

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), page 224

"...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), page 366

"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 e 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), page 182

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), page 82

"The parameter C controls the trade-off between the margin and the size of the slack variables."

Shawe-Taylor and Cristianini (2004) p220

"[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), page 40

"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 e and [...] Both parameters are chosen by the user. [...] An increase in e means a reduction in requirements for the accuracy of approximation. It also decreases the number of SVs, leading to data compression."

Kecman (2001), pages 182-183

"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 \square flat \square 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, page 79

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.

Bibliography

- ALI, S. and K.A. SMITH, 2003. [Automatic parameter selection for polynomial kernel](#), *Proceedings of the IEEE International Conference on Information Reuse and Integration (IRI 2003)*, pages 243-249. [Cited by 2] (0.62/year)
abstract = {Kernel is the heart of kernel based learning. To choose an appropriate parameter for a specific kernel is an important research issue in the data mining area. In this paper, we propose an automatic parameter selection approach for polynomial kernel. The algorithm is tested on support vector machines (SVM). The parameter selection is considered on the basis of prior information of the data distribution and Bayesian inference. The new approach is tested on different sizes of benchmark datasets with binary class problems as well as multi class classification problems.} AliSmith03 proposed an automatic parameter selection approach for the polynomial kernel.
- ALPAYDIN, Ethem, 2004. [Introduction to Machine Learning](#). books.google.com. [Cited by 28] (12.60/year)
- ANCONA, N., *et al.*, 2002. [Object detection in images: Run-time complexity and parameter selection of Support Vector Machines](#), *Proceedings of the 16th International Conference on Pattern Recognition (ICPR'02) - Volume 2*, pages 426-429. [not cited] (0/year)
abstract = {In this paper we address two aspects related to the exploitation of Support Vector Machines (SVM) for classification in real application domains, such as the detection of objects in images. The first one concerns the reduction of the run-time complexity of a reference classifier, without increasing its generalization error. In fact we show that the

complexity in test phase can be reduced by training SVM classifiers on a new set of features obtained by using Principal Component Analysis (PCA). Moreover, due to the small number of features involved, we explicitly map the new input space in the feature space induced by the adopted kernel function. Since the classifier is simply a hyperplane in the feature space, then the classification of a new pattern involves only the computation of a dot product between the normal to the hyperplane and the pattern. The second issue concerns the problem of parameter selection. In particular we show 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. We address these two issues for the particular application of detecting goals during a football match.} 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.

- BOARDMAN, Matthew and Thomas TRAPPENBERG, 2006. [A Heuristic for Free Parameter Optimization with Support Vector Machines](#), *Proceedings of the 2006 IEEE International Joint Conference on Neural Networks (IJCNN 2006)*, pp. 1337-1344. [not cited] (0/year)

abstract = {A heuristic is proposed to address free parameter selection for Support Vector Machines, with the goals of improving generalization performance and providing greater insensitivity to training set selection. The many local extrema in these optimization problems make gradient descent algorithms impractical. The main point of the proposed heuristic is the inclusion of a model complexity measure to improve generalization performance. We also use simulated annealing to improve parameter search efficiency compared to an exhaustive grid search, and include an intensity-weighted centre of mass of the most optimum points to reduce volatility. We examine two standard classification problems for comparison, and apply the heuristic to bioinformatics and retinal electrophysiology classification.} [C] proposed a heuristic (inclusion of a model complexity measure) to address the selection of C and the width γ of the Radial Basis Function (RBF) kernel. We focus on optimal selection of the free parameters in classification using Support Vector Machines (SVM). Specifically, we target optimization of the cost parameter C , which controls the tradeoff between maximization of the margin width and minimizing the number of misclassified samples in the training set [22], and the width γ of the Radial Basis Function (RBF) kernel. controls the width of the Gaussian kernel.

- CASSABAUM, Mary L., *et al.*, 2004. [Unsupervised optimization of support vector machine parameters](#), *Automatic Target Recognition XIV. Proceedings of the SPIE, Volume 5426* edited by Firooz A. Sadjadi, pages 316-325. [Cited by 2] (0.91/year)
- abstract = {Selection of the kernel parameters is critical to the performance of Support Vector Machines (SVMs), directly impacting the generalization and classification efficacy of the SVM. An automated procedure for parameter selection is clearly desirable given the intractable problem of exhaustive search methods. The authors' previous work in this area involved analyzing the SVM training data margin distributions for a Gaussian kernel in order to guide the kernel parameter selection process. The approach entailed several iterations of training the SVM in order to minimize the number of support vectors. Our continued investigation of unsupervised kernel parameter selection has led to a scheme employing selection of the parameters before training occurs. Statistical methods are applied to the Gram matrix to determine kernel optimization in an unsupervised fashion. This preprocessing framework removes the requirement for iterative SVM training. Empirical results will be presented for the "toy" checkerboard and quadboard problems.} investigate unsupervised

kernel parameter selection by applying statistical methods to the Gram matrix.

