SVM - Support Vector Machines

Menu:	QCExpert	SVM

The SVM, or Support Vector Machines employs a machine learning method developed in 1990's by Vladimir Naumovich Vapnik (Soviet Academy of Science, Stanford University, Royal Holloway College London, AT&T Bell Labs New Jersey, NEC Labs Princeton, Columbia University New York and Alexey Jakovlevich Chervonenkis. This method was formally used primarily as a classification tool but later was adapted also for regression and distribution density modeling. SVM models make use of the theory of empirical risk *R* and Vapnik-Chervonenkis (VC) dimension of the model. It has been proven that the following inequality holds with probability $(1 - \eta)$:

$$R(\alpha) \le R_{emp}(\alpha) + \sqrt{\frac{h\left(\ln\left(2l/h\right)+1\right) - \ln\left(\eta/4\right)}{l}}, \qquad \textbf{(0-1)}$$

where $R(\alpha) = \int \frac{1}{2}\left|y - f(\mathbf{x},\alpha)\right| p(\mathbf{x},y)\, d\mathbf{x}\, dy$ is risk (or actual mean error of the model), *l* is number of data rows, α is the model parameters vector, $R_{emp}(\alpha) = \frac{1}{2l}\sum_{i=1}^{l}\left|y_i - f(x,\alpha)\right|$ is empirical risk and *h* is non-negative integer VC-dimension of the model. The last term on the right-hand side (the square root) is called VC-confidence. The following text is a brief and simplified description. For more detailed information, we refer the reader to some of the resources given at the end of this chapter.

SVM-C – SVM Classification models

SVM models minimize suitably defined error (misclassification rate in classification or deviation in some metric in regression). For example, in classification of a linearly separable task in two dimensions (with two independent numerical variables and one two-level factor response variable defining one of two classes such as "A" and "B" for each value) we look for a line which separates (discriminates) both classes and allows for maximal distance of the different classes from the separating line thus generally minimizing risk of misclassification for any new data, see Fig. 1. The SVM model can then be used to predict the class from a given set of independent variable values including probabilities for each class.

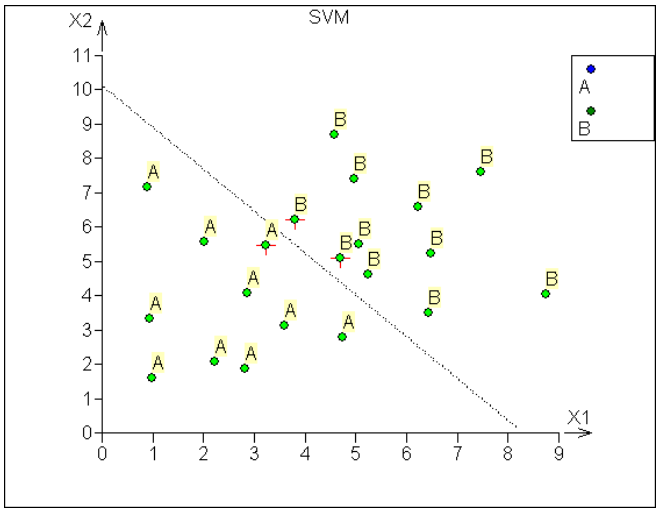


Fig. 1 SVM, separable data, linear model

In a non-separable case (like that on Fig. 2), a line is sought that minimizes the misclassification "distance" of misclassified (or incorrectly classified) data. On Fig. 2 the separating line minimizes the sum of distances of incorrectly classified point "A" and one incorrectly classified point "B" from the separating line and maximizes distance of the correctly classified data from the separating line.

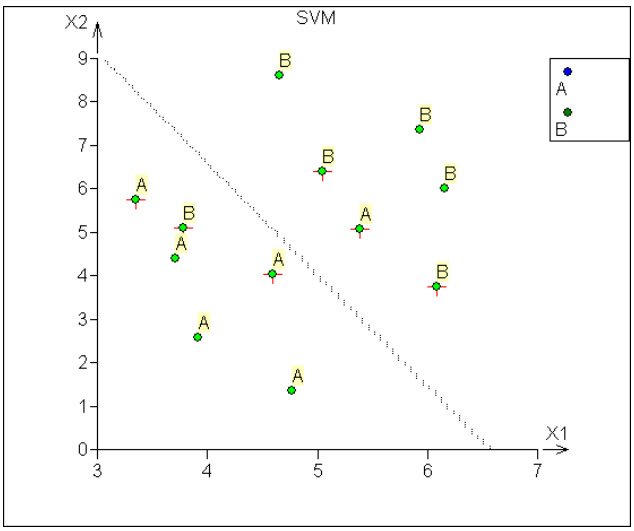


Fig. 2 SVM, linearly non-separable data, linear model

In case of separable data with binary response ($y_i = -1$ or 1) the length of the normal vector **w** to the separation line (or generally separation hyper plane) is minimized:

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

which maximizes the width of the gap between the two classes (lines H_1 and H_2 in Fig. 3). In the case of non-separable data, a term for misclass penalization with a user-defined tuning "cost" parameter C is added.

$$f(x) = \frac{\|\mathbf{w}\|^2}{2} + C\sum_{i=1}^{l}\xi_i \text{ subject to } \begin{matrix} y_i(\mathbf{w}\cdot\mathbf{x}_i + b) \geq 1 - \xi_i, i = 1,...,l \\ \xi_i \geq 0, i = 1,...,l \end{matrix}. \quad \textbf{(0-2)}$$

Geometrical interpretation of non-separable case is illustrated on Fig. 3. The points that lie on (or "support") the separation zone lines H_1 and H_2 are called "support vectors" – hence the name of the whole method. The support vectors are circled on Fig. 3.

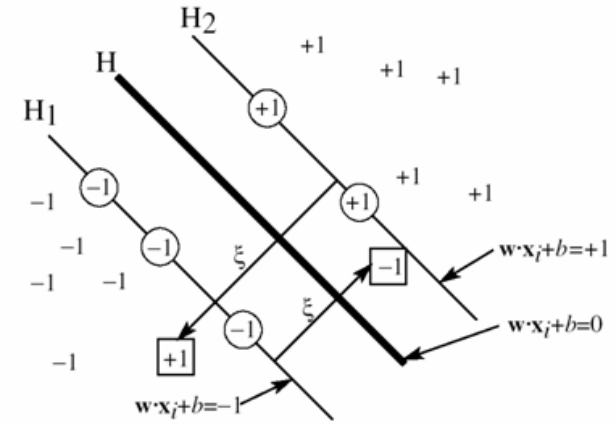


Fig. 3 Separation by SVM-hyper plane H for non-separable data

Alternatively, instead of the loss coefficient C, a ratio v ($0 < v < 1$) may be employed,

$$f(x) = \frac{\|\mathbf{w}\|^2}{2} - v\rho + \frac{1}{l}\sum_{i=1}^{l}\xi_i \text{ subject to } \begin{matrix} y_i(\mathbf{w}\cdot\mathbf{x}_i + b) \geq \rho - \xi_i, i = 1,...,l \\ \xi_i \geq 0, i = 1,...,l; \rho \geq 0 \end{matrix}. \quad \textbf{(0-3)}$$

v corresponds to an expected ratio of misclassified cases.

SVM-R – SVM Regression models

The approach in SVM-Classification was extended to regression by defining a new criterion containing an acceptable error of a regression model ε. Points outside this interval are penalized with a linear loss function with the loss coefficient $C > 0$. The model should then minimize

$$\frac{\|w\|^2}{2} + C\sum_{i=1}^{l}\left(\xi_i^+ + \xi_i^-\right) \text{ subject to } \begin{matrix} \mathbf{w}\cdot\mathbf{x}_i + b - y_i \leq \varepsilon + \xi_i^+, i = 1,...,l \\ y_i - \mathbf{w}\cdot\mathbf{x}_i + b \leq \varepsilon + \xi_i^-, i = 1,...,l \end{matrix} \quad \textbf{(0-4)}$$

