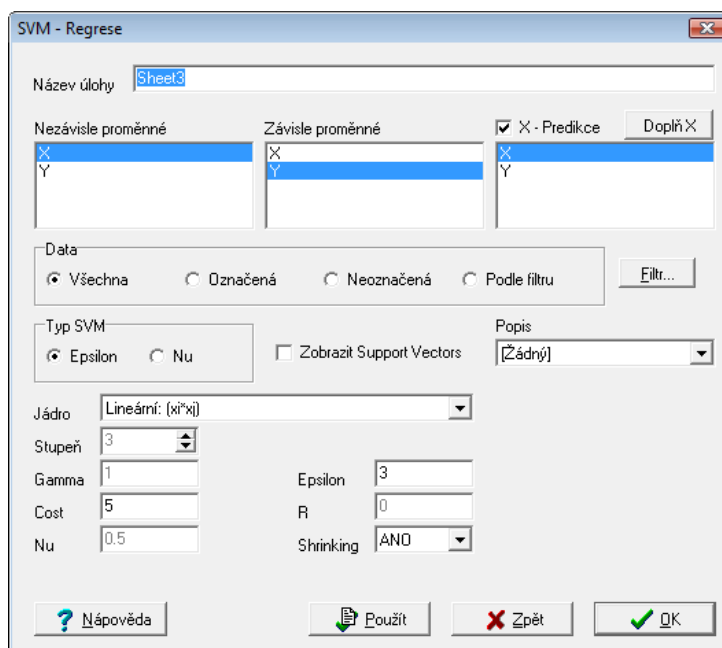
Lineární: $(x_i * x_j)$

Polynomické: $(-\gamma * (x_i * x_j) + r)^d$

Radial Base Function: $\exp(-\gamma * |x_i - x_j|^2)$

Sigmoida: $\tanh(\gamma * (x_i * x_j) + r)$

a parametry výpočtu: *Stupeň* (pouze u polynomického jádra), hodnotu parametru γ *Gamma*, který souvisí se strmostí jádra a tím i s nelinearitou modelu, tento parametr není nutno zadávat implicitně se použije hodnota $1/n$, ztrátový koeficient *Cost*, očekávaný podíl vybočujících hodnot ν *Nu*, očekávaná maximální chyba ϵ v poli *Epsilon* a parametr jádra *R*. Zvolíme, zda chceme provést závěrečnou korekci modelu po výpočtu (*Shrinking*, většinou nemá na výsledek velký vliv). Po stisku *OK* se zahájí výpočet.



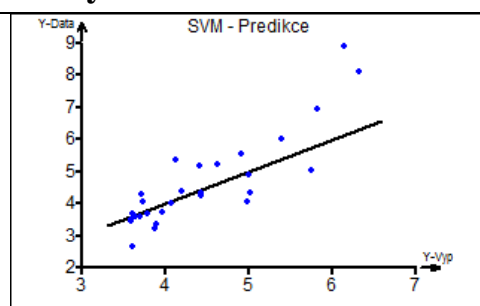
Obrázek 8 Dialogový panel SVM - Regrese

Protokol

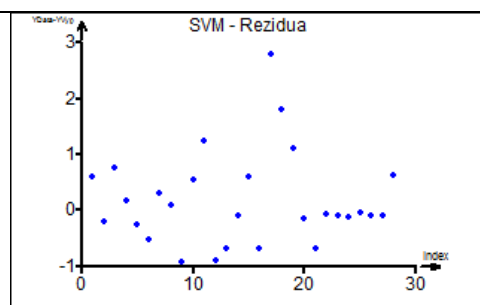
Název úlohy	Zadaný název úlohy
Nezávisle proměnná	Seznam použitých nezávisle proměnných
Závisle proměnná	Název sloupce se závisle proměnnou
Predikce	Sloupce použité pro predikci
Typ SVM	Zvolený typ SVM regrese, Regrese pomocí zadaného <i>Epsilon</i> podle rov. (0-4) , nebo Regrese podle podílu chybných dat <i>Nu</i> podle (0-5)
Jádro	Zvolený typ jádra: <i>Lineární:</i> $(x_i * x_j)$