- CHALIMOURDA, Athanassia, Bernhard SCHOLKOPF and Alex J. SMOLA, 2000. [Choosing \$\nu\$ in support vector regression with different noise models—theory and experiments](#), IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks, Volume 5, edited by Shun-Ichi Amari *et al.*, Pages 199-204. [Cited by 1] (0.16/year)

abstract = {In support vector (SV) regression, a parameter ν controls the number of support vectors and the number of points that come to lie outside of the so-called ϵ -insensitive tube. For various noise models and SV parameter settings, we experimentally determine the values of ν that lead to the lowest generalization error. We find good agreement with the values that had previously been predicted by a theoretical argument based on the asymptotic efficiency of a simplified model of SV regression}

investigated the choice of ν in support vector regression with different noise models.

- CHALIMOURDA, Athanassia, Bernhard SCHÖLKOPF and Alex J. SMOLA, 2004. [Experimentally optimal \$\nu\$ in support vector regression for different noise models and parameter settings](#), *Neural Networks*, Volume 17, Issue 1, January 2004, Pages 127-141. [Cited by 9] (4.07/year)

abstract = {In Support Vector (SV) regression, a parameter ν controls the number of Support Vectors and the number of points that come to lie outside of the so-called ϵ -insensitive tube. For various noise models and SV parameter settings, we experimentally determine the values of ν that lead to the lowest generalization error. We find good agreement with the values that had previously been predicted by a theoretical argument based on the asymptotic efficiency of a simplified model of SV regression. As a side effect of the experiments, valuable information about the generalization behavior of the remaining SVM parameters and their dependencies is gained. The experimental findings are valid even for complex 'real-world' data sets. Based on our results on the role of the ν -SVM parameters, we discuss various model selection methods.} experimentally determined the optimal ν in support vector regression for different noise models and parameter settings.

- CHALIMOURDA, Athanassia, Bernhard SCHÖLKOPF and Alex J. SMOLA, 2005. [Letter to the Editor: Experimentally optimal \$\nu\$ in support vector regression for different noise models and parameter settings](#), *Neural Networks*, Volume 18, Issue 2 (March 2005), Page 205.

abstract = {n/a} do not use

- CHAPELLE, Olivier, *et al.*, 2002. [Choosing multiple parameters for support vector machines](#), *Machine Learning*, Volume 46, Numbers 1-3 (January 2002), Pages 131-159. [Cited by 257] (61.07/year)

abstract = {The problem of automatically tuning multiple parameters for pattern recognition Support Vector Machines (SVMs) is considered. This is done by minimizing some estimates of the generalization error of SVMs using a gradient descent algorithm over the set of parameters. Usual methods for choosing parameters, based on exhaustive search become intractable as soon as the number of parameters exceeds two. Some experimental results assess the feasibility of our approach for a large number of parameters (more than 100) and demonstrate an improvement of generalization performance.} used gradient descent to automatically tune C and the kernel parameters in SVM classification. considered the problem of automatically tuning multiple parameters for pattern recognition SVMs ADD MORE

Chapelle, Olivier, *et al.*,

Uses gradient descent.

KERNEL AND c

- CHERKASSKY, Vladimir and Yunqian MA, 2002. [Selection of meta-parameters for support vector regression](#), *Artificial Neural Networks -- ICANN 2002: International Conference, Madrid, Spain, August 2002, Proceedings*, edited by José R. Dorronsoro, pages 687-693. [Cited by 8] (1.90/year)

abstract = {We propose practical recommendations for selecting meta-parameters for SVM regression (that is, ϵ -insensitive zone and regularization parameter C). The proposed methodology advocates analytic parameter selection directly from the training data, rather than resampling approaches commonly used in SVM applications. Good generalization performance of the proposed parameter selection is demonstrated empirically using several low-dimensional and high-dimensional regression problems. In addition, we compare generalization performance of SVM regression (with proposed choice ϵ) with robust regression using 'least-modulus' loss function ($\epsilon = 0$). These comparisons indicate superior generalization performance of SVM regression.} proposed practical recommendations for selecting meta-parameters for SVM regression (ϵ and C) and advocated analytic parameter selection directly from the training data, rather than the usual resampling approaches.
- CHERKASSKY, Vladimir and Yunqian MA, 2003. [Comparison of Model Selection for Regression](#), *Neural Computation*, Volume 15, Issue 7 (July 2003), Pages 1691-1714. [Cited by 13] (4.05/year)

abstract = {We discuss empirical comparison of analytical methods for model selection. Currently, there is no consensus on the best method for finite-sample estimation problems, even for the simple case of linear estimators. This article presents empirical comparisons between classical statistical methods---Akaike information criterion (AIC) and Bayesian information criterion (BIC)---and the structural risk minimization (SRM) method, based on Vapnik-Chervonenkis (VC) theory, for regression problems. Our study is motivated by empirical comparisons in Hastie, Tibshirani, and Friedman (2001), which claims that the SRM method performs poorly for model selection and suggests that AIC yields superior predictive performance. Hence, we present empirical comparisons for various data sets and different types of estimators (linear, subset selection, and k -nearest neighbor regression). Our results demonstrate the practical advantages of VC-based model selection; it consistently outperforms AIC for all data sets. In our study, SRM and BIC methods show similar predictive performance. This discrepancy (between empirical results obtained using the same data) is caused by methodological drawbacks in Hastie et al. (2001), especially in their loose interpretation and application of SRM method. Hence, we discuss methodological issues important for meaningful comparisons and practical application of SRM method. We also point out the importance of accurate estimation of model complexity (VC-dimension) for empirical comparisons and propose a new practical estimate of model complexity for k -nearest neighbors regression.} [NOT PARAMETERS]
- CHERKASSKY, Vladimir and Yunqian MA, 2004. [Practical selection of SVM parameters and noise estimation for SVM regression](#), *Neural Networks*, Volume 17, Issue 1, January 2004, Pages 113-126. [Cited by 54] (24.45/year)