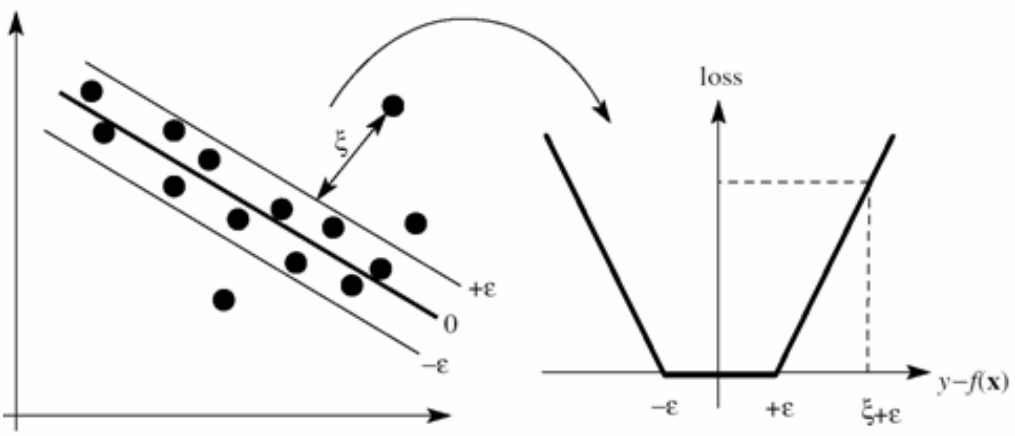
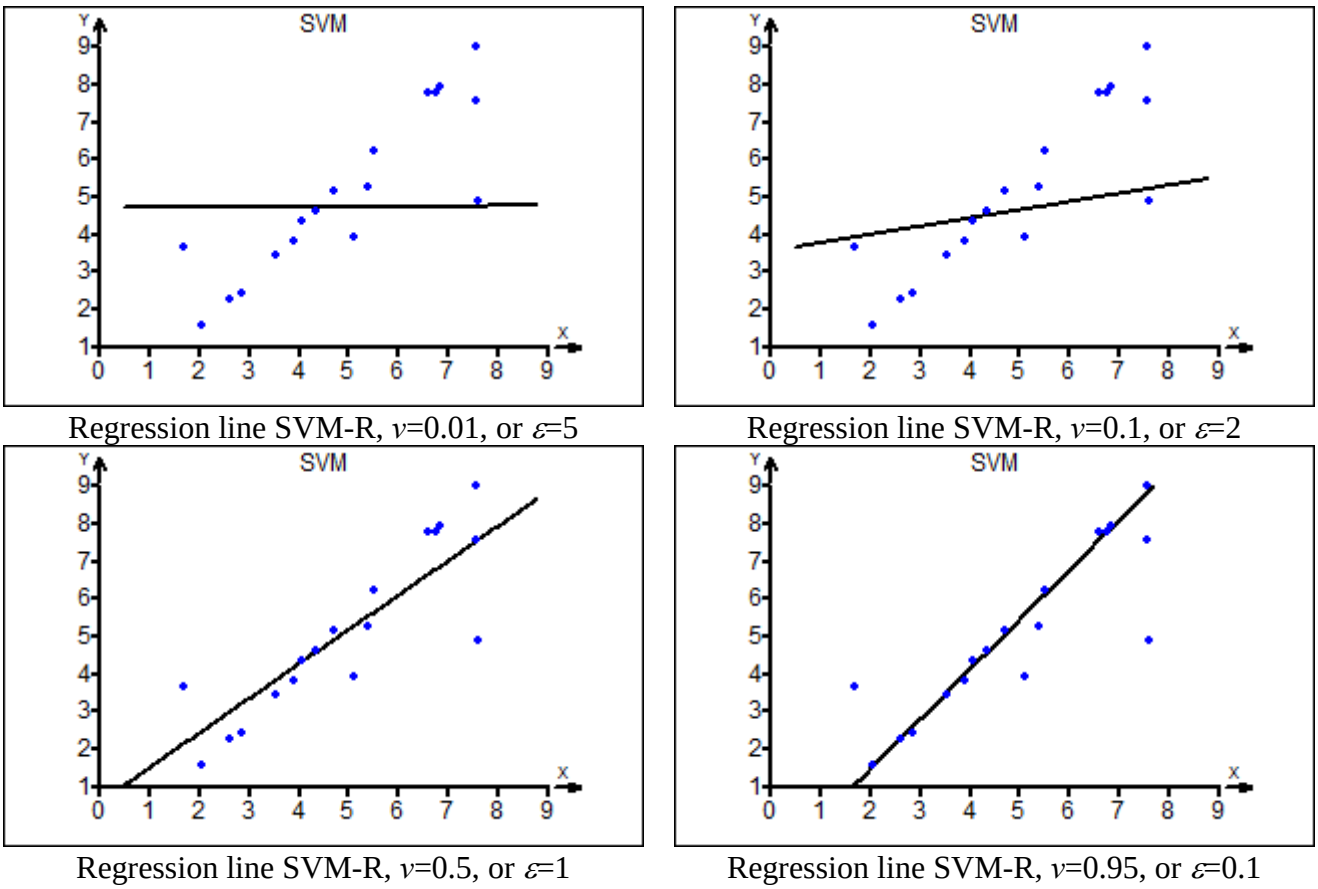


Fig. 4 Principle of SVM regression (SVM-R)

The coefficient vector **w** corresponds to regression parameters of a linear model, b is the absolute term. The user-tuned parameter ε is the half-width of the band in which errors are acceptable, ε is the maximal acceptable absolute error. This criterion has some "robustness" in it since it forces the model to "squeeze" as much data as possible into a narrow band $f(x) \pm \varepsilon$ around the model and discard the data that do not fit in the band. This makes SVM regression a suitable alternative to robust regression methods in case of heavily contaminated data. The criterion (0-4) can be rewritten using parameter v ($0<v<1$), corresponding to the probability of a given point to lie inside the acceptable region $f(x) \pm \varepsilon$. The resulting criterion can be written as a constrained minimization with respect to **w**, b, ξ, ξ^*, ε:

$$f(x) = \frac{\|\mathbf{w}\|^2}{2} + C\left(v\varepsilon + \frac{1}{l}\sum_{i=1}^{l}\left(\xi_i + \xi_i^*\right)\right) \text{ subject to } \begin{matrix} (\mathbf{w}\cdot\mathbf{x}_i + b) - y_i \leq \varepsilon + \xi_i \\ y_i - (\mathbf{w}\cdot\mathbf{x}_i + b) \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, i = 1,...,l; \varepsilon \geq 0 \end{matrix} \quad \textbf{(0-5)}$$

As with any regression, the regression model can be used to predict expected value of the independent (response) variable for given values of the independent (predictor) variables. The following plot shows the effect on the parameter ε on the regression model in a univariate case and the effect of robustness.



Regression line SVM-R, v=0.01, or ε=5

Regression line SVM-R, v=0.1, or ε=2

Regression line SVM-R, v=0.5, or ε=1

Regression line SVM-R, v=0.95, or ε=0.1

Fig. 5 Effect of loss coefficient on robustness of the SVM-R regression model for C = 1

SVM-OneClass – Distribution density

The SVM can be used to find a boundary of a model-free distribution in a multivariate sample space using the criterion (B. Schoelkopf et al., 2001)

$$\frac{\|w\|^2}{2} - \rho + \frac{1}{\nu l} C \sum_{i=1}^{l} (\xi_i) \text{ subject to } \begin{matrix} \mathbf{w} \cdot \mathbf{x}_i \geq \rho - \xi_i, i = 1,...,l \\ \xi_i \geq 0, i = 1,...,l \end{matrix} \qquad \textbf{(0-6)}$$

This boundary is again a linear hyper plane in **x** and corresponds approximately to a ν-quantile of the distribution. The coefficient v, $0 < \nu \leq 1$ can be thought of as the ratio of the "outliers" in the data sample that will be outside of the distribution (and thus will not influence the model). The user however does not label any specific points to be excluded from the model. The model defines the "correct data" that belong to the distribution and can also be used to decide whether new points belong to this distribution. Robustness of this method makes it suitable for detecting stability and change points in complex multivariate processes with any type of distribution.

SVM-kernel transformations

The above described methods lead only to linear models (discrimination, regression) of type $\mathbf{w} \cdot \mathbf{x}$ and as such can hardly be too much useful. One of major achievements of SVM theory is implementation of transformation of l-dimensional sample space $\mathfrak{R}^l$ spanned by **x** with a system of nonlinear functions $\boldsymbol{\varphi}(\mathbf{x})$, into a new, n-dimensional space $\mathfrak{L}^n$. Dimensionality n of $\boldsymbol{\varphi}(\mathbf{x})$, has generally no connection to dimensionality of data l, Typically, $n > l$ and n can also be infinite. The linear SVM model is created in the new space. Since the relationship between $\mathfrak{R}$ and $\mathfrak{L}$ is non-linear, the linear SVM models created in $\mathfrak{L}$

are nonlinear in $\mathcal{K}$. This gives SVM models a remarkable flexibility. If transformations are defined using quadratic forms

$$K(\mathbf{x}_i, \mathbf{x}_j) = \boldsymbol{\varphi}(\mathbf{x}_i)^T \boldsymbol{\varphi}(\mathbf{x}_j),$$

(where K is a kernel function), the optimization tasks can be formulated as a convex quadratic constrained optimization which can be effectively solved with use of Lagrange multipliers.

