Letter to the editor

Experimentally optimal ν in support vector regression for different noise models and parameter settings

In our Neural Networks paper Chalimourda, Schölkopf, and Smola (2004) we presented results on the optimal choice of the parameter ν in regression support vector machines (SVMs). In Section 5 we extended our experiments discussing the optimal choice of the other SVM parameters, C and the σ_{kernel} of the Gaussian kernel. For the parameter C we found that it shows significant effect on the risk only when changed by orders of magnitude. We suggested that $C = y_{\text{max}} \cdot l$ is a good starting value, where l is the number and y_{max} the maximum output value of the training data.

With respect to our suggestion for C, we argued in Section 7.1 that the prescriptions $C = \max(|\bar{y} + 3\sigma_y|, |\bar{y} - 3\sigma_y|)$, where $\bar{y}$ and σ_y are the mean and the standard deviation of the output values of the training data and $C = y_{\text{max}}$ given by Cherkassky and Ma (2002, 2004) and Mattera and Hakin (1999), respectively, are too low. Unfortunately, we overlooked the fact that the other authors use a different scaling in the objective function optimized by the SVM. We optimized the functional

$$\tau(\mathbf{w}, \xi^{(*)}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\frac{1}{l}\sum_{i=1}^{l}(\xi_i + \xi_i^*), \tag{1}$$

see Eq. (4), Section 2 in Chalimourda et al. (2004). In the above equation the sum of the slack variables that stands for the empirical risk is weighted by C/l. In Cherkassky and Ma (2002, 2004) and Mattera and Hakin (1999) the empirical risk is weighted only by C, see for example Eq. (10) in Cherkassky and Ma (2004). The different scaling is the reason why the factor l appears in our suggested C value while it does not in Cherkassky and Ma (2002, 2004) and Mattera and Hakin (1999). In fact, all prescriptions of C work well depending on the scaling of the empirical risk one uses.

We would like to apologize for overlooking the different scaling used putting the results of Cherkassky and Ma (2002, 2004) and Mattera and Hakin (1999) in an unfavorable light.

References

Chalimourda, A., Schölkopf, B., & Smola, A. J. (2004). Experimentally optimal ν in support vector regression for different noise models and parameter settings. *Neural Networks*, *17*, 127–141.

Cherkassky, V., & Ma, Y. (2002). *Selection of meta-parameters for support vector regression Proceedings of the International Conference on Artificial Neural Networks (ICANN), Madrid, Spain.*

Cherkassky, V., & Ma, Y. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks*, *17*, 113–126.

Mattera, D., & Haykin, S. (1999). Support vector machines for dynamic reconstruction of a chaotic system. In B. Schölkopf, C. Burges, & A. Smola (Eds.), *Advances in kernel methods-support vector learning* (pp. 211–214). Cambridge, MA: MIT Press, 211–214.

Athanassia Chalimourda*
Ruhr-Universität Bochum, Institut für Neuroinformatik,
Bochum D-44780, Germany
E-mail address: athanassia.chalimourda@
neuroinformatik.ruhr-uni-bochum.de

Bernhard Schölkopf[1]
Max-Planck-Institut for Biological Cybernetics
Spemannstraße 38, Tübingen D-72076, Germany
E-mail address: bernhard.schoelkopf@tuebingen.mpg.de

Alex J. Smola[2]
Australian National University,
Canberra ACT 0200, Australia
E-mail address: alex.smola@anu.edu.au

*Corresponding author. Tel.: +49 6104 941784.
[1] Tel.: +49 7071 601 551; fax: +49 7071 601 552.
[2] Tel.: +61 2 6125 8652; fax: +61 2 6125 8651.