abstract = {We investigate practical selection of hyper-parameters for support vector machines (SVM) regression (that is, ϵ -insensitive zone and regularization parameter C). The proposed methodology advocates analytic parameter selection directly from the training data, rather than re-sampling approaches commonly used in SVM applications. In particular, we describe a new analytical prescription for setting the value of

insensitive zone ϵ ; as a function of training sample size. Good generalization performance of the proposed parameter selection is demonstrated empirically using several low- and high-dimensional regression problems. Further, we point out the importance of Vapnik's ϵ -insensitive loss for regression problems with finite samples. To this end, we compare generalization performance of SVM regression (using proposed selection of ϵ -values) with regression using 'least-modulus' loss ($\epsilon = 0$) and standard squared loss. These comparisons indicate superior generalization performance of SVM regression under sparse sample settings, for various types of additive noise.} investigate practical selection of hyper-parameters for SVM regression (ϵ and C).

- CHERKASSKY, Vladimir and Filip MULIER, 1998. *Learning from Data: Concepts, Theory, and Methods*, John Wiley & Sons, Inc. New York, N.Y., USA. [[Cited by 367](#)] (44.71/year)
- CORTES, C. and V. VAPNIK, 1995. [Support-vector networks](#). *Machine Learning*. [[Cited by 1802](#)] (160.55/year)
- DEBNATH, R. and H. TAKAHASHI, 2004. [An efficient method for tuning kernel parameter of the support vector machine](#), *Proceeding of the IEEE International Symposium on Communications and Information Technology, 2004 (ISCIT 2004)*, Volume 2, pp. 1023-1028. [not cited] (0/year)

abstract = {We propose a new method for searching the kernel parameter of the support vector machine on the basis of the distribution of data in the feature space. Although the distribution (structure) of data is unknown in the feature space, it depends on the kernel parameter. The distribution of data is characterized by the principal component analysis method. Thus, simple eigenanalysis method is applied to the matrix of the same dimension as the kernel matrix to find the kernel parameter. Therefore, this method is very fast. The proposed method can obtain the kernel parameter graphically.} proposed an efficient method for tuning the kernel parameter.

- DUAN, Kaibo, S. Sathiya KEERTHI and Aun Neow POO, 2003. [Evaluation of simple performance measures for tuning SVM hyperparameters](#), *Neurocomputing*, Volume 51, April 2003, Pages 41-59. [[Cited by 77](#)] (24.00/year)
- abstract = {Choosing optimal hyperparameter values for support vector machines is an important step in SVM design. This is usually done by minimizing either an estimate of generalization error or some other related performance measure. In this paper, we empirically study the usefulness of several simple performance measures that are inexpensive to compute (in the sense that they do not require expensive matrix operations involving the kernel matrix). The results point out which of these measures are adequate functionals for tuning SVM hyperparameters. For SVMs with L1 soft-margin formulation, none of the simple measures yields a performance uniformly as good as k -fold cross validation; Joachims' ξ -Alpha bound and the GACV of Wahba et al. come next and perform reasonably well. For SVMs with L2 soft-margin formulation, the radius margin bound gives a very good prediction of optimal hyperparameter values.} [class] evaluated simple performance measures for tuning SVM hyperparameters and concluded that "[f]or SVMs with L1 soft-margin formulation, none of the simple measures yields a performance uniformly as good as k -fold cross validation; Joachims' ξ -Alpha bound and the GACV of Wahba et al. come next and perform reasonably well. For SVMs with L2 soft-margin formulation, the radius margin bound gives a very good prediction of optimal hyperparameter values."

"For SVMs with L1 soft-margin formulation, none of the simple measures yields a performance uniformly as good as k -fold cross validation; Joachims' ξ -Alpha bound and the GACV of Wahba et al. come next and perform reasonably well. For SVMs with L2 soft-margin formulation, the radius margin bound gives a very good prediction of optimal hyperparameter values."