The most commonly used kernel functions are RBF type functions (Radial Base Functions) defined as

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \left\| \mathbf{x}_i - \mathbf{x}_j \right\|\right) \qquad \textbf{(0-7)}$$

Further often used transformations are:

Polynomial kernel $\quad K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(\gamma\, \mathbf{x}_i{}^T \mathbf{x}_j + r\right)^d; \quad \gamma > 0$

Sigmoid kernel $\quad K(\mathbf{x}_i, \mathbf{x}_j) = \tanh\left(\gamma\, \mathbf{x}_i{}^T \mathbf{x}_j + r\right)$

Linear kernel $\quad K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i{}^T \mathbf{x}_j$

Parameters γ and d are set by the user, r is computed. The parameter γ is the steepness of the kernel. Higher values of γ give generally more detailed (often also overdetermined and less stable) models. With use of kernel transformations, highly non-linear models can be created to describe the data **x**. Stability and prediction capability can be diagnosed with some validation tool,

Examples

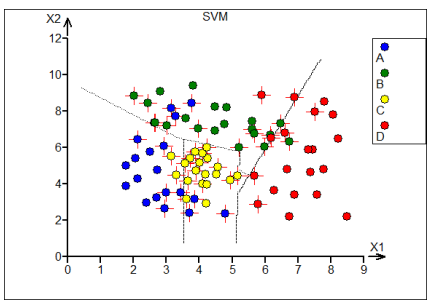
Here we provide several simple examples to illustrate common SVM models and use and sense of the parameters. Despite the fact that SVM are usually employed in high-dimensional problems and rather extensive data sets, we restrict ourselves to two-dimensional small samples for easier visualization.

Example 1 – Classification

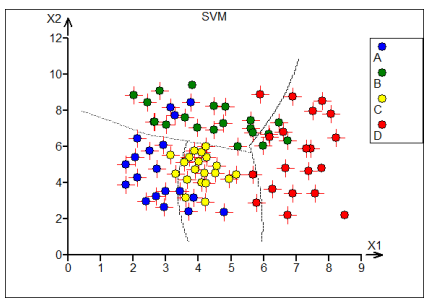
For two continuous variables, *X* and *Y* we have four possible categorial outputs: A, B, C, D. The different levels (values) of the categorial variable are not linearly separable in the plane *X*, *Y*. This example shows the difference between linear and RBF-transformed SVM classification model. The model is trained on the data shown at the figures below. The plots show the separating hyper planes (in this case ordinary lines) for the linear model (the first plot) and separating non-linear hypersurfaces (in this case curves) for the RBF-SVM models with different value of parameter γ from γ=0.01 to γ=10. Models are based on (0-2) and (0-3). Misclass is he number of incorrectly classified cases. Too big value of γ will result in overdetermined models strongly dependent on the particular training data.

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

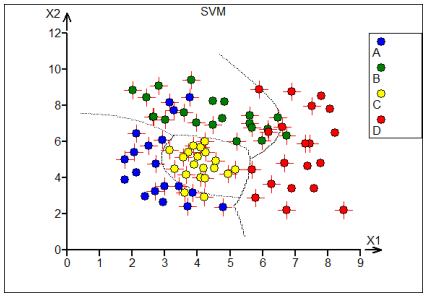
Example of data table



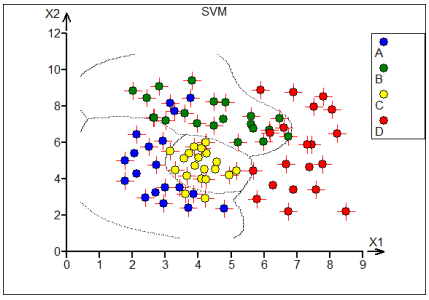
Linear, Misclass=15



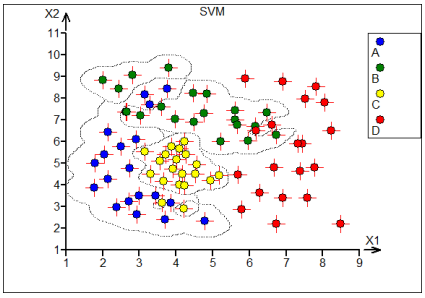
RBF, γ=0.01, Misclass=14



RBF, γ=0.1, Misclass=8



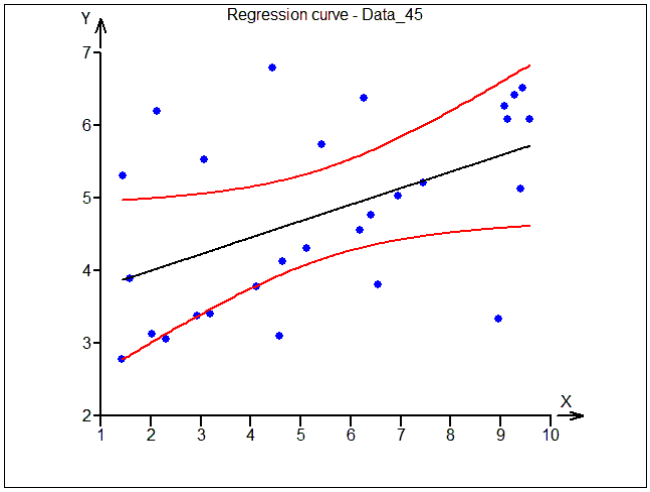
RBF, γ=1, Misclass=5



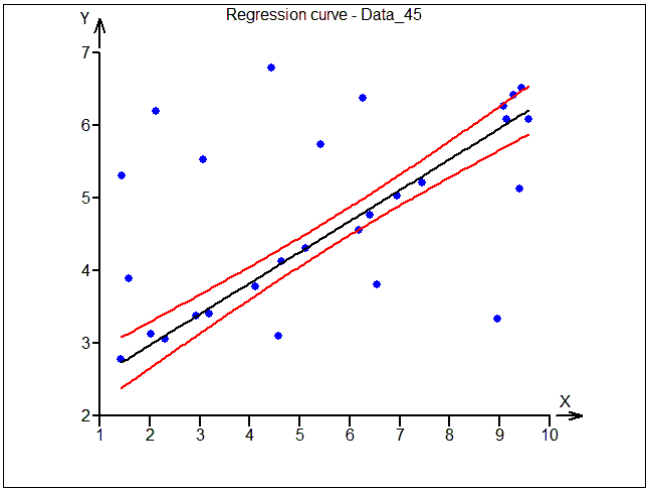
RBF, γ=10, Miclass=1

Example 3 – Classical Robust and SVM-ε regression

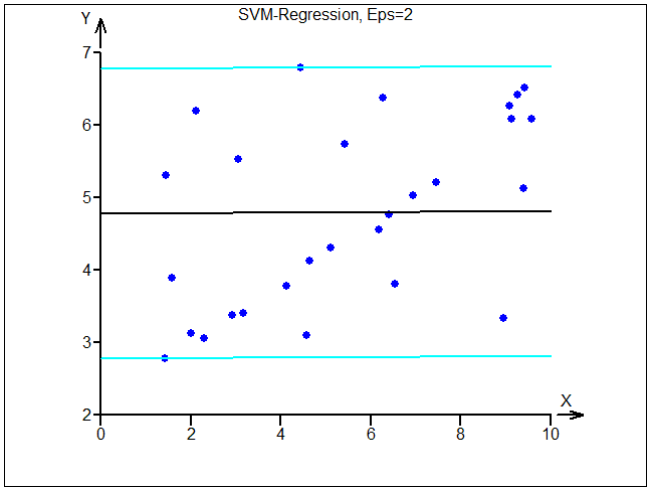
The parameter ε in eq. (0-4) sets the width of an acceptable band around the regression model, as illustrated in Fig. 4. Decreasing this parameter at a constant value of γ will increase robustness of the model against outlying values with respect to the regression model $f(x)$. In SVM-regression, the data points outside the interval $\langle f(x)-\varepsilon; f(x)+\varepsilon \rangle$ are considered outliers. With decreasing ε, we can thus obtain models in a certain sense similar to robust regression (like regression M-estimates) which may be used to detect outliers and to filter contaminated data. The following plots illustrate behavior of classical regression and SVM regression with varying ε and γ. SVM tries to "squeeze" as much data as possible into $f(x) \pm \varepsilon$. The sufficiently low parameter γ prevents the model to "go through all points", as is (nearly) the case on the plot (J) below.



