CrossMark

Committee polyhedral separability: complexity and polynomial approximation

Michael Khachay[1,2,3]

Abstract We consider the minimum affine separating committee (MASC) combinatorial optimization problem, which is related to ensemble machine learning techniques on the class of linear weak classifiers combined by the rule of simple majority. Actually, the MASC problem is a mathematical formalization of the famous Vapnik–Chervonenkis principle of structural risk minimization in the mentioned class of classifiers. According to this principle, it is required to construct a best performance ensemble classifier belonging to a family of the least possible VC-dimension. It is known that the MASC problem is *NP*-hard and remains intractable in spaces of any fixed dimension $n > 1$ even under an additional constraint on the separated sets to be in general position. This special case of the MASC problem called MASC-GP(n) is the main subject of interest of the present paper. To design polynomial-time approximation algorithms for a class of combinatorial optimization problems containing the MASC problem, we propose a new framework, adjusting the well-known Multiple Weights Update method. Following this approach, we construct polynomial-time approximation algorithms with state-of-the-art approximation guarantee for the MASC-GP(n) problem. The results obtained provide a theoretical framework for learning a high-performance ensembles of affine classifiers.

Keywords Polyhedral separability · Affine committees · Computational complexity · Approximability

✉ Michael Khachay
   mkhachay@imm.uran.ru

1 Krasovsky Institute of Mathematics and Mechanics UB RAS, Yekaterinburg, Russia

2 Ural Federal University, Yekaterinburg, Russia

3 Omsk State Technical University, Omsk, Russia

🖉 Springer

1 Introduction

A system of constraints (algebraic or differential equations, inequalities, inclusions, etc.) is a mathematical model, which is widely involved in economic modeling, decision making, machine learning, and data mining. In many cases, the formulation of an initial problem contains contradictory conditions; this leads the corresponding system of constraints to be infeasible; i.e., to have no solutions in the usual sense. To this end, before solving the problem, it is required to generalize the concept of a solution itself. In machine learning, there are two known significant approaches to such a generalization. Namely, they are: (i) the kernel trick (Cortes and Vapnik 1995; Schölkopf and Smola 2002; Hofmann et al. 2008) resolving the infeasibility of an initial instance of the SVM-classifier learning problem by an implicit mapping (of this instance) into a new feature space and (ii) ensemble learning techniques, among them are boosting (Schapire 1990; Freund 1995; Schapire and Freund 2012), bagging (Breiman 1996, 2001), etc.

During past decades, along with the famous (Eremin et al. 1983; Cortes and Vapnik 1995; Eremin 2002) Chebyshev ε-relaxation continuous approach to the optimal correction of infeasible systems, an alternative discrete approach to such a generalization has been developed (Mazurov 1990; Khachai et al. 2002; Khachai 2006; Mazurov and Khachai 2004, 2007; Kobylkin 2012). According to this approach, the concept of a solution for a system of constraints (in case of its infeasibility) is refined using the concept of a finite ensemble of appropriately chosen quasi-solutions. The theory of committee generalized solutions based on fundamental results of Vl. Mazurov (Mazurov 1990) provides a mathematical framework for this generalization, where partial quasi-solutions, also known as committee members, are combined by the majority voting.

We consider the Minimum Affine Separating Committee (MASC) combinatorial optimization problem closely related to the special case of a two-class classification problem, for which piecewise linear classifiers also take their decisions by the rule of simple majority voting.

According to the classic setting of the classification problem, for a finite sample

$$(x_1, y_1), \ldots, (x_m, y_m), \tag{1}$$

where x_i are points in an n-dimensional Euclidean feature space E, and $y_i \in \{-1, 1\}$, it is required to construct (learn) a *committee* piecewise linear classifier

$$h(x) = \text{sign} \sum_{j=1}^{k} \alpha_j \text{sign}(c_j^T x - d_j) \tag{2}$$

for some integers $\alpha_j \geq 0$. Since any partial classifier $\text{sign}(c_j^T x - d_j)$ is a composition of an *affine* and threshold functions, the classifier h is also called *affine separating committee*. If a classifier h classifies correctly the instances of sample (1), we call such a classifier *perfect* with respect to this sample.