	<i>Polynomické:</i> $(-\gamma * (x_i * x_j) + r)^d$ <i>Radial Base Function:</i> $\exp(-\gamma * x_i - x_j ^2)$ <i>Sigmoida:</i> $\tanh(\gamma * (x_i * x_j) + r)$
Stupeň	Požadovaný stupeň polynomického jádra
Gamma	Zadaný parametr jádra γ
Cost	Zadaný ztrátový koeficient pro optimalizaci C
Epsilon	Zadaná hodnota maximální chyby ϵ (pouze pro <i>Typ SVM = Epsilon</i>)
R	Zadaný parametr jádra R (pouze pro polynom a tanh)
Nu	Zadaný parametr ν (pouze pro <i>Typ SVM = Nu</i>)
Shrinking	Zvolená možnost dodatečné stabilizace modelu pomocí techniky Shrinking
Support Vectors	Pokud byla tato možnost vybrána v dialogovém panelu, vypíší se všechny použité support vektory.
Predikce	Tabulka predikovaných hodnot pro trénovací data.
Index	Pořadové číslo
Predikce	Predikovaná hodnota závisle proměnné z modelu
Data	Skutečná pozorovaná hodnota závisle proměnné z tabulky dat
Rezidua	
Predikce	Pokud byla zvolena <i>Predikce</i> v dialogovém panelu, obsahuje tato tabulka predikované hodnoty závisle proměnné pro zadané hodnoty nezávisle proměnné.
Index	Pořadové číslo predikce
(nezávisle proměnné)	Hodnoty zadaných nezávisle proměnných
Predikce	Predikovaná hodnota závisle proměnné z modelu

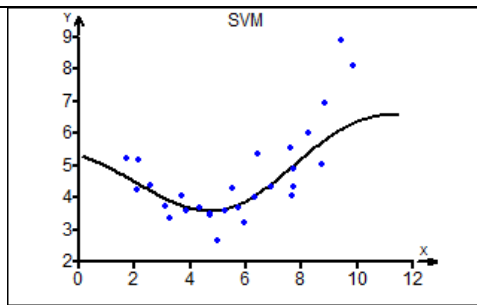
Grafy



Graf predikce SVM Regrese vynáší na vodorovnou osu vypočtenou (predikovanou) hodnotu Y -Vyp závisle proměnné, na svislou osu skutečnou hodnotu Y -Data z datové tabulky. Čím blíže k přímce leží body, tím těsněji se model přibližuje k datům.



Graf reziduí, svislé vzdálenosti od přímky v předchozím grafu, rozdíl mezi skutečnou a predikovanou hodnotou závisle proměnné. Tvary a interpretace tohoto grafu jsou ilustrovány v odst. 0 – Příklady.



Pokud je nezávisle proměnná jednorozměrná (jeden sloupec), zobrazí se regresní funkce, tedy skutečný SVM-regresní model a data.

SVM – Hustota rozdělení

Menu:	QCExpert	SVM	SVM-Hustota
-------	----------	-----	-------------

Data a parametry

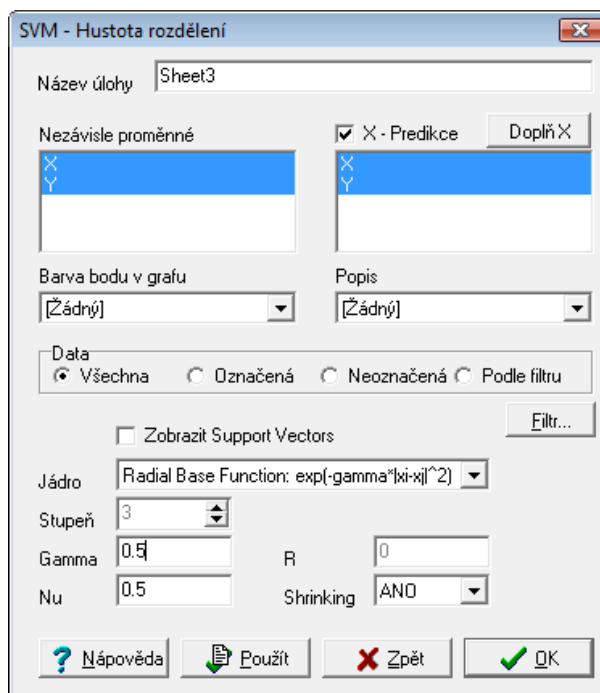
Data jsou uspořádána do jednoho nebo více sloupců proměnných. Proměnné jsou číselné hodnoty. Každý řádek musí obsahovat platné hodnoty ve všech sloupcích. Neúplné řádky nejsou použity. V dialogovém panelu (Obrázek 9) se označí jeden nebo více sloupců proměnné. Chceme-li počítat predikci, označíme ještě políčko *Predikce* a vybereme sloupec, pro něj chceme predikovat příslušnost k rozdělení. Stiskem tlačítka „Doplň X“ lze označit sloupce odpovídající vybraným proměnným a predikce se pak počítá pro táž data, z nichž se vytvořil model. Ve skupině *Data* se zvolí chceme-li pro výpočet použít jen nějakou podmnožinu dat, či všechna data. Zaškrtneme-li políčko *Zobrazit Support Vectors*, vypíší se do protokolu všechny support vektory modelu a rovněž se označí v grafu křížkem, pokud se graf konstruuje. V políčku *Popis* lze vybrat sloupec, který se pak použije k identifikaci jednotlivých bodů (řádků) v grafech. Dále se zvolí jádro pro transformaci z možností

