Preface

This book deals with mathematical constructions that are foundational in such an important area of *data mining* as *pattern recognition*. A closer look is taken at *infeasible systems of linear inequalities*, whose generalized solutions act as building blocks of geometric decision rules for recognition.

Infeasible systems of linear inequalities proved to be a key object in pattern recognition problems described in geometrical terms thanks to *the committee method*.

Infeasible systems of inequalities represent an important special subclass of *infeasible systems of constraints with monotonicity property* – systems whose multi-indices of feasible subsystems form *abstract simplicial complexes* (*independence systems*), fundamental objects of combinatorial topology. In discrete mathematics, the faces of such complexes are interpreted as zeros of *monotone Boolean functions*. Chapter 1 of the book deals with simplicial complexes and monotone Boolean functions related to common infeasible systems of constraints. The graph-theoretic methods represent a very productive way to study combinatorial and structural properties of infeasible systems of constraints. From the applied point of view, the most important property is the *connectedness* of a specific graph assigned to a family of *maximal feasible subsystems*. For instance, the set of solutions taken one by one for each of the maximal feasible subsystems of an infeasible system, which constitute an *odd cycle* in such a graph, represents a committee for an infeasible system of linear inequalities over $\mathbb{R}^n$ formally describing a pattern recognition problem. Thus, graph-theoretic methods that help us to solve one of the main tasks of committee theory – searching for a committee with the minimal number of elements can be taken as a basis for efficient algorithms of constructing decision rules for pattern recognition. The connectedness of graphs discussed is actually determined by the connectedness of the space $\mathbb{R}^n$; moreover, the connectedness of similar graphs in the context of common topological spaces is also determined by the connectedness of these spaces. The subject matter of Chapter 2 is (hyper)graphs corresponding to facets of common simplicial complexes and to maximal feasible subsystems of infeasible systems of linear inequalities.

Equally interesting results are obtained from an analysis of infeasible systems of linear inequalities by methods of combinatorial geometry. In Chapter 3, the notion of *diagonal* of a polytope, which is traditional for plane geometry, is generalized to multidimensional convex polytopes. A dual correspondence between diagonals and facets of polytopes, on the one hand, and multi-indices of maximal feasible and minimal infeasible subsystems of inequalities, on the other hand, is described. This duality is used, in particular, to obtain different estimates of the number of subsystems.

In Chapter 4, the correspondence between infeasible systems of inequalities and monotone Boolean functions motivates us to construct *algorithms for optimal inference* of functions. Several criteria for optimality of algorithms of inference are considered, and algorithms satisfying these criteria are constructed.

DOI 10.1515/9783110481068-001, BY-NC-ND © 2020 Damir Gainanov, published by De Gruyter.
This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.

In Chapter 5, the algorithmic approach to constructing an optimal committee of an infeasible system of linear inequalities is considered; it is based on such principal features of graphs as the connectedness and the existence of odd cycles. A brief review of *alternative covers* in the second half of this chapter provides a new look at collective solutions to infeasible systems of constraints.

The aim of this book is to present a mathematical toolset finding an application to the construction of pattern recognition complexes that solve the recognition problem in its geometric setting.

Such complexes of pattern recognition start their work with preprocessing of a training sample, that is, a massive collection of vectors from a high-dimensional feature space. Because the vectors of the training sample are preliminarily divided into groups that partially represent logically uniform classes or categories, they reflect a certain knowledge domain in the boundaries of which every new unclassified vector entering into the complex must be referred to one of the classes. At consecutive stages of preprocessing, the groups from the training sample are aggregated, with the use of hierarchical tree-like structures, into two extended groups that partially represent the corresponding generalized classes. The task of the recognition complex consists in the search for a geometric object that has a relatively simple formal description and, at the same time, strictly separates the vectors from distinct extended groups of the training sample. In the context of the book, the above-mentioned task can be interpreted, for example, as the search for a separating hyperplane in an Euclidean feature space. In practice, information contained in almost any training sample leads to a situation where a unique separating hyperplane cannot be found, because the linear inequality system underlying the problem of the discrimination of the two extended groups turns out to be infeasible. By means of some dimensional increase of the input data, the inequalities become homogeneous; their strictness is motivated by the stability demands that must be satisfied by the decision rules generated by the pattern recognition complex. This is how the infeasible system of homogeneous strict linear inequalities comes to the stage in the contradictory two-class pattern recognition problem, which has to be solved by the complex. The system as a whole has no solution, but any of its feasible subsystems can be solved by the software of the recognition complex that implements modern powerful techniques of linear optimization. The smart committee strategy of the recognition complex consists in the finding of solutions to a few maximal feasible subsystems and in their combining into a committee decision rule which operates with arrangements of separating hyperplanes. On the one hand, such a rule always allows the complex to correctly discriminate the vectors from the two extended groups of the training sample and, on the other hand, it makes it possible to apply the procedure of committee voting to a new vector entering into the complex; the majority decision rule, governed by the committee, refers the new vector to a generalized class. The recognition complex implements various effective techniques for constructing the separating committees, by exploiting specific properties of the (hyper)graphs of the maximal feasible subsystems of infeasible

systems of linear inequalities. With the help of these techniques, the complex repeatedly solves the two-class pattern recognition problem for each higher level extended group of vectors from the training sample, adding at every step some committee decision rule to a resulting hierarchical tree-like structure. This structure represents the machine for recognition of new vectors, and it correctly recognizes any vector of the training sample.

This edition is the extended translation of the book *Combinatorial Geometry and Graphs in an Analysis of Infeasible Systems and Pattern Recognition* published by *Nauka*, Moscow, in 2014.

Moscow and *Ekaterinburg* *Damir N. Gainanov*
October 2016