

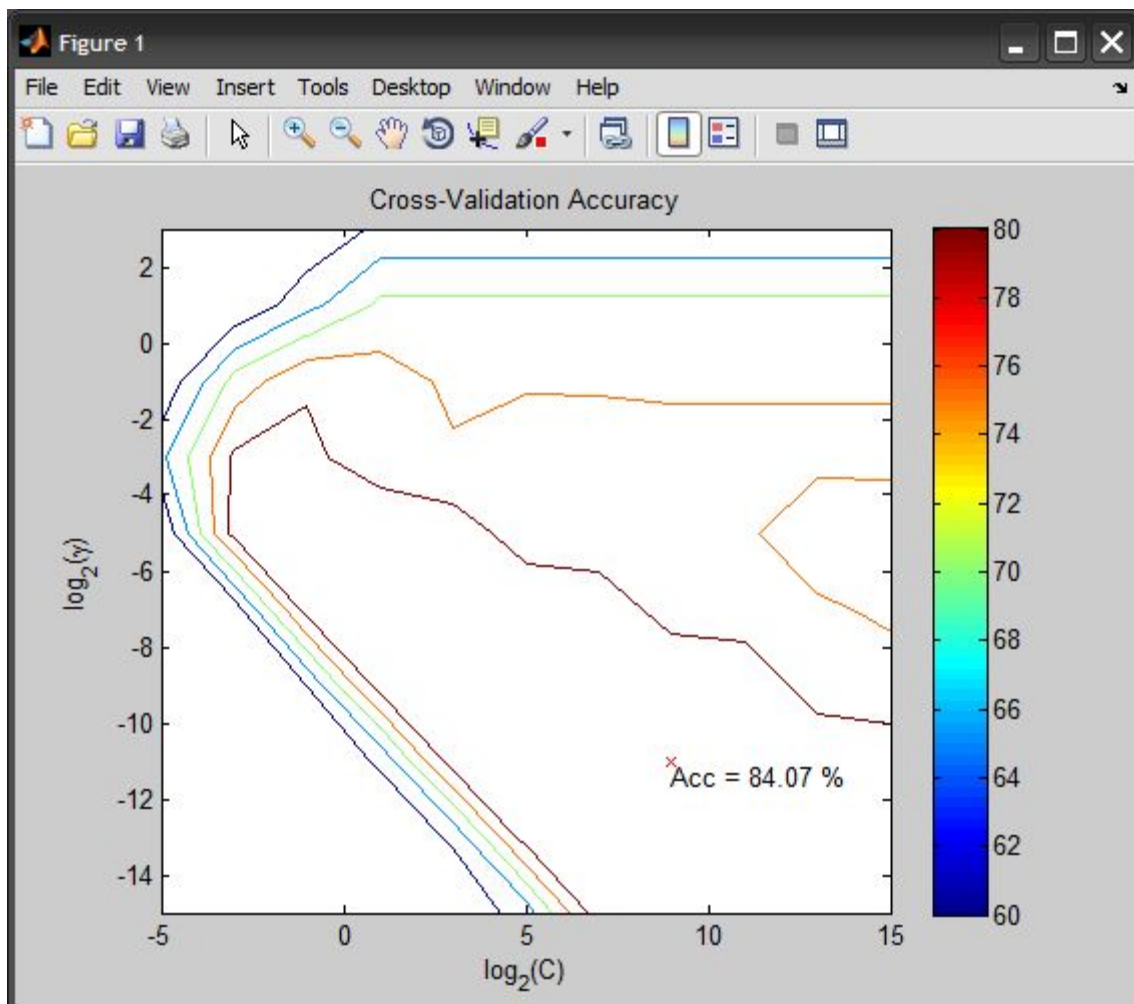
Check out [A practical guide to SVM Classification](#) for some pointers, particularly page 5.

We recommend a "grid-search" on C and γ using cross-validation. Various pairs of (C, γ) values are tried and the one with the best cross-validation accuracy is picked. We found that trying exponentially growing sequences of C and γ is a practical method to identify good parameters (for example,

$$C=2^{-5}, 2^{-3}, \dots, 2^{15};$$

$$\gamma=2^{-15}, 2^{-13}, \dots, 2^3).$$

This shouldn't be hard to implement: create a grid of values using MESHGRID, iterate overall all pairs (C, γ) training an SVM model with say 5-fold cross-validation, and choosing the values with the best CV-accuracy...



SVM Parameters

C

"However, it is critical here, as in any regularization scheme, that a proper value is chosen for C , the penalty factor. If it is too large, we have a high penalty for nonseparable points and we may store many support vectors and overfit. If it is too small, we may have underfitting." Alpaydin (2004)

"...the coefficient C affects the trade-off between complexity and proportion of nonseparable samples and must be selected by the user." Cherkassky and Mulier (1998)

"Selecting parameter C equal to the range of output values [6]. This is a reasonable proposal, but it does not take into account possible effect of outliers in the training data." "Our empirical results suggest that with optimal choice of ε , the value of regularization parameter C has negligible effect on the generalization performance (as long as C is larger than a certain threshold analytically determined from the training data)." Cherkassky and Ma (2002)

"In the support-vector networks algorithm one can control the trade-off between complexity of decision rule and frequency of error by changing the parameter C ,..." Cortes and Vapnik (1995)

"There are a number of learning parameters that can be utilized in constructing SV machines for regression. The two most relevant are the insensitivity zone ε and the penalty parameter C , which determines the trade-off between the training error and VC dimension of the model. Both parameters are chosen by the user." Kecman (2001)