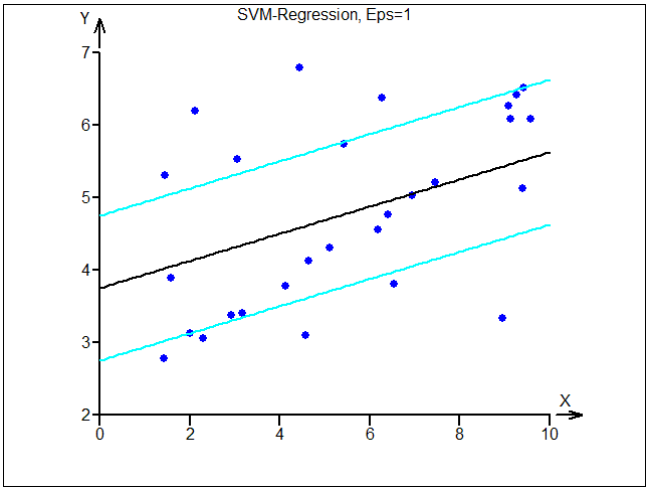
(A) Classical regression, least squares



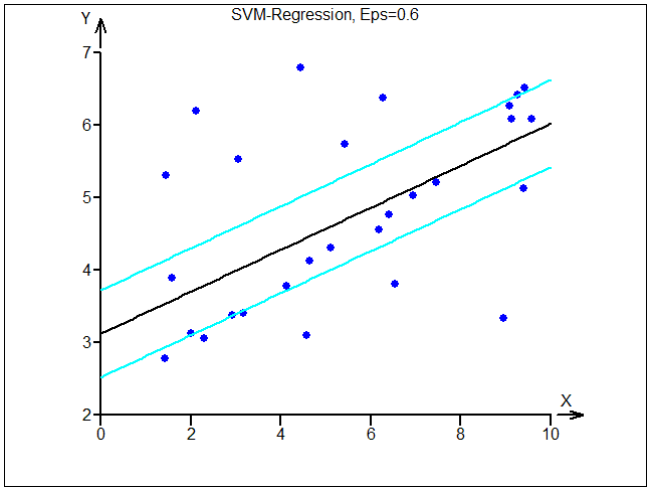
(B) Classical regression, robust BIR method



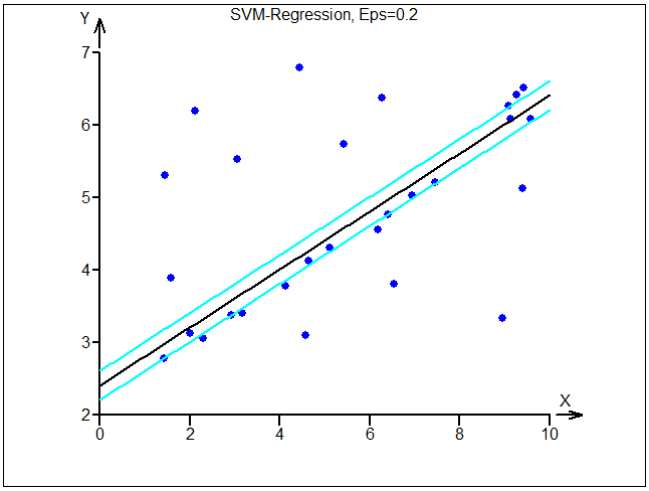
(C) SVM-R, ε = 2



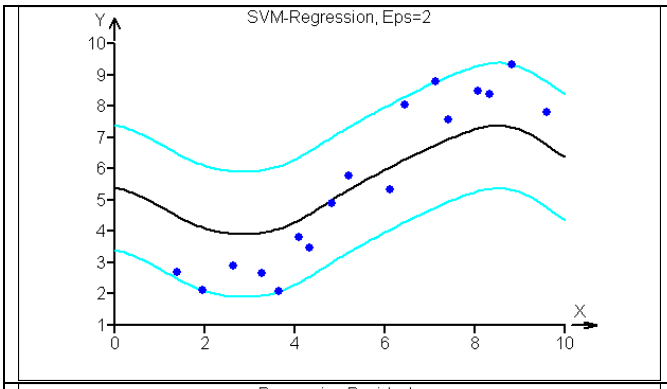
(D) SVM-R, ε = 1

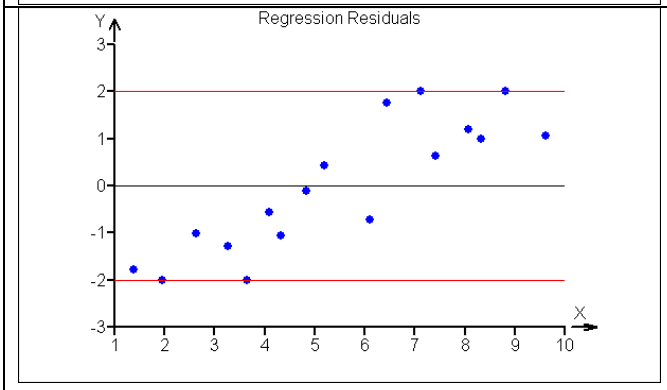


(E) SVM-R, ε = 0.5

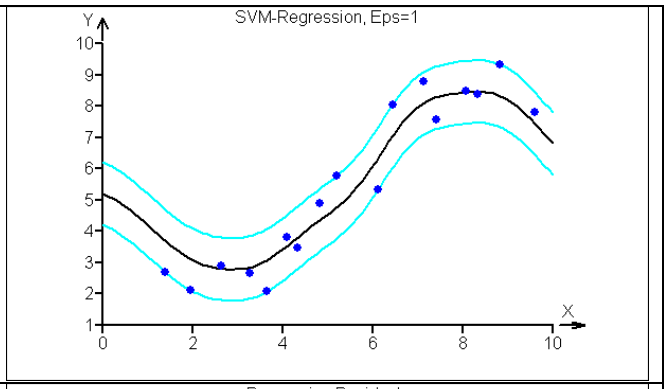


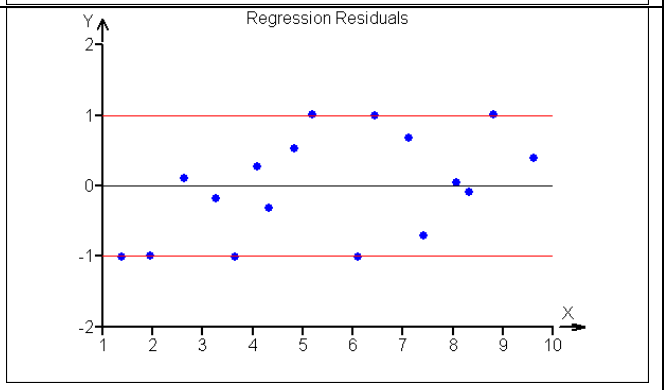
(F) SVM-R, ε = 0.2



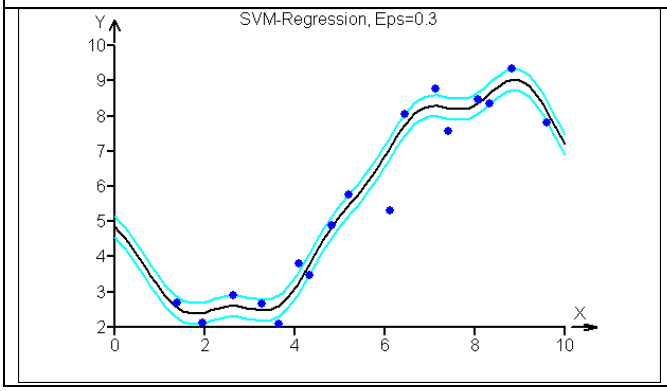


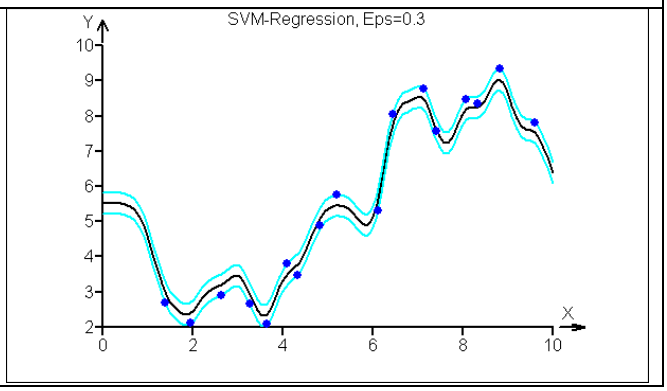
(G) RBF kernel, ε=2; γ=0.5, cost=10
(Too loose model – too big ε)

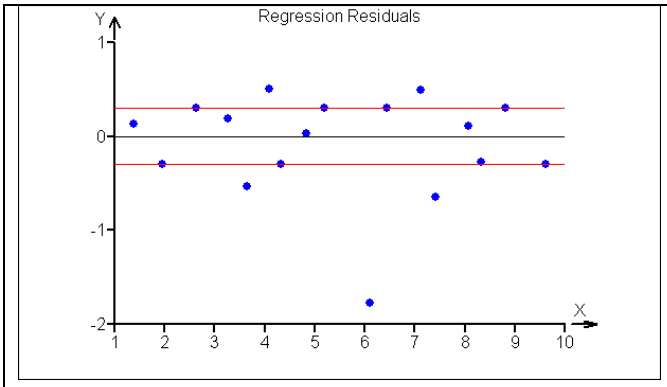




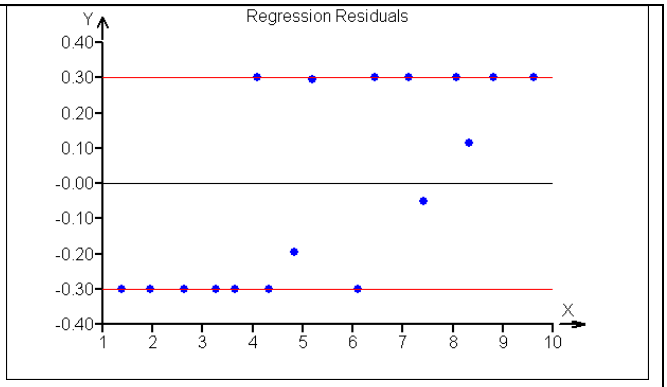
(H) RBF kernel, ε=1; γ=0.5, cost=10
(satisfies all data)