Xi-Alpha bound and the GACV of Wahba et al. come next and perform reasonably well. For SVMs with L2 soft-margin formulation, the radius margin bound gives a very good prediction of optimal hyperparameter values." Duan, Keerthi and Poo, 2002

- FROHLICH, H. and A. ZELL, 2005. [Efficient parameter selection for support vector machines in classification and regression via model-based global optimization](#), *Proceedings of the 2005 IEEE International Joint Conference on Neural Networks (IJCNN '05)*, Volume 3, pages 1431-1436. [not cited] (0/year)

abstract = {Support vector machines (SVMs) have become one of the most popular methods in machine learning during the last years. A special strength is the use of a kernel function to introduce nonlinearity and to deal with arbitrarily structured data. Usually the kernel function depends on certain parameters, which, together with other parameters of the SVM, have to be tuned to achieve good results. However, finding good parameters can become a real computational burden as the number of parameters and the size of the dataset increases. In this paper we propose an algorithm to deal with the model selection problem, which is based on the idea of learning an online Gaussian process model of the error surface in parameter space and sampling systematically at points for which the so called expected improvement is highest. Our experiments show that on this way we can find good parameters very efficiently.} proposed an algorithm to facilitate parameter selection based on the idea of learning an online Gaussian process model of the error surface in parameter space.
- GOLD, Carl and Peter SOLLICH, 2005. [Fast Bayesian Support Vector Machine Parameter Tuning with the Nystrom Method](#), *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN '05)*, Volume 5, pages 2820-2825. [Cited by 1] (0.82/year)

abstract = {We experiment with speeding up a Bayesian method for tuning the hyperparameters of a support vector machine (SVM) classifier. The Bayesian approach gives the gradients of the evidence as averages over the posterior, which can be approximated using hybrid Monte Carlo simulation (HMC). By using the Nystrom approximation to the SVM kernel, our method significantly reduces the dimensionality of the space to be simulated in the HMC. We show that this speeds up the running time of the HMC simulation from $O(n^2)$ (with a large prefactor) to effectively $O(n)$, where n is the number of training samples. We conclude that the Nystrom approximation has an almost insignificant effect on the performance of the algorithm when compared to the full Bayesian method, and gives excellent performance in comparison with other approaches to hyperparameter tuning.} experimented with speeding up a Bayesian method for tuning the hyperparameters in SVM classification.
- HORVATH, G., 2003. Neural Networks in Measurement Systems (an engineering view). *NATO SCIENCE SERIES SUB SERIES III COMPUTER AND SYSTEMS ...* [not cited] (0/year)
- IMBAULT, F. and K. LEBART, 2004. [A stochastic optimization approach for parameter tuning of support vector machines](#), *Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004)*, Volume 4, Pages 597-600. [Cited by 1] (0.45/year)

abstract = {Support vector machines (SVMs) are both mathematically well-founded and efficient in a large number of real-world applications. However, the classification results highly depend on the parameters of the model: the scale of the kernel and the regularization parameter. Estimating these parameters is referred to as tuning. Tuning requires to estimate the generalization error and to find its minimum over the parameter space. Classical methods

use a local minimization approach. After empirically showing that the tuning of parameters presents local minima, we investigate in this paper the use of global minimization techniques, namely genetic algorithms and simulated annealing. This latter approach is compared to the standard tuning frameworks and provides a more reliable tuning method.} investigated the application of genetic algorithms and simulated annealing to tuning SVM parameters.

- JENG, Jin-Tsong, 2006. [Hybrid approach of selecting hyperparameters of support vector machine for regression](#), *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, Volume 36, Issue 3, June 2006, Pages 699-709. [not cited] (0/year)
abstract = {To select the hyperparameters of the support vector machine for regression (SVR), a hybrid approach is proposed to determine the kernel parameter of the Gaussian kernel function and the epsilon value of Vapnik's ϵ -insensitive loss function. The proposed hybrid approach includes a competitive agglomeration (CA) clustering algorithm and a repeated SVR (RSVR) approach. Since the CA clustering algorithm is used to find the nearly "optimal" number of clusters and the centers of clusters in the clustering process, the CA clustering algorithm is applied to select the Gaussian kernel parameter. Additionally, an RSVR approach that relies on the standard deviation of a training error is proposed to obtain an epsilon in the loss function. Finally, two functions, one real data set (i.e., a time series of quarterly unemployment rate for West Germany) and an identification of nonlinear plant are used to verify the usefulness of the hybrid approach.} proposed a hybrid approach to selecting the parameter of the Gaussian kernel and epsilon in SVMs for regression.
- JENG, Jing-Tsong and Chen-Chia CHUANG, 2002. [A novel approach for the hyperparameters of support vector regression](#), *Proceedings of the 2002 International Joint Conference on Neural Networks (IJCNN '02)*, Volume 1, Pages 642-647. [Cited by 1] (0.24/year)
abstract = {In order to determine the hyperparameters of support vector regression (SVR), an approach with a two structured method is proposed to determine the kernel parameter s and ϵ in the ϵ -insensitive loss function. Firstly, the kernel parameter s of a Gaussian kernel function is determined by the competitive agglomeration (CA) clustering algorithm. The CA clustering algorithm incorporates the advantage of both hierarchical and partitioned clustering algorithms. Besides, it can find the nearly "optimum" number of clusters as well as its center of clusters in the clustering process. Secondly, the repeated SVR approach is proposed to obtain a proper ϵ in the ϵ -insensitive loss function that can be included in most of the data. Based on the efficiently structured way for choosing the hyperparameters s and ϵ , the simulation results have shown that the proposed approach comes close to the "optimum" hyperparameter region} describe a novel approach for determining the Gaussian kernel parameter s and ϵ SVM regression.
- JOACHIMS, T., 2002. [Learning to Classify Text Using Support Vector Machines: Methods, Theory and Algorithms](#). books.google.com. [Cited by 263] (62.26/year)
- JORDAAN, E.M. and G.F. SMITS, 2002. [Estimation of the regularization parameter for support vector regression](#), *Proceedings of the 2002 International Joint Conference on Neural Networks (IJCNN '02)*, Volume 3, Pages 2192-2197. [Cited by 5] (1.19/year)
abstract = {Support vector machines use a regularization parameter C to regulate the trade-off between the complexity of the model and the empirical risk of the model. Most of the techniques available for determining the optimal value of C are very time consuming. For industrial applications of the SVM method, there is a need for a fast and robust method to estimate C . A method based on the characteristics of the kernel, the range of output values and the size of the ϵ -insensitive zone, is proposed} proposed a method of determining the optimal value of the regularization parameter, C , which is based on the