The parameter C controls the trade off between errors of the SVM on training data and margin maximization ($C = \infty$ leads to hard margin SVM). Rychetsky (2001)

"The parameter C controls the trade-off between the margin and the size of the slack variables." Shawe-Taylor and Cristianini (2004)

"[Tuning the parameter C] In practice the parameter C is varied through a wide range of values and the optimal performance assessed using a separate validation set or a technique known as cross-validation for verifying performance using only a training set." Shawe-Taylor and Cristianini (2004) p220

"...the parameter C has no intuitive meaning." Shawe-Taylor and Cristianini (2004) p225

"The factor C in (3.15) is a parameter that allows one to trade off training error vs. model complexity. A small value for C will increase the number of training errors, while a large C will lead to a behavior similar to that of a hard-margin SVM." Joachims (2002)

"Let us suppose that the output values are in the range $[0, B]$. [...] a value of C about equal to B can be considered to be a robust choice." Mattera and Haykin (1999), pages 226-227 in *Advances in Kernel Methods*

Epsilon (ε)

"Similarly, Mattera and Haykin [6] propose to choose ε - value so that the percentage of SVs in the SVM regression model is around 50% of the number of samples. However, one can easily show examples when optimal generalization performance is achieved with the number of SVs larger or smaller than 50%." "Smola et al [8] and Kwok [9] proposed asymptotically optimal ε -

values proportional to noise variance, in agreement with general sources on SVM [2,7]. The main practical drawback of such proposals is that they do not reflect sample size. Intuitively, the value of ϵ should be smaller for larger sample size than for small sample size (with same noise level)."

"Optimal setting of ϵ requires the knowledge of noise level. The noise variance can be estimated directly from training data, i.e. by fitting very flexible (high-variance) estimator to the data. Alternatively, one can first apply least-modulus regression to the data, in order to estimate noise level." Cherkassky and Ma (2002)

"For an SVM the value of ϵ in the ϵ -insensitive loss function should also be selected. ϵ has an effect on the smoothness of the SVM's response and it affects the number of support vectors, so both the complexity and the generalization capability of the network depend on its value. There is also some connection between observation noise in the training data and the value of ϵ . Fixing the parameter ϵ can be useful if the desired accuracy of the approximation can be specified in advance." Horváth (2003), page 392 in Suykens *et al.*

"There are a number of learning parameters that can be utilized in constructing SV machines for regression. The two most relevant are the insensitivity zone ϵ and [...] Both parameters are chosen by the user. [...] An increase in ϵ means a reduction in requirements for the accuracy of approximation. It also decreases the number of SVs, leading to data compression." Kecman (2001)

"Under the assumption of asymptotically unbiased estimators we show that there exists a nontrivial choice of the insensitivity parameter in Vapnik's ϵ -insensitive loss function which scales linearly with the input noise of the training data. This finding is backed by experimental results." Smola, *et al.* (1998)

"The value of epsilon determines the level of accuracy of the approximated function. It relies entirely on the target values in the training set. If epsilon is larger than the range of the target values we cannot expect a good result. If epsilon is zero, we can expect overfitting. Epsilon must therefore be chosen to reflect the data in some way. Choosing epsilon to be a certain accuracy does of course only guarantee that accuracy on the training set; often to achieve a certain accuracy overall, we need to choose a slightly smaller epsilon."

"Parameter ϵ controls the width of the ϵ -insensitive zone, used to fit the training data. The value of ϵ can affect the number of support vectors used to construct the regression function. The bigger ϵ , the fewer support vectors are selected. On the other hand, bigger ϵ -values results in more flat estimates. Hence, both C and ϵ -values affect model complexity (but in a different way)." [Support Vector Machine Regression](#)

"A robust compromise can be to impose the condition that the percentage of Support Vectors be equal to 50%. A larger value of ϵ can be utilized (especially for very large and/or noisy training sets)..." Mattera and Haykin (1999)

"the optimal value of ϵ scales linearly with σ [variance of Gaussian noise]." Learning with Kernels

Kernel Parameters

"For classification problems, the optimal σ can be computed on the basis of Fisher discrimination. And for regression problems, based on scale space theory, we demonstrate the existence of a certain range of σ , within which the generalization performance is stable. An

appropriate σ within the range can be achieved via dynamic evaluation. In addition, the lower bound of iterating step size of σ is given." Wang, *et al.*, 2003.

Ali and Smith (2003) proposed an automatic parameter selection approach for the polynomial kernel.

General

[c] Ancona-etal02 showed that the Receiver Operating Characteristic (ROC) curves, measured on a suitable validation set, are effective for selecting, among the classifiers the machine implements, the one having performances similar to the reference classifier.

"The selection of appropriate values for the three parameters (C , e , s) in the above expressions has been proposed by various researchers. Cherkassky & Mulier (1998) suggested the use of cross-validation for the SVM parameter choice. Mattera & Haykin (1999) proposed the parameter C to be equal to the range of output values. They also proposed the selection of the e value to be such that the percentage of support vectors in the SVM regression model is around 50% of the number of samples. Smola et al. (1998) assigned optimal e values as proportional to the noise variance, in agreement with general sources on SVM. Cherkassky & Ma (2004) proposed the selection of e parameters based on the estimated noise. Different approaches yield different values for the three parameters. As shown later, this study finds the optimal parameter set simultaneously by minimising the prediction error as the objective function." Yu, Liong and Babovic (2004) found the optimal parameter set simultaneously by minimising the prediction error as the objective function.