In the context of Vapnik–Chervonenkis structural risk minimization framework, it is significant to design learning algorithms, which, given sample (1), construct perfect committee classifiers of form (2) belonging to a family of the minimum VC-dimension.

The motivation of our paper is based on the following two facts:

(i) for any non-contradictory sample (1) (i.e., a sample, for which inequality $y_{i_1} \neq y_{i_2}$ implies $x_{i_1} \neq x_{i_2}$), there exists a committee classifier h, which is perfect w.r.t. this sample (Mazurov 1971);

(ii) the family of classifiers (2) defined on the n-dimensional space E and sharing the property $\sum_{j=1}^{k} \alpha_j = q$ has VC-dimension $O(qn)$ (Khachay and Poberii 2009).

In the MASC problem, the goal is to search for perfect committee (2) w.r.t. sample (1) having the minimum value[1] of the total weight $\sum_{j=1}^{k} \alpha_j$. Therefore, the MASC problem can be considered as a mathematical formalization of the structural risk minimization principle in the class of committee classifiers. Any optimal (or even a good approximate) solution of this problem produces a classifier with a high generalization performance.

Unfortunately, the MASC problem is strongly NP-hard and remains intractable even in a fixed-dimensional space (for any dimension $n > 1$) and under the additional constraint of general position for a training set (Definition 2, Sect. 3). This special case denoted by MASC-GP(n) is the main subject of interest of this paper. Since this problem is intractable, the question of its effective approximability remains still open. To date, the best known (Khachay and Poberii 2009) ratio (state-of-the-art) for polynomial-time approximation algorithms for the MASC-GP(n) problem is $O(m/n)$ (in the general case) and $O(\ln m)$ for some special instances (in Sect. 4, we call them *nice*).

Using a suitable generalization of the famous boosting by sampling (Freund 1995) learning strategy, we construct new polynomial-time approximation algorithms for this problem with improved accuracy. In particular, we present an algorithm with a general approximation ratio of $O((\frac{m \ln m}{n})^{1/2})$ and a ratio of $O(\ln m)$ for the same nice instances, where m is the length of the training sample. Hereinafter, w.l.o.g., we suppose that $m > n$, since otherwise the initial sample can be polynomially reduced to an equivalent one located in the $(m-1)$-dimensional space.

The paper has the following structure. In Sect. 2, we give an overview of related work. In Sect. 3, we introduce the basic notation and the Minimum Affine Separating Committee (MASC) combinatorial optimization problem. Section 4 contains a brief survey of known results our considerations are based on. These results stem from the theory of linear inequalities and matrix games. In Sect. 5, we present our own results concerning the adaptation of the Multiple Weights Update algorithm to finding approximate solutions of the MASC problem. Our approximation algorithms (one of which has the best known approximation ratio) are proposed in Sect. 6. Finally, in Sect. 7, we summarize the results presented and discuss questions that still remain open.

2 Related work

The notion of a *committee solution* for the pattern recognition problem (*committee classifier, committee machine, committee*) was introduced in the middle of the 1960s (Ablow 1965; Nilsson 1965). These kinds of learnable machines are closely related to the geometric separation problem for finite sets by ensembles of hyperplanes. The works above mostly deal with illustrative examples of the committee concept and present several learning heuristics in the class of committee piecewise linear classifiers. Arguably, the first theoretical result in this field is Mazurov's theorem (Mazurov 1971) providing a criterion of the perfect learnability in the class of affine committees in terms of properties of the given training sample. We discuss this result in Sect. 3, where the criterion is referred as Theorem 1.