characteristics of the kernel, the range of output values and the size of the ϵ -insensitive zone. Jordaan, E.M. Smits, G.F. Dept. of Math. & Comput. Sci., Eindhoven University of Technology.;

- KECCMAN, V., 2001. Learning and soft computing. MIT Press Cambridge, Mass. [[Cited by 53](#)] (10.15/year)
- KLINKENBERG, Ralf, 2002. "Informed Parameter Setting for Support Vector Machines: Using Additional User Knowledge in Classification Tasks" (Report Number CI-126/02, Collaborative Research Center on Computational Intelligence (SFB CI) (SFB 531), University of Dortmund, Dortmund, Germany, 2002; ISSN 1433-3325). Informed Parameter Setting for Support Vector Machines: Using Additional User Knowledge in Classification Tasks Many applications of machine learning involve the learning of classifiers. Given a set of labeled training examples, the task is to learn a classifier for predicting the labels of previously unseen examples. By providing the labels of the training examples, the user already specifies a lot of her or his knowledge about the classification problem at hand. In some cases, however, the user may not be satisfied with the result provided by the learning method. Hence the user may want to specify additional knowledge about the problem or constraints on the desired solution and she or he may want the learner to provide a classifier that fits better to her or his needs. The goal of this research is to allow the user to specify additional knowledge about the classification problem and to incorporate this knowledge into the learning process. In this work, support vector machines (SVMs) were chosen as learning methods for classifiers. Two methods for integrating user knowledge into the learning process of an SVM for classification tasks are discussed. Example weighting allows the user to set individual weights for individual training examples, which are then used in SVM training. Kernel modification allows the incorporation of a user's knowledge about similarities and dissimilarities of examples into the SVM training by modified kernel functions. \citeasnoun{Klinkenberg02} discussed two methods for integrating user knowledge into the learning process of an SVM for classification. Example weighting allows the user to set individual weights for individual training examples, whilst kernel modification allows the incorporation of a user's knowledge about similarities and dissimilarities of examples.
- KUBA, Petr, *et al.*, 2002. Exploiting sampling and meta-learning for parameter setting support vector machines, *Proceedings of the Workshop de Minería de Datos Y Aprendizaje of (IBERAMIA 2002)*, edited by F.J. Garijo and J.C. Riquelme and M. Toro, pages 217-225. [[Cited by 1](#)] (0.24/year)
abstract = {It is a known fact that good parameter settings affect the performance of many machine learning algorithms. Support Vector Machines (SVM) and Neural Networks are particularly affected. In this paper, we concentrate on SVM and discuss some ways to set its parameters. The first approach uses small samples, while the second one exploits meta-learning and past results. Both methods have been thoroughly evaluated. We show that both approaches enable us to obtain quite good results with significant savings in experimentation time.} [regression] exploited sampling and meta-learning for parameter setting.
- KULKARNI, Abhijit, V.K. JAYARAMAN and B.D. KULKARNI, 2004. [Support vector classification with parameter tuning assisted by agent-based technique](#), *Computers and Chemical Engineering* [[Cited by 7](#)] (3.15/year)
abstract = {This paper describes a robust support vector machines (SVMs) classification methodology, which can offer superior classification performance for important process engineering problems. The method incorporates efficient tuning procedures based on minimization of radius/margin and span bound for leave-one-out errors. An agent-based asynchronous teams (A-teams) software framework, which combines Genetic-Quasi-

Newton algorithms for the optimization is highly successful in obtaining the optimal SVM hyper-parameters. The algorithm has been applied for classification of binary as well as multi-class real world problems.} described an agent-based technique for SVMs for classification parameter tuning.