Lineární: $(x_i * x_j)$

Polynomické: $(-\gamma * (x_i * x_j) + r)^d$

Radial Base Function: $\exp(-\gamma * |x_i - x_j|^2)$

Sigmoida: $\tanh(\gamma * (x_i * x_j) + r)$



Obrázek 9 Dialogový panel SVM – Hustota rozdělení

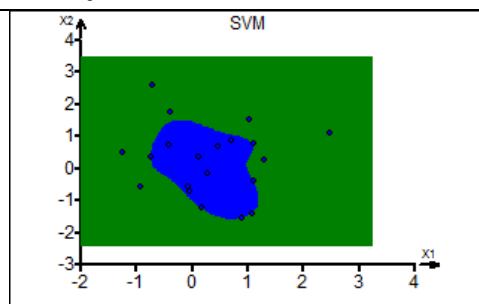
a parametry výpočtu: *Stupeň* (pouze u polynomického jádra), hodnotu parametru γ *Gamma*, který souvisí se strmostí jádra a tím s nelinearitou modelu, tento parametr není nutno zadávat implicitně se použije hodnota $1/n$, podíl dat mimo rozdělení v *Nu* a parametr jádra *R*. Zvolíme, zda chceme provést závěrečnou korekci modelu po výpočtu (*Shrinking*, většinou nemá na výsledek vliv. Po stisku *OK* se zahájí výpočet.

Protokol

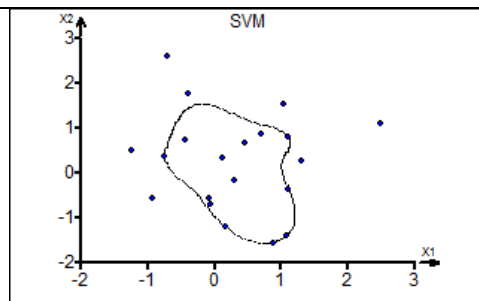
Název úlohy	Zadaný název úlohy
Nezávisle proměnná	Seznam použitých nezávisle proměnných
Predikce	Sloupce použité pro predikci
Typ SVM	Zvolený typ SVM klasifikace, pomocí zadaného ztrátového koeficientu <i>C Klasifikace - Cost</i> , nebo podle předpokládaného podílu chybných klasifikací <i>Klasifikace - Nu</i>
Jádro	Zvolený typ jádra: <i>Lineární: $(x_i * x_j)$</i> <i>Polynomické: $(-gamma * (x_i * x_j) + r)^d$</i> <i>Radial Base Function: $exp(-gamma * x_i - x_j ^2)$</i> <i>Sigmoida: $tanh(gamma * (x_i * x_j) + r)$</i>
Stupeň	Požadovaný stupeň polynomického jádra (je-li použito)
Gamma	Zadaný parametr jádra γ
Nu	Zadaný parametr <i>v</i> (pouze pro zvolený typ <i>Nu</i>)
R	Zadaný parametr jádra <i>R</i>
Shrinking	Zvolená možnost dodatečné stabilizace modelu pomocí techniky <i>Shrinking</i>
Support Vectors	Pokud byla vybrána tato možnost v dialogovém panelu, vypíší se všechny použité support vektory.

Predikce	Predikovaná příslušnost trénovacích (původních) dat k rozdělení při daném ν (Nu)
Index	Pořadové číslo
Predikce	Příslušnost k rozdělení při daném ν (Nu), 1 = uvnitř rozdělení, -1 = vně rozdělení.
Predikce	Pokud byla zvolena <i>Predikce</i> v dialogovém panelu, obsahuje tato tabulka predikované příslušnosti zadaných hodnot proměnné k rozdělení při daném ν (Nu), 1 = uvnitř rozdělení, -1 = vně rozdělení.
Index	Pořadové číslo
proměnné	Zadané hodnoty proměnné
Predikce	Predikce příslušnosti k rozdělení při daném ν (Nu)

Grafy



Graf oblasti příslušnosti k rozdělení při daném ν (Nu), tento graf se konstruuje pouze pro 2-rozměrnou proměnnou.



Graf hranice oblasti příslušnosti k rozdělení při daném ν (Nu), tento graf se konstruuje pouze pro 2-rozměrnou proměnnou.