To the best of our knowledge, most of the related results come from several areas of computer science, among them are neural networks, computational geometry, combinatorial optimization, and ensemble learning theory. Since an affine separating committee can

---

[1] This value is also called *committee length*.

be represented in terms of a simple feed-forward neural network (e.g., see Khachay 2007), complexity results for the MASC problem can be considered as a progress in the complexity analysis of more general learning problems for the class of k-layered perceptrons. In particular, the algorithms proposed in our paper improve the complexity results obtained in (Lin and Vitter 1991; Blum and Rivest 1992) for the perceptron loading problem.

In computational geometry, the k-Polyhedral Separability Problem is studied from the late 1980s (Megiddo 1988). In this problem, for a given boolean formula $\varphi(\xi_1, \ldots, \xi_k)$ and finite sets A and B located in an finite-dimensional space, it is required to find a collection of k hyperplanes separating A and B according to the formula φ. It can be easily verified (e.g., see Khachay and Poberii 2009) that an affine separating committee is a solution of this problem for the special type of φ. Therefore, any affine committee construction algorithm can be considered as an approximation algorithm for this special case of the k-Polyhedral Separability Problem.

Also, the MASC problem is closely related to several geometrical problems. In particular, in (Khachay and Poberii 2009), NP-hardness of the MASC and MASC-GP(n) problems is proved by the reduction (to these problems) of the well-known Point Covering problem, for which intractability was established in (Megiddo and Tamir 1982).

Further, the inapproximability of the MASC problem is proved in (Khachai 2006) on the basis of intractability results for hypergraph coloring problems obtained in (Dinur et al. 2002) and (Guruswami 2004), which are originated from the famous PCP theorem (Arora and Safra 1998) and approximation threshold bounds for NP-hard problems (Feige 1998).

The multi-stage game considered in Sect. 5 is extending the game-theoretic approach to construction learning algorithms proposed in (Freund 1995) and adopted by many researchers (see, e.g. (Freund and Schapire 1999; Syed and Schapire 2007) and a survey in (Arora et al. 2012)). In turn, this approach is sourcing from the theory of positional statistical games (Wald 1949) and, in particular, from the well-known *fictitious play* approximation method (introduced in Brown 1951) for solving such games.

Learning methods for affine committees are close to other learning techniques for ensemble classifiers such as bagging (Breiman 1996) and boosting (Schapire and Freund 2012). Our main contribution can be considered as an extension of the boosting by sampling technique (Freund 1995) in a way proposed in (Arora et al. 2012).

3 Problem statement

We use the traditional mathematical notation $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{R}$ for the sets of natural, integer, and real numbers, respectively; $|S|$ for the cardinality of a set S; aff X and conv X for the affine and convex hulls of a subset $X \subset \mathbb{R}^n$; and dim L for the dimension of a linear or affine subspace $L \subset \mathbb{R}^n$. For convenience, we introduce subsets A, $B \subset \mathbb{R}^n$ consisting of x_i from sample (1) and defined by the equations

$$A = \{x_i \colon y_i = 1\}, \quad B = \{x_i \colon y_i = -1\}. \tag{3}$$

In addition, to any committee classifier $h(x)$ (defined by equation (2)), we assign the finite sequence $K[h] = (f_1, \ldots, f_q)$ such that $q = \sum_{j=1}^{k} \alpha_j$ and

$$f_1(x) \equiv \cdots \equiv f_{\alpha_1}(x) \equiv c_1^T x - d_1,$$
$$f_{\alpha_1+1}(x) \equiv \cdots \equiv f_{\alpha_1+\alpha_2}(x) \equiv c_2^T x - d_2,$$
$$\cdots$$

$$f_{q-\alpha_k+1}(x) \equiv \cdots \equiv f_q(x) \equiv c_k^T x - d_k.$$

Definition 1 Let A and B be finite non-empty subsets of $\mathbb{R}^n$, and $f_1, \ldots, f_q : \mathbb{R}^n \to \mathbb{R}$. A finite sequence $K = (f_1, \ldots, f_q)$ is called *a committee* separating the subsets A and B if

$$|\{i \in \mathbb{N}_q \ : \ f_i(a) > 0\}| > \frac{q}{2} \quad (a \in A),$$

$$|\{i \in \mathbb{N}_q \ : \ f_i(b) < 0\}| > \frac{q}{2} \quad (b \in B).$$

The number q is called *a length* and the functions $f_1, \ldots, f_q$ are called *members* of the committee K. For brevity, we name a committee of length q as *q-committee*. In this paper, we restrict ourselves to the case of affine committee members f_j; in this case, the committee K is also called *affine*.