- KWOK, James T. and Ivor W. TSANG, 2003. [Linear Dependency between \$\epsilon\$ and the Input Noise in \$\epsilon\$ -Support Vector Regression](#), *IEEE Transactions on Neural Networks* [Cited by 10] (1.92/year)

abstract = {In using the ϵ -support vector regression (ϵ -SVR) algorithm, one has to decide a suitable value for the insensitivity parameter ϵ . Smola et al.} considered its "optimal" choice by studying the statistical efficiency in a location parameter estimation problem. While they successfully predicted a linear scaling between the optimal ϵ and the noise in the data, their theoretically optimal value does not have a close match with its experimentally observed counterpart in the case of Gaussian noise. In this paper, we attempt to better explain their experimental results by studying the regression problem itself. Our resultant predicted choice of ϵ is much closer to the experimentally observed optimal value, while again demonstrating a linear trend with the input noise.} demonstrated a linear trend with ϵ and the input noise.
- LIM, Hojung, 2004. Support vector parameter selection using experimental design based generating set search (SVEG) with application to predictive software data modeling, Doctoral Thesis, Syracuse University. [not cited] (0/year)

abstract = {Predictive data modeling is germane to many engineering and scientific applications. Recently, a new type of learning machine, called support vector machine (svm), has gained prominence for predictive modeling of classification and regression problems. However, the solution of svm requires some user specified parameters called hyperparameters. In practice these are determined by a computationally intensive grid search. In this research, we develop a principled approach for the selection of svm hyperparameters. The proposed three step methodology consists of determination of parametric ranges based on their interrelationships, setting up experimental designs for an efficient exploration of the error surface, and pursuing generating set search for local refinement. We demonstrate its efficacy for software module classification and effort prediction problems.} developed a principled approach for the selection of SVM hyperparameters.
- LIN, Pao-Tsun, Shun-Feng SU and Tsu-Tian LEE, 2005. [Support vector regression performance analysis and systematic parameter selection](#), *Proceedings of the International Joint Conference on Neural Networks (IJCNN '05), Volume 2*, pages 877-882. [not cited] (0/year)

abstract = {Support vector regression (SVR) based on statistical learning is a useful tool for nonlinear regression problems. The SVR method deals with data in a high dimension space by using linear quadratic programming techniques. As a consequence, the regression result has optimal properties. However, if parameters were not properly selected, overfitting and/or underfitting phenomena might occur in SVR. Two parameters σ , the width of Gaussian kernels and ϵ , the tolerance zone in the cost function are considered in this research. We adopted the concept of the sampling theory into Gaussian filter to deal with parameter σ . The idea is to analyze the frequency spectrum of training data and to select a cut-off frequency by including 90% of power in spectrum. The corresponding σ can then be obtained through the sampling theory. In our simulations, it can be found that good performances are observed when the selected frequency is near the cut-off frequency. For another parameter ϵ , it is a tradeoff between the number of support vectors and the RMSE. By introducing the confidence interval concept, a suitable selection

of ϵ can be obtained. The idea is to use the L_1 -norm (i.e., when $\epsilon = 0$) to estimate the noise distribution of training data. When ϵ is obtained by selecting the 90% confidence interval, simulations demonstrated superior performance in our illustrative example. By our systematical design, proper values of σ and ϵ can be obtained and the resultant system performances are nice in all aspects.} consider the selection of σ , the width of Gaussian kernels and ϵ in SVM regression.

- MATTERA, Davide and Simon HAYKIN, 1999. [Support vector machines for dynamic reconstruction of a chaotic system](#). In: *Advances in Kernel Methods: Support Vector Learning*, edited by Bernhard Schölkopf and Christopher J. C. Burges and Alexander J. Smola, Pages 209-241. [Cited by 26] (3.61/year)

abstract = {Dynamic reconstruction is an inverse problem that deals with reconstructing the dynamics of an unknown system, given a noisy time-series representing the evolution of one variable of the system with time. The reconstruction proceeds by utilizing the time-series to build a predictive model of the system and, then, using iterated prediction to test what the model has learned from the training data on the dynamics of the system. In this paper, we review the details of the theoretical derivation of the Support Vector Machine (SVM); this allows us to derive its close relationship with the regularized radial basis function. The dependence of the SVM performance on the choice of its parameters is investigated both by means of theoretical analysis and numerical experiments performed on the well-known Lorenz system. The results obtained show the effectiveness of the SVM in performing the nonlinear reconstruction; its main advantage consists in the possibility of trading off the required accuracy with the number of Support Vectors.} [reg] considered the choice of parameters and kernels
- QUAN, Yong and Jie YANG, 2003. [An improved parameter tuning method for support vector machines](#), *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, 9th International Conference, RSFDGrC 2003, Chongqing, China, May 26-29, 2003, Proceedings*, Pages 607-610. [not cited] (0/year)

abstract = {Support vector machines (SVMs) is a very important tool for data mining. However, the problem of tuning parameters manually limits its application in practical environment. In this paper, under analyzing the limitation of these existing approaches, a new methodology to tuning kernel parameters, based on the computation of the gradient of penalty function with respect to the RBF kernel parameters, is proposed. Simulation results reveal the feasibility of this new approach and demonstrate an improvement of generalization ability.} proposed a new methodology to tuning kernel parameters, based on the computation of the gradient of penalty function with respect to the RBF kernel parameters.
- RYCHETSKY, Matthias, 2001. *Algorithms and Architectures for Machine Learning based on Regularized Neural Networks and Support Vector Approaches*, Shaker Verlag. [not cited] (0/year)
- SCHITTKOWSKI, K., 2005. [Optimal parameter selection in support vector machines](#), *Journal of Industrial and Management Optimization*, Volume 1, Number 4, November 2005, pp. 465-476. [not cited] (0/year)

abstract = {The purpose of the paper is to apply a nonlinear programming algorithm for computing kernel and related parameters of a support vector machine (SVM) by a two-level approach. Available training data are split into two groups, one set for formulating a quadratic SVM with L_2 -soft margin and another one for minimizing the generalization error, where the optimal SVM variables are inserted. Subsequently, the total generalization error is evaluated for a separate set of test data. Derivatives of functions by which the optimization problem is defined, are evaluated in an analytical way, where an existing