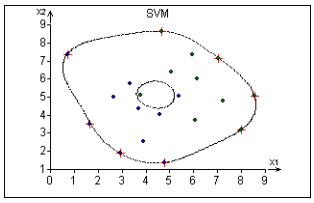
(I) RBF kernel, ε=0.3; γ=0.5, cost=10 (identifies suspicious outlier)



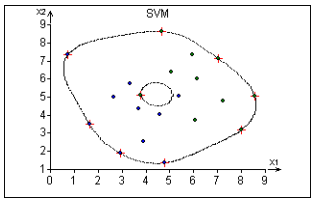
(J) RBF kernel, ε=0.3; γ=5, cost=10 (useless model, too big γ – just copies the data)

Example 5 – Unsupervised learning, distribution density, influence of γ and ν

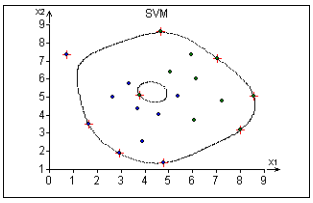
The following table of plots illustrates the influence of γ (can be viewed as "stiffness") and ν (ratio of the "discarded" part of the distribution), roughly said - the model will describe 100(1- ν)% of the distribution with highest density. Observe the following plots to understand the role of the two parameters.



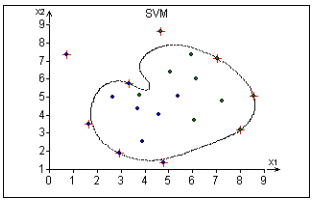
γ = 0.1, ν = 0.1



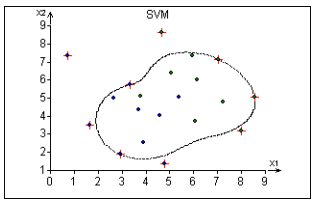
γ = 0.1, ν = 0.2



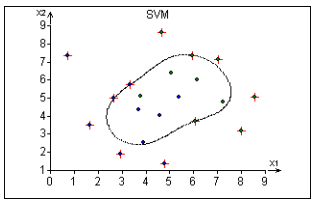
γ = 0.1, ν = 0.3



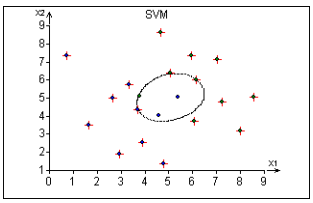
γ = 0.1, ν = 0.4



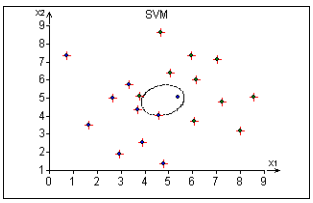
γ = 0.1, ν = 0.5



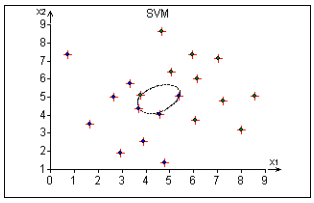
γ = 0.1, ν = 0.6



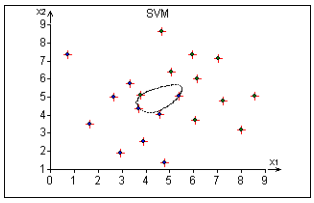
γ = 0.1, ν = 0.8



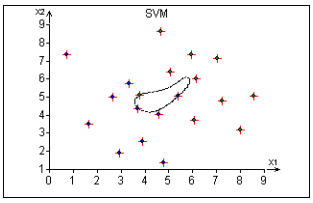
γ = 0.1, ν = 0.9



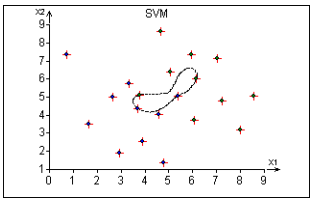
γ = 0.2, ν = 0.9



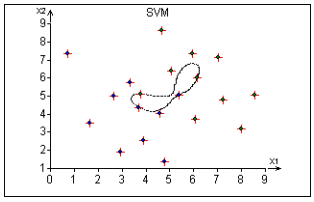
γ = 0.3, ν = 0.9



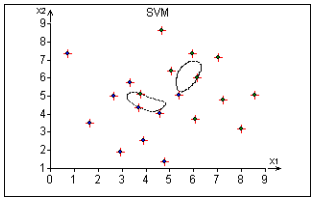
γ = 0.4, ν = 0.9



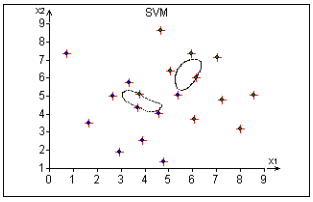
γ = 0.5, ν = 0.9



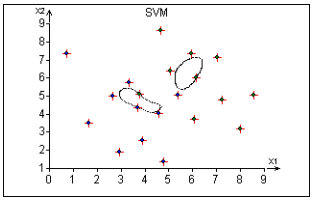
γ = 0.6, ν = 0.9



γ = 0.7, ν = 0.9



γ = 0.8, ν = 0.9



γ = 0.9, ν = 0.9

The SVM implementation in QCExpert is based on the code LIB-SVM (c) by Chih-Chung Chang and Chih-Jen Lin developed on National Taiwan University, see Chih-Chung Chang and Chih-Jen Lin: Library for Support Vector Machines

References

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). ["Vintage" book - 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 – Classification

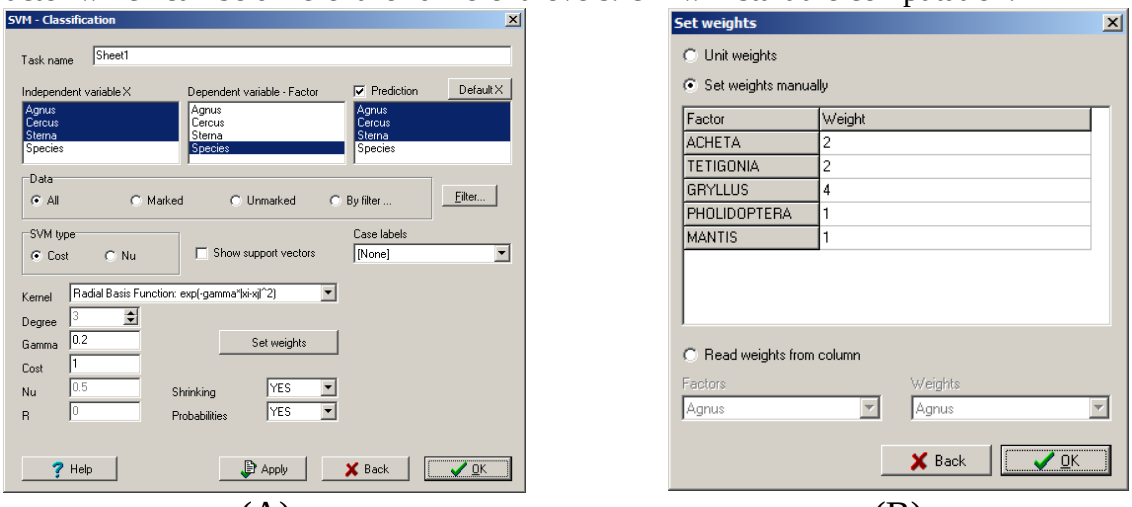
Menu:	QCExpert	SVM	SVM-Classification
DARWin:	**SVMC**		

Data and parameters

Data are expected in two or more columns. One column contains the factor levels (numerical or text). The other columns (the independent variables, or predictors) are numerical. The factor column (dependent variable, or response) must have at least two different values. The independent and dependent variables are selected in the dialog window (Fig. 6 A). If prediction is required, the independent variables for prediction are selected in the *Prediction* field. The number of predictors must be the same in *Independent variable* and *Prediction* fields. The button *Default X* will select the same predictors as selected in *Independent variables* field. In the *SVM Type* field select the type of the model criterion: *C* in (0-2) or *v* in (0-3) (*Cost* or *Nu*). Checking the checkbox *Show Support* vectors will mark supporting points in 2d-plots (the 2d plots are constructed only in the case of 2 independent variables). Column selected in *Case labels* field will be used to label points in plots. Select the required SVM kernel transformation in the *Kernel* field:

*Linear: (xi*xj)*	*Polynomial: (-gamma*(xi*xj)+r)^d*
Radial Base Function: exp(-gamma|xi-xj|^2)*	*Sigmoid: tanh(gamma*(xi*xj)+r)*

and select further SVM parameters: *Degree* (degree of the polynomial in case of *Polynomial kernel* – recommended value 2 or 3), *Gamma*, *Cost* (in case of Cost method), *Nu* (in case of Nu method), R (in polynomial or sigmoid kernel). Choose if the shrinkage correction is to be used. If *YES* is chosen in the *Probabilities* field classification probabilities are computed for all data rows and all existing levels of the response factor. Individual weights for each factor level can be set in the child dialog window invoked by clicking on the *Set weights* button (Fig. 6 B). By default, all weights are set to 1 (*Unit weights*). If there are many levels in the response factor, it may become feasible to read the weights from a pre-prepared separate column in the data sheet by checking *Read weights from column* and selecting the column. The weights can be understood as (financial) loss raised from incorrectly predicted (classified) particular level of the factor which can be different for different levels. *OK* will start the computation.



(A) (B)

Fig. 6 Dialog window for SVM – Classification

Length	Width	Type
6.25	2.11	A
6.05	2.46	A
6.19	2.37	A
5.28	2.55	B
5.28	2.9	B
5.54	2.58	B
6.4	2.9	C
6.23	3.31	C
6.44	3.16	C
6.17	2.14	A
5.28	2.14	B
6.46	2.72	C

New samples to be classified

Length	Width
6.25	2.11
6.05	2.46
6.19	2.37
5.28	2.55
5.28	2.9

(A)

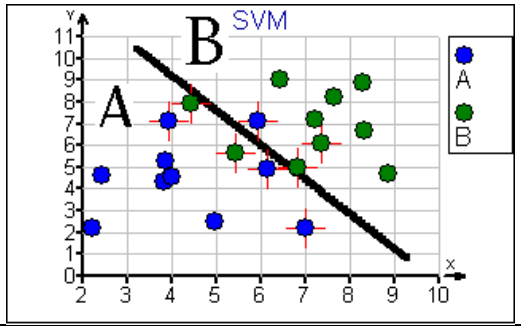
Na	K	Zn	Fe	Ni	Cr	Al	Origin
10.15	0.79	5.31	0.81	0.12	0.57	8.13	Afg
9.93	0.53	3.04	1.15	0.1	0.32	6.17	Colu
8.12	0.63	4.78	0.9	0.13	0.43	6.47	Colu
9.29	0.52	3.7	0.6	0.15	0.3	6.66	Mex
10.08	0.51	5.26	0.89	0.14	0.44	7.79	Tur
13.21	0.83	4.05	0.77	0.17	0.57	8.61	Tur
11.32	0.59	5.82	0.98	0.2	0.57	8.99	Mex
13.72	0.52	4.69	0.71	0.13	0.37	5.84	Afg
10.72	0.98	4.23	0.86	0.11	0.33	5.71	Afg
13.31	0.66	5.54	0.67	0.13	0.37	5.85	Morr
13.53	0.65	3.15	1.01	0.13	0.55	9.8	Morr

(B)

Fig. 7 Typical data for SVM-classification (shortened) – (A) Identification of biological species according to several morphological measures, (B) Determination of a drug based on the trace elementary chemical analysis

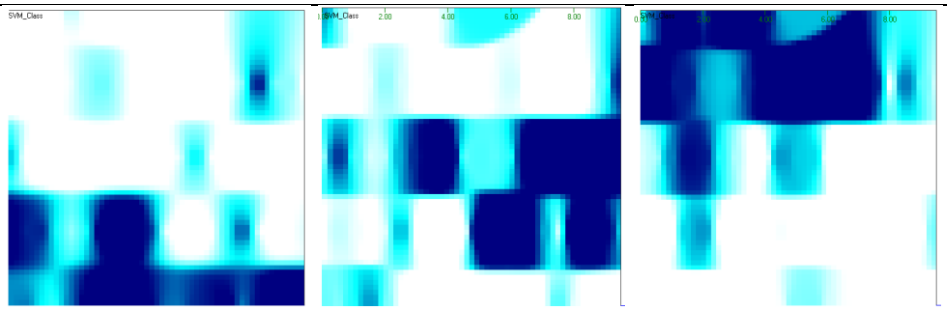
Protocol

Task name	Name from dialog window
Independent variable	Independent variable columns
Dependent variable	Factor response column
Prediction	Columns for prediction (if selected)
SVM Type	Used type of SVM classification, *Classification - Cost*, or by expected rate of misclassification *Classification - Nu*
Kernel	Selected kernel type: *Linear: (xi*xj)* *Polynomial: (-gamma*(xi*xj)+r)^d* *Radial Base Function: exp(-gamma*|xi-xj|^2)* *Sigmoidal: tanh(gamma*(xi*xj)+r)*
Degree	Used polynomial degree (for Polynomial kernel only)
Gamma	Kernel parameter γ
Cost	Loss coefficient *C* (only for *SVM-C*)
R	Kernel parameter *R*
Nu	SVM parameter nu ν (only for *SVM-Nu*)
Shrinking	Model shrinking option used (*Yes* or *No*)
Probabilities	Should probabilities for each level be computed? (*Yes / No*)
Weights	
Primal parameters	Primal parameters are valuable numbers computed only in case of linear kernel. They define the resulting model quantitatively and make the model generally applicable even without any SVM software, therefore are suitable e.g. for official publication. Primal parameters define the linear boundary $\mathbf{w}^T\mathbf{x} + \beta = w_1x_1 + w_2x_2 + \ldots w_mx_m + \beta = 0$ between all pairs of factor levels in the sample space defined by columns X. Absolute terms β are in column Beta, other columns contain coefficients w_i associated with predictors. Rows of the table are labeled with pairs of the discriminated factor levels, such as "A – B". When predictor values are substituted in the linear boundary equation, the sign of the result decides of the predicted value A or B: If the result is positive the predicted value is the first of the pair. If the result is negative the predicted value is the second of the pair. In the example shown below, prediction for the predictor value (x=4, y=5) will be: $\mathbf{w}^T\mathbf{x} = 5*(-0.7061) + 4*(-0.445) + 6.9123 = 1.602,$ therefore for the predictors (5,4) the model will predict „A".

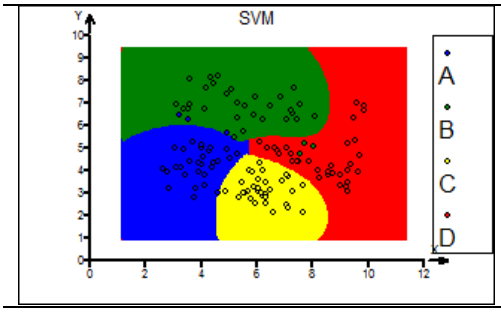
	
Misclassification table	Summary table of the correct and incorrect classifications for all response levels and a contingency table of misclassifications. In the bottom row are the actual occurrences of the levels, in the rightmost column are the predicted occurrences of levels. The bottom-right number of the sum of the right column, or the bottom row and equals the number of cases (rows) in the original data sheet. Other values in the table are numbers of correctly or incorrectly classified cases. For example, the value "9" in the table below says that the level "A" was nine times misclassified as "B" and the value "17" means that out of 30 occurrences (bottom line) of "B", only (or "as many as" – it depends…) 17 was classified correctly.
Classification affectivity	This table summarizes false positives rate (FP) and true positives rate (TP) and area under Receiver Operating Characteristic curve (ROC-AUC) calculate as [ROC AUC] = (TP*(1 – FP)+1/2*(TP*FP+(1 – TP)*(1 – FP)) for each factor level.
Support Vectors	If checked in the dialog window, lists all detected support vectors.
Prediction	Table of predicted values for the original data.
Index	Row index
Prediction	Predicted factor level from the model
Data	Actual factor level from data
Residuals	0 in case of correct classification, 1 in case of incorrect classification.
Level Probabilities	Probability of each possible factor level for the given predictor values. If, for example, we have two predictors and the factor has 3 levels (say, A, B, C), maps of probable outcome can be easily constructed from this table using the QC.Expert™ module Plotting/3DSpline as shown below.