It can be easily seen that the classifier $h(x)$ (defined by Eq. (2)) is perfect w.r.t. sample (1) if and only if the sequence $K[h]$ is an affine separating committee for the sets A and B.

The concept of an affine separating committee is a natural discrete generalization of the concept of a separating hyperplane in Euclidean spaces. Indeed, by virtue of the famous separability theorem (see, e.g., Elster et al. 1977), the equality

$$\text{conv}(A) \cap \text{conv}(B) = \varnothing$$

implies the existence of some hyperplane

$$H = \{x \in \mathbb{R}^n : f(x) \equiv c^T x - d = 0\}$$

separating the sets A and B such that

$$f(a) > 0 \quad (a \in A),$$
$$f(b) < 0 \quad (b \in B).$$

Therefore, if sets A and B are separable in the usual sense, then there exists an affine 1-committee separating them. In the general case, sets A and B, for which there exists an affine separating q-committee, are called *q-committee separable*. The following Mazurov's criterion (Mazurov 1971) of separability by affine committees is a natural generalization of the separability theorem[2] mentioned above.

Theorem 1 *Finite sets A and B can be separated by an affine committee if and only if $A \cap B = \varnothing$.*

The proof of Theorem 1 is mostly constructive. In its sufficiency part, for the given sets A and B, an affine q-committee with $q \leq |A \cup B|$ is constructed.

Among all affine committees separating sets A and B, the committees of minimum length are of most interest. By virtue of their extremality, these committees are called *minimum*. Since every q-committee separating A and B can be easily transformed into $(q + 2t)$-committee for the same subsets and an arbitrary $t \in \mathbb{N}$, the length of a minimum committee can be considered as a quantitative characteristic of separability (or *inseparability*) of these sets.

Problem 1 (Minimum Affine Separating Committee (MASC)) For given sets $A, B \subset \mathbb{R}^n$, it is required to find a minimum affine separating committee.

Hereinafter, we consider the special case MASC-GP(n) of the MASC problem, where the dimension $n > 1$ is fixed in advance[3] and the set $A \cup B$ is in *general position*.

---

[2] In case of finite sets.

[3] Is not an input entry of the problem.

Definition 2 A set $D \subset \mathbb{R}^n$, $|D| > n$, is said to be in general position, if, for any subset $D' \subseteq D$, $|D'| = n + 1$, the equality $\dim \operatorname{aff} D' = n$ is valid.

In particular, a set $D \subset \mathbb{R}^2$ is in general position if there does not exist a subset $D' = \{x, y, z\} \subseteq D$ such that the points x, y, and z belong to the same straight line.

4 Known complexity and approximability results

Known results concerning computational complexity and approximation issues of the MASC problem and its subclasses can be divided into two types according to the involved theoretical approach. The results of the first type (we present a brief survey in Sect. 4.1) are mostly based on the theory of linear inequalities, the duality principle for linear programs, and recent results in computational complexity theory (including the PCP theorem Arora and Safra 1998). The results of the second type (we discuss it in Sect. 4.2) originated from the game-theoretic foundations: the von Neumann MinMax theorem, fictitious play technique (Littlestone and Warmuth 1994), etc.

4.1 Approach based on linear inequalities

We start with the overview of known complexity results. The Minimum Affine Separating Committee problem is intractable and is inapproximable unless $P \neq NP$.