Cholesky decomposition needed for solving the quadratic SVM, is exploited. The approach is implemented and tested on a couple of standard data sets with up to 4,800 patterns. The results show a significant reduction of the generalization error, an increase of the margin, and a reduction of the number of support vectors in all cases where the data sets are sufficiently large. By a second set of test runs, kernel parameters are assigned to individual features. Redundant attributes are identified and suitable relative weighting factors are computed.} applied a nonlinear programming algorithm for computing kernel and related parameters of an SVM.

- SCHÖLKOPF, Bernhard and Alexander J. SMOLA, 2002. [Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond](#). [Cited by 328] (62.78/year)
- SHAWE-TAYLOR, John and Nello CRISTIANINI, 2004. [Kernel Methods for Pattern Analysis](#). books.google.com. [Cited by 215] (96.66/year)
- SHITONG, Wang, *et al.*, 2005. [Theoretically Optimal Parameter Choices for Support Vector Regression Machines with Noisy Input](#), *Soft Computing - A Fusion of Foundations, Methodologies and Applications* [not cited] (0/year)
 abstract = {With the evidence framework, the regularized linear regression model can be explained as the corresponding MAP problem in this paper, and the general dependency relationships that the optimal parameters in this model with noisy input should follow is then derived. The support vector regression machines Huber-SVR and Norm-r r-SVR are two typical examples of this model and their optimal parameter choices are paid particular attention. It turns out that with the existence of the typical Gaussian noisy input, the parameter ϵ in Huber-SVR has the linear dependency with the input noise, and the parameter r in the r-SVR has the inversely proportional to the input noise. The theoretical results here will be helpful for us to apply kernel-based regression techniques effectively in practical applications.} consider theoretically optimal parameter choices for SVMs for regression and conclude that with the existence of the typical Gaussian noisy input, the parameter ϵ in Huber-SVR has the linear dependency with the input noise, and the parameter r in Norm- r SVM regression is inversely proportional to the input noise.
- SMOLA, A., *et al.*, 1998. [Asymptotically optimal choice of \$\epsilon\$ -loss for support vector machines](#), *Proceedings of the International Conference on Artificial Neural Networks*, pp. 105-110, Springer, Berlin. [Cited by 34] (4.14/year)
 abstract = {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.} show that under the assumption of asymptotically unbiased estimators, 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.
- SOARES, Carlos, Pavel B. BRAZDIL and Petr KUBA, 2004. [A meta-learning method to select the kernel width in support vector regression](#), *Machine Learning*, Volume 54, Number 3 (March 2004), Pages 195-209. [Cited by 9] (4.04/year)
 abstract = {The Support Vector Machine algorithm is sensitive to the choice of parameter settings. If these are not set correctly, the algorithm may have a substandard performance. Suggesting a good setting is thus an important problem. We propose a meta-learning methodology for this purpose and exploit information about the past performance of different settings. The methodology is applied to set the width of the Gaussian kernel. We carry out an extensive empirical evaluation, including comparisons with other methods (fixed default ranking; selection based on cross-validation and a heuristic method commonly used

to set the width of the SVM kernel). We show that our methodology can select settings with low error while providing significant savings in time. Further work should be carried out to see how the methodology could be adapted to different parameter setting tasks.} propose a meta-learning method to select the width of the Gaussian kernel in support vector regression