Primal parameters	Beta	X	Y
A - B	6.9123	-0.7061	-0.445

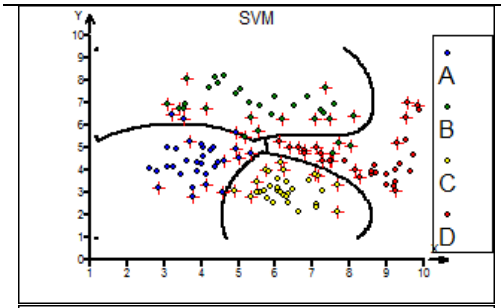
	A	B	C	Total
A	19	6	1	26
B	9	17	2	28
C	2	7	27	36
Total	30	30	30	**90**

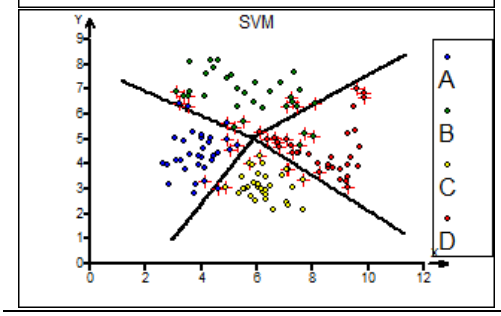
	 ———Probability of A ———— Probability of B ———— Probability of C———
Prediction	If the *Prediction* was checked in the dialog window, this table gives the predicted levels of the factor for new predictor values selected in the dialog window including probabilities for each level. Predicted level is the level with the highest (not necessarily "high" probability (for many levels the highest probability may still be well below 0.5).
Index	Row index
(independent variables)	Values of the new predictors
Prediction	Predicted (the most probable) factor level
Level Probabilities	Computed probabilities for all levels

Graphs

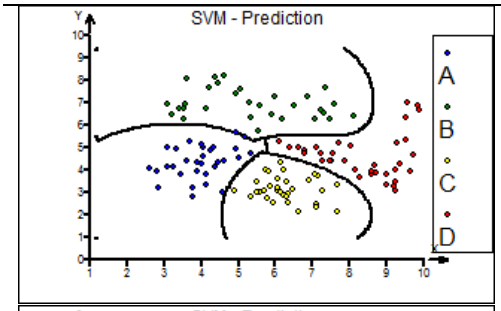


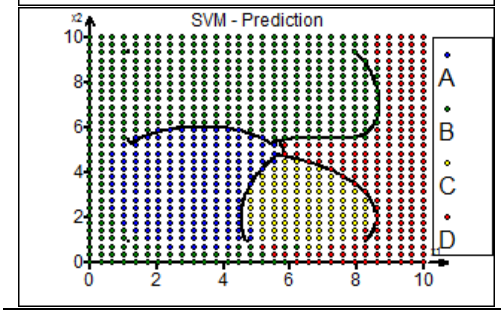
This plot is constructed only in case two independent variables (two independent columns). In the colored areas, the corresponding factor level is more probable than any other level. Every color represents one level, as given in the legend. If a factor has *m* levels the probability of a level to be predicted in a particular place must be greater than 1/*m*.



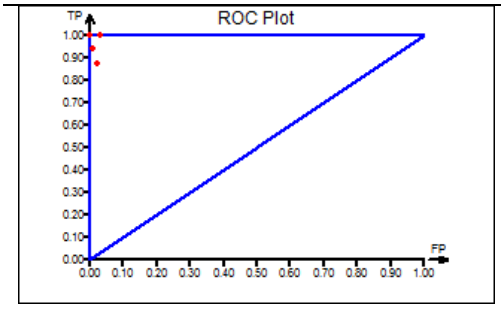


Plot of prediction boundary is the same plot as the previous one, but the boundary between different predictions is plotted instead of the areas. The boundary is a place in X where the probabilities of the adjacent levels are equal.

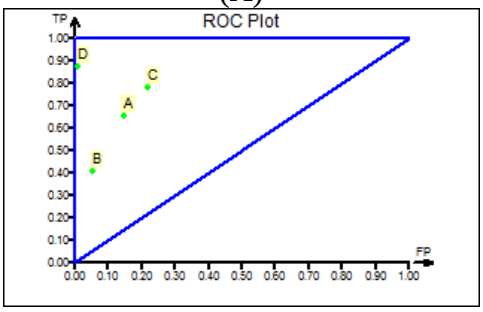




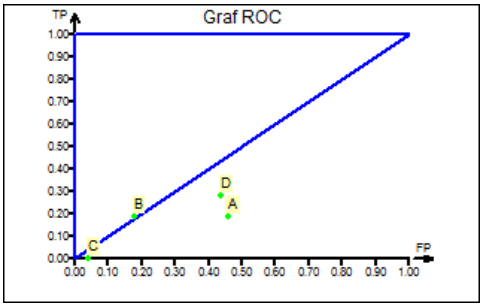
Projection of the new values of predictor into the prediction scheme from the previous plot. This plot is constructed if the checkbox *Prediction* was checked in the *SVM-Classification* dialog window.



(A)



(B)



(C)

Points of the ROC (Receiver Operating Characteristic) for an assessment of the affectivity of prediction of each factor level. True Positives rate (TP) are on the vertical axis, false positives rate, or FP are on the horizontal axis. In the left-upper corner, near (0,1) the prediction is very good (plot A). Moderately predicted levels are inside the triangle (B). Points below the diagonal (C) correspond to poor prediction where prediction is worse than throwing coin.

SVM – Regression

Menu:	QCExpert	SVM	SVM-Regression
DARWin	**SVMR**		

Data and parameters

Data are expected in two or more columns. One column contains the numerical dependent variable. The other one or more columns (the independent variables, or predictors) are also numerical. The independent and dependent variables are selected in the dialog window (Fig. 8). If the prediction is required, the independent variables for prediction are selected in the *Prediction* field. The number of predictors must be the same in *Independent variable* and *Prediction* fields. The button *Default X* will select the same predictors as selected in *Independent variables* field. In the *SVM Type* field select the type of the model criterion: ε in (0-5) or v in (0-4) (*Epsilon* or *Nu*). Checking the checkbox *Show Support* vectors will mark supporting points in 2d-plots (the 2d plots are constructed only in the case of 2 independent variables). Column selected in *Case labels* field will be used to label points in plots. Select the required SVM kernel transformation in the *Kernel* field.

*Linear: (xi*xj)*
Polynomial: (-gamma(xi*xj)+r)^d*
Radial Base Function: exp(-gamma|xi-xj|^2)*
Sigmoid: tanh(gamma(xi*xj)+r)*

and select further SVM parameters: *Degree* (degree of the polynomial in case of *Polynomial kernel* – recommended value 2 or 3), Gamma, *Cost* (in case of Epsilon method), *Nu* (in case of Nu method), *Epsilon* (in case of Epsilon method), *R* (in polynomial or sigmoid kernel). Choose if the shrinkage correction is to be used. *OK* will start the computation.

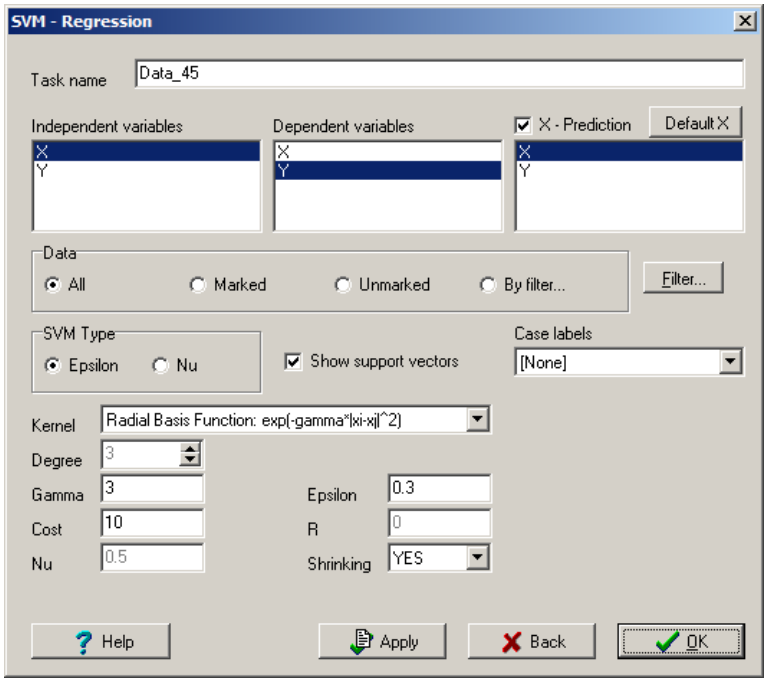


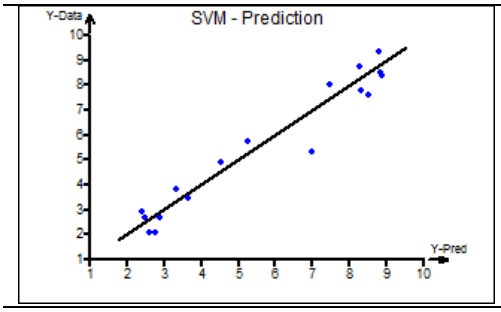
Fig. 8 Dialog window for SVM - Regression

