UDC 004.8

M.I. SCHLESINGER
International Research and Training Center for Information Technologies and Systems of the NAS of Ukraine and MES of Ukraine, Kyiv, Ukraine, e-mail: *schles@irtc.org.ua.*

E.V. VODOLAZSKIY
International Research and Training Center for Information Technologies and Systems of the NAS of Ukraine and MES of Ukraine, Kyiv, Ukraine, e-mail: *waterlaz@gmail.com.*

MINIMAX DEVIATION STRATEGIES FOR MACHINE LEARNING AND RECOGNITION WITH SHORT LEARNING SAMPLES

Abstract. The article analyses risk-oriented formulation of pattern recognition and machine learning problems. Based on arguments from multicriteria optimization, a class of improper strategies is defined that are dominated by some other strategy. A general form of strategies that are not improper is derived. It is shown that some widely used approaches are improper in the defined sense, including the maximum likelihood estimation approach. This drawback is especially apparent when dealing with short learning samples of fixed length. A unified formulation of pattern recognition and machine learning problems is presented that embraces the whole range of sizes of the learning sample, including zero size. It is proven that solutions to problems in the presented formulation are not improper. The concept of minimax deviation recognition and learning is formulated, several examples of its implementation are presented and compared with the widely used methods based on the maximal likelihood estimation.

Keywords: pattern recognition, machine learning, short learning sample.

INTRODUCTION

The short learning sample problem has been around in machine learning under different names during its whole life. The learning sample is used to compensate for the lack of knowledge about the recognized object when its statistical model is not completely known. Naturally, the longer the learning sample, the better the subsequent recognition. However, when the learning sample becomes too small (2, 3, 5 elements) the effect of small samples becomes evident. In spite of the fact that any learning sample (even a very small one) provides some additional information about the object, it may be better to ignore the learning sample than to utilize it with the commonly used methods.

Example 1. Let us consider an object that can exist in one of two random states $y=1$ and $y=2$ with equal probabilities. In each state the object generates two independent Gaussian random signals x_1 and x_2 with variances equal 1. Mean values of signals depend on the state as it is shown in Fig. 1. In the first state, the mean value is $(2, 0)$. In the second state, the mean value depends on an unknown parameter θ and is $(0, \theta)$. Even when no learning sample is given a minimax strategy can be used to make a decision about the state y. The minimax strategy ignores the second signal and makes decision $y^* = 1$ when $x_1 > 1$ and decision $y^* = 2$ when $x_1 \le 1$.

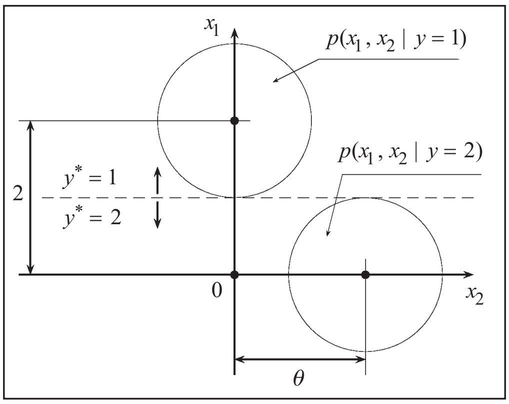

Fig. 1. Example 1: $(x_1, x_2) \in \mathbb{R}^2$ — signal, $y \in \{1, 2\}$ — state

Now let us assume that there is a sample of signals generated by