Support Vector Machine Model Selection Using Strangeness

David R. Hardoon, Zakria Hussain & John Shawe-Taylor
University College London, Dept. of Computer Science, U.K.
{D.Hardoon, Z.Hussain, jst}@cs.ucl.ac.uk

Abstract

Model Selection is the task of choosing the best model for a particular data analysis task. It generally makes a compromise between fit with the data and the complexity of the model. Currently the most popular techniques used by practitioners are Cross-Validation (CV) and Leave-One-Out (LOO). In this study we concentrated on the Support Vector Machine (SVM) (Boser et al., 1992) model.

Recently, Özöğür-Akyüz et al. (In Press), following on work by Özöğür et al. (2008), show that selecting a model whose hyperplane achieves the maximum separation from a test point obtains comparable error rates to those found by selecting the SVM model through CV. In other words, while methods such as CV involve finding one SVM model (together with its optimal parameters) that minimises the CV error, Özöğür-Akyüz et al. (In Press) keep all of the models generated during the model selection stage and make predictions according to the model whose hyperplane achieves the maximum separation from a test point. The main advantage of this approach is the computational saving when compared to CV or LOO. However, their method is only applicable to large margin classifiers like SVMs.

We continue this line of research, but rather than using the distance of each test point from the hyperplane we explore the idea of using the *nonconformity measure* (Vovk et al., 2005; Shafer & Vovk, 2008) of a test sample to a particular label set. The nonconformity measure is a function that evaluates how 'strange' a prediction is according to the different possibilities available. The notion of nonconformity has been proposed in the on-line learning framework of conformal prediction (Shafer & Vovk, 2008), and is a way of scoring how different a new sample is from a bag[1] of old samples. The premise is that if the observed samples are well-sampled then we should have high confidence on correct prediction of new samples, given that they *conform* to the observations.

We take the nonconformity measure and apply it to the SVM algorithm during testing in order to gain a time advantage over CV and to generalise the algorithm of Özöğür-Akyüz et al. (In Press). Hence we are not restricted

[1]A *bag* is a more general formalism of a mathematical *set* that allows repeated elements.

to SVMs (or indeed a measure of the margin for prediction) and can apply our method to a broader class of learning algorithms. However, due to space constraints we only address the SVM technique and leave the application to other algorithms (and other nonconformity measures not using the margin) as a future research study. Furthermore we also derive a novel learning theory bound that uses nonconformity as a measure of complexity. To our knowledge this is the first attempt at using this type of measure to upper bound the loss of learning algorithms.

Acknowledgements

The authors would like to acknowledge financial support from the EPSRC project Le Strum[2], EP-D063612-1 and from the EU project PinView[3], FP7-216529.

References

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory* (pp. 144–152). Pittsburgh ACM.

Özöğür, S., Shawe-Taylor, J., Weber, G. W., & Ögel, Z. B. (2008). Pattern analysis for the prediction of fungal pro-peptide cleavage sites. *Discrete Applied Mathematics, Special Issue on Networks in Computational Biology, doi:10.1016/j.dam.2008.06.043.*

Özöğür-Akyüz, S., Hussain, Z., & Shawe-Taylor, J. (In Press). Prediction with the SVM using test point margins. *Annals of Information Systems, Special Issue on Optimization methods in Machine Learning.*

Shafer, G., & Vovk, V. (2008). A tutorial on conformal prediction. *Journal of Machine Learning Research*, *9*, 371–421.

Vovk, V., Gammerman, A., & Shafer, G. (2005). *Algorithmic learning in a random world.* New York: Springer.

[2]http://www.lestrum.org
[3]http://www.pineview.eu