Protocol

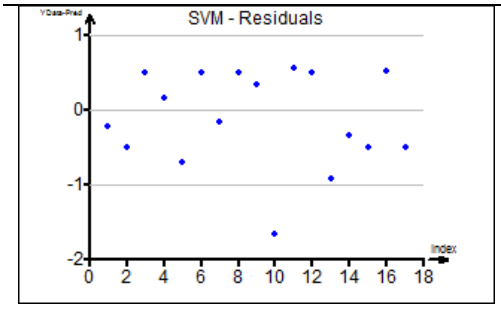
Task name	Name from dialog window
Independent variable	Independent variable columns
Dependent variable	Dependent variable (response) column
Prediction	Columns for prediction (if selected)
SVM Type	Used type of SVM regression, *Epsilon*, or expected rate of outliers *Nu*.
Kernel	Selected kernel type: *Linear: (xi*xj)* *Polynomial: (-gamma*(xi*xj)+r)^d* *Radial Base Function: exp(-gamma*|xi-xj|^2)*

	Sigmoidal: tanh(gamma(xi*xj)+r)*
Degree	Used polynomial degree (for Polynomial kernel only)
Gamma	Kernel parameter γ
Cost	Loss function coefficient *C*
Epsilon	Maximal error ε (only for *SVM* type = *Epsilon*)
R	Kernel parameter *R*
Nu	SVM parameter nu ν (only for *SVM-Nu*)
Shrinking	Model shrinking option used (*Yes* or *No*)
Support Vectors	If checked in the dialog window, lists all detected support vectors.
Prediction	Table of predicted values for the original data.
Index	Row index
Prediction	Predicted dependent variable value from the model
Data	Actual dependent variable value from data
Residuals	Regression residuals $y - f(x)$
Prediction	If the Prediction was checked in the dialog window, this table gives the predicted values of the response for new predictor values selected in the dialog window.
Index	Row index
(independent variables)	Values of the new predictors
Prediction	Predicted values of the response

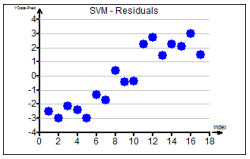
Graphs



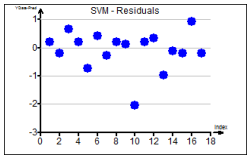
SVM Prediction plot plots measured response (Y-data) against predicted response (Y-Pred). In case of perfect fit (which should be avoided as perfect fit with very small or zero residuals is typically a bad model as it just copies the actual data, makes no generalization and provides usually very poor prediction) all points lie on the line.



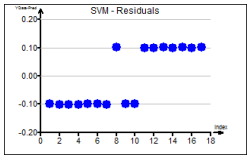
SVM Residuals plot visualizes vertical distances of the points from the line in the Prediction plot (above). This plot can help to diagnose the model and fit, as suggested at the following pictures.



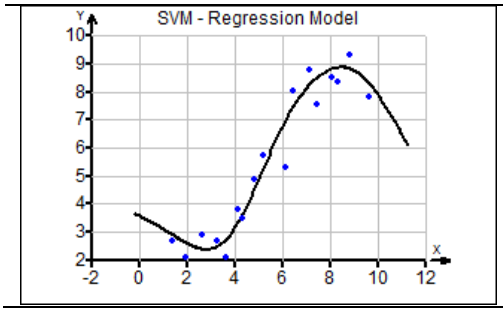
Probably underfitted: increase γ or decrease ε



maybe a reasonable fit



Probably overfitted: decrease γ or increase ε



If the independent variable is univariate (one column only) the data and regression function is plotted to assess the model visually.

SVM – Density

Menu:	QCExpert	SVM	SVM-Density

Data and parameters

Numerical data are expected in one or more columns. There is no dependent variable. Rows with missing data (empty cells) are discarded. Variables are selected in the dialog window (Fig. 9). If prediction is required the variables for prediction are selected in the *Prediction* field (*Prediction* checkbox must be checked). The number of predictors must be the same in *Independent variable* and *Prediction* fields. The button *Default X* will select the same predictors as selected in *Independent variables* field. Checking the checkbox *Show Support* vectors will mark suporting points in 2d-plots (the 2d plots are constructed only in the case of 2 independent variables). Column selected in *Case labels* field will be used to label points in plots. One column can be selected in *Point color in plot* field to distinguish color of the points in plot (different values or texts in this column will color the points with different colors). Select the required SVM kernel transformation in the *Kernel* field:

*Linear: (xi*xj)*
Polynomial: (-gamma(xi*xj)+r)^d*
Radial Base Function: exp(-gamma|xi-xj|^2)*
Sigmoid: tanh(gamma(xi*xj)+r).*

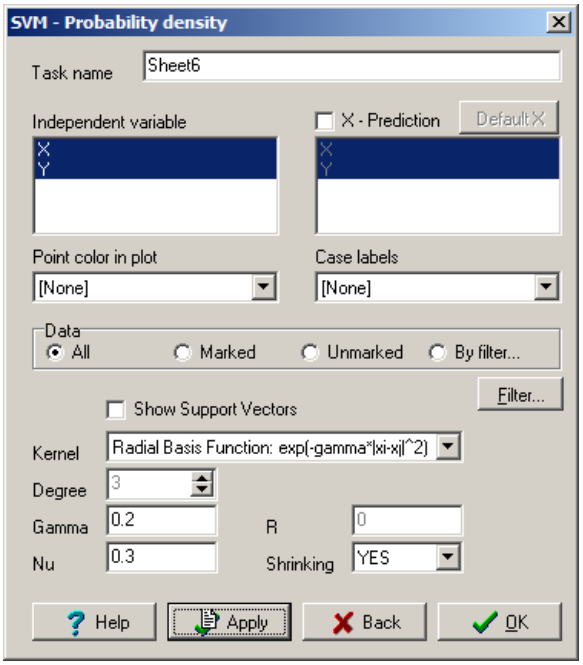


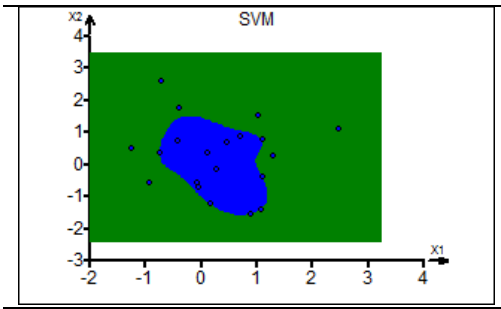
Fig. 9 Dialog window for SVM – Probability Density

If needed, change further parameters: *Degree* (degree of the polynomial kernel – for polynomial kernel only), value of Gamma kernel parameter γ, parameter ν (*Nu*), another kernel parameter *R* and select the *Shrinking* option. Then, press OK to start computation.

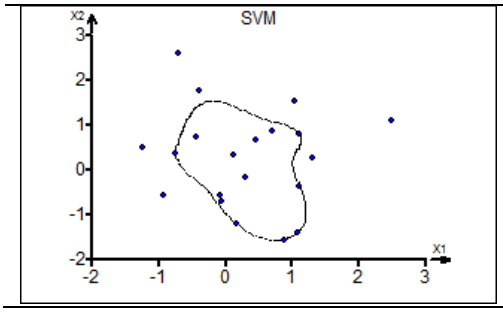
Protocol

Task name	Name from dialog window
Independent variable	Independent variable columns
Prediction	Columns for prediction (if selected)
SVM Type	"Probability density"
Kernel	Selected kernel type: *Linear: (xi*xj)* *Polynomial: (-gamma*(xi*xj)+r)^d* *Radial Base Function: exp(-gamma*|xi-xj|^2)* *Sigmoidal: tanh(gamma*(xi*xj)+r)*
Degree	Used polynomial degree (for Polynomial kernel only)
Gamma	Kernel parameter γ
Nu	The specifies parameter ν (only for type *Nu*)
R	Kernel parameter *R*
Shrinking	Model shrinking option used (*Yes* or *No*)
Support Vectors	If checked in the dialog window, lists all detected support vectors.
Prediction	Table of predicted values for the original data.
Index	Row index
Prediction	Membership of individual original data points (data rows) to the inner ν – quantile of the density model: 1 = inside the density, – 1 = outside the density.
Prediction	If the *Prediction* was checked in the dialog window, this table gives the predicted membership of individual "new" data points (data rows selected in the *Prediction* field of the dialog window) to the inner ν – quantile of the density model: 1 = inside the density, – 1 = outside the density.
Index	Row index
(variables)	Variables used for prediction
Prediction	Predicted membership

Graphs



For 2 columns of independent variables this plot visualizes an area of 100(1 – *v*)% volume of data with the highest density.



For 2 columns of independent variables this plot visualizes the border of an area of $100(1 - \nu)\%$ volume of data with the highest density.