- STAELIN, Carl, 2003. [Parameter selection for support vector machines](#), *HP Laboratories Israel, Tech. Rep. HPL-2002-354 (R. 1), Nov.* [Cited by 10] (3.12/year)
 abstract = {We present an algorithm for selecting support vector machine (SVM) meta-parameter values which is based on ideas from design of experiments (DOE) and demonstrate that it is robust and works effectively and efficiently on a variety of problems.}
 presented an algorithm for selecting SVM meta-parameter values which is based on ideas from design of experiments (DOE). [both]
- STEINWART, Ingo, 2003. [On the optimal parameter choice for \$\nu\$ -support vector machines](#), *IEEE Transactions on Pattern Analysis and Machine Intelligence*, October 2003 (Vol. 25, No. 10), pp. 1274-1284. [Cited by 8] (2.49/year)
 abstract = {We determine the asymptotically optimal choice of the parameter ν for classifiers of ν -support vector machine (ν -SVM) type which has been introduced by Scholkopf et al. [14]. It turns out that ν should be a close upper estimate of twice the optimal Bayes risk provided that the classifier uses a so-called universal kernel such as the Gaussian RBF kernel. Moreover, several experiments show that this result can be used to implement some modified cross validation procedures which improve standard cross validation for ν -SVMs.} conclusion = "In this paper we have shown that an asymptotical optimal choice of the regularization parameter ν for ν -SVM's is an arbitrary close upper bound of twice the Bayes risk of the considered optimization problem. [...]" \cite{Steinwart03} considered ν -SVMs for classification and showed that an asymptotical optimal choice of the regularization parameter ν is an arbitrary close upper bound of twice the optimal Bayes risk provided that the classifier uses a so-called universal kernel such as the Gaussian RBF kernel. ["new support vector algorithms" *Neural computation* 12 (2000) 1207-1245]
- WANG, Wenjian, *et al.*, 2003. [Determination of the spread parameter in the Gaussian kernel for classification and regression](#), *Neurocomputing*, Volume 55, Number 3, October 2003, pp. 643-663. [Cited by 21] (6.52/year)
 abstract = {Based on statistical learning theory, Support Vector Machine (SVM) is a novel type of learning machine, and it contains polynomial, neural network and radial basis function (RBF) as special cases. In the RBF case, the Gaussian kernel is commonly used, while the spread parameter σ in the Gaussian kernel is essential to generalization performance of SVMs. In this paper, determination of σ is studied based on discussions of the influence of σ on generalization performance. 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. Simulation results show the effectiveness of the presented method.} consider the spread parameter in the Gaussian kernel for classification and regression. They found that for classification problems, the optimal σ can be computed on the basis of Fisher discrimination and for regression problems, based on scale space theory, they demonstrated the existence of a certain range of σ , within which the generalization performance is stable.
- WANG, Xin, *et al.*, 2005. [Parameter selection of support vector regression based on hybrid optimization algorithm and its application](#), *Journal of Control Theory and*

Applications, pages 371-376. [not cited] (0/year)

abstract = {Choosing optimal parameters for support vector regression (SVR) is an important step in SVR design, which strongly affects the performance of SVR. In this paper, based on the analysis of influence of SVR parameters on generalization error, a new approach with two steps is proposed for selecting SVR parameters. First the kernel function and SVM parameters are optimized roughly through genetic algorithm, then the kernel parameter is finely adjusted by local linear search. This approach has been successfully applied to the prediction model of the sulfur content in hot metal. The experiment results show that the proposed approach can yield better generalization performance of SVR than other methods.} used a two-step process for parameter selection for SVMs for regression. They first optimized the kernel function and SVM parameters roughly using genetic algorithms, then they adjusted the kernel parameter finely using local linear search.

- WANG, S., *et al.*, 2006. [Experimental study on parameter choices in norm- \$r\$ support vector regression machines with noisy input](#), *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, Volume 10, Number 3 / February, 2006, pages 219-223. [not cited] (0/year)

abstract = {In [1], with the evidence framework, the almost inversely linear dependency between the optimal parameter r in norm- r support vector regression machine r -SVR and the Gaussian input noise is theoretically derived. When r takes a non-integer value, r -SVR cannot be easily realized using the classical QP optimization method. This correspondence attempts to achieve two goals: (1) The Newton-decent-method based implementation procedure of r -SVR is presented here; (2) With this procedure, the experimental studies on the dependency between the optimal parameter r in r -SVR and the Gaussian noisy input are given. Our experimental results here confirm the theoretical claim in [1].} performed an experimental study on parameter choices in SVM regression.

- YU, Xinying, Shie-Yui LIONG and Vladan BABOVIC, 2004. [EC-SVM approach for real-time hydrologic forecasting](#), *Journal of Hydroinformatics*, Volume 6, Number 3, July 2004, Pages 209-223. [Cited by 1] (0.45/year)

abstract = {This study demonstrates a combined application of chaos theory and support vector machine (SVM) in the analysis of chaotic time series with a very large sample data record. A large data record is often required and causes computational difficulty. The decomposition method is used in this study to circumvent this difficulty. The various parameters inherent in chaos technique and SVM are optimised, with the assistance of an evolutionary algorithm, to yield the minimal prediction error. The performance of the proposed scheme, EC-SVM, is demonstrated on two daily runoff time series: Tryggev{\ae} lde catchment, Denmark and the Mississippi River at Vicksburg. The prediction accuracy of the proposed scheme is compared with that of the conventional approach and the recently introduced inverse approach. This comparison shows that EC-SVM yields a significantly lower normalised RMSE value of 0.347 for the Tryggev{\ae} lde catchment runoff and 0.0385 for the Mississippi River flow compared to 0.444 and 0.2064, respectively, resulting from the conventional approach. A slight improvement in accuracy was obtained by analysing the first difference or the daily flow difference time series. It should be noted, however, that the computational speed in analysing the daily flow difference time series is significantly much faster than that of the daily flow time series.} [reg]

"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 ϵ 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 ϵ values as proportional to the noise variance, in agreement with general sources on SVM. Cherkassky & Ma (2004) proposed the selection of ϵ 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, Liang and Babovic (2004)

found the optimal parameter set simultaneously by minimising the prediction error as the objective function.