

Randomized Machine Learning

Yuri S. Popkov

In recent years there has been a trend toward formation of scientific research activities aimed at developing of software for solving clusters of problems which varies by content but similar by their underlying mathematical models. Clusters like “Data Mining”, “Data Science”, “Big Data” and others in general are about technology and differ only in terms.

Machine Learning [1-3] as a scientific area stands separately from areas noticed above. Its main trait is that there are defined problem classes, data structures and mathematical methods of their solving. In addition it has almost 80-year history, where there were increase and decrease periods which mostly correlated with progress of computational hardware. I will not an exaggeration if we claim that general case of machine learning problem is based on the triple: data, models, algorithms which aim at classification, clustering, regression and recognition problems. Main feature of modern ML-procedures corresponds to models with non-random parameters and targets to obtaining of their estimates by minimization of empirical risk [4].

We propose a new class of ML-procedures – randomized ML-procedures (RML-procedures), basic idea of which is based on models with random parameters, obtaining estimates of parameters’ densities which are entropy-optimal for existed data and generation of ensembles of models with these probability characteristics [5, 6]. RML-algorithms reduce to functional entropy-linear programming problem for which general solution has been obtained. There have been developed methods of generation of ensembles with entropy-optimal PDFs. Proposed examples of randomized text classification and randomized regression.

References

1. Rosenblatt M. The Perceptron – Perceiving and Recognizing Automaton. Report 85-460-1, 1957.
2. Tsypkin Ya. Z. Foundations of the Theory of Learning Systems. – Academic Press, 1973.
3. Vapnik V.N. Statistical Learning Theory, NY, John Willey, 1998.
4. Bishop C.,M. Pattern Recognition and Machine Learning. Springer, 2006.
5. Popkov Y.S. , Popkov A.Y., Lysak Y.N. Estimation of Characteristics of Randomized Static Models of Data (Entropy-robust Approach) // Automation and Remote Control, 2013, Vol. 74, No. 11, pp. 1863–1877.
6. Popkov Y. S., Popkov A.Y. New Methods of Entropy-Robust Estimation for Randomized Models under Limited Data // Entropy 2014, 16, 675-698.