Estimating Correlations of Random Classifiers on Large Data Sets

Abstract for the MLDM Workshop of the AI*IA Conference

Věra Kůrková[1] and Marcello Sanguineti[2 *]

1 Institute of Computer Science of the Czech Academy of Sciences
Pod Vodárenskou věží 2 - 182 07 Prague 8, Czech Republic
2 Department of Computer Science, Bioengineering,
Robotics, and Systems Engineering (DIBRIS)
University of Genova
Via Opera Pia 13 - 16145 Genova, Italy
* corresponding author
vera@cs.cas.cz, marcello.sanguineti@unige.it

Abstract

A probabilistic model of relevance of classification tasks is developed to investigate classification of large data sets by feedforward neural networks. Optimization of networks is studied in terms of correlations of randomly-chosen classifiers with network input-output functions. Effects of increasing sizes of sets of data to be classified are analyzed by exploiting geometrical properties of high-dimensional spaces. Consequences on concentrations of values of sufficiently smooth functions around their mean values are considered. It is shown that the critical factor for suitability of a class of networks for computation of randomly-chosen classifiers is the size of the mean value of their correlations with network input-output functions. This research has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 824160.

References

[1] D. Dubhashi and A. Panconesi. Concentration of Measure for the Analysis of Randomized Algorithms. Cambridge University Press, 2009.

[2] V. Kůrková and M. Sanguineti. Model complexities of shallow networks representing highlyvarying functions. Neurocomputing 171:598-604, 2016.

[3] V. Kůrková and M. Sanguineti. Probabilistic lower bounds for approximation by shallow perceptron networks. Neural Networks 91:34-41,2017.

[4] V. Kůrková and M. Sanguineti. Classification by sparse neural networks. IEEE Trans. on Neural Networks and Learning Systems 30(9):2746-2754, 2019.