Theorem 2 (Khachai 2008) *The MASC problem is strongly NP-hard[4] and remains intractable under the additional constraint*

$$A \cup B \subset \{x \in \{0, 1, 2\}^n : \|x\|_2 \leq 2\}.$$

The MASC problem does not belong to Apx *approximability class[5] unless $P \neq NP$.*

The MASC-GP(n) problem preserves the intractability in the worst case, but this problem has untrivial polynomial-time solvable subclasses and can be approximated much better.

Theorem 3 (Khachay and Poberii 2009) *The MASC-GP(n) problem is polynomially solvable for $n = 1$ and NP-hard for any fixed $n > 1$.*

Further, we proceed with approximability the MASC problem and it's descendants. As mentioned above, Theorem 1 brings us an approximation algorithm for the MASC problem, which is surprisingly efficient but has a poor accuracy in the worst case.

Remark 1 The proof of Theorem 1 provides an approximation algorithm finding a $2m$-approximate solution of the MASC problem in time of $O(m \times T_2)$, where T_2 is the running time of solving a system of two linear equations.

A major part of known approximability results for the MASC-GP(n) problem is based on Theorem 4 (which can be considered as an adaptation of Theorem 1 to the special case of sets in general position).

---

[4] i.e., The existence of a pseudo-polynomial time algorithm for the problem implies the equality $P = NP$.

[5] The class of combinatorial optimization problems that can be approximated in polynomial time within a fixed ratio.

Theorem 4 (Khachai and Rybin 1998) *Suppose* $A, B \subset \mathbb{R}^n$, $A \cap B = \varnothing$, *and* $A \cup B$ *is in general position. For the sets* A *and* B, *there exists an affine q-committee with*

$$q \leq 2 \left\lceil \frac{\lfloor (m-n)/2 \rfloor}{n} \right\rceil + 1, \tag{4}$$

where $m = |A \cup B|$.

The proof of Theorem 4 is based on the following helpful properties of linear inequalities.

Property 1 For any n-subset Z' of a finite set Z located in the n-dimensional Euclidean space E and being in general position, there are two open half-spaces $L_1 = \{x \in E : c_1^T x - d_1 < 0\}$ and $L_2 = \{x \in E : c_2^T x - d_2 < 0\}$ such that $Z \subset L_1 \cup L_2$ and $Z' \subset L_1 \cap L_2$.

Property 2 Let $A' \subset A$ and $B' \subset B$, where $A \cup B$ is in general position and $|A' \cup B'| = n \leq m = |A \cup B|$. Then, there is an affine function $f(x) = c^T x - d$ such that

$$f(a) > 0, \quad (a \in A''), \quad f(b) < 0, \quad (b \in B'')$$

for some $A' \subset A'' \subset A$ and $B' \subset B'' \subset B$, and

$$|A'' \cup B''| \geq \left\lceil \frac{m+n}{2} \right\rceil. \tag{5}$$

Note that Property 2 easily follows from Property 1.

Algorithm 1 Mazurov's affine separating committee construction algorithm

---

Input: a finite set $Z = A \cup B \in \mathbb{R}^n$ being in general position.
Output: an affine separating committee K for the sets A and B.
1: for an arbitrary n-subset $A' \cup B'$ of the given set $A \cup B$, find an affine function f_0 according to Property 2;
2: assign $Z = (A \setminus A') \cup (B \setminus B')$ and $k = 0$;
3: **while** $Z \neq \varnothing$ **do**
4:    assign $k = k + 2$
5:    for some (at most) n-subset $Z'' = A'' \cup B''$, $A'' \subset A$, $B'' \subset B$, take an arbitrary pair of affine functions $f_{k-1} = c_{k-1}^T x - d_{k-1}$ and $f_k = c_k^T x - d_k$ satisfying the system

$$\begin{cases} (f_{k-1}(a) > 0) \wedge (f_k(a) > 0), \; (a \in A''), \\ (f_{k-1}(b) < 0) \wedge (f_k(b) < 0), \; (b \in B''), \\ (f_{k-1}(a) > 0) \vee (f_k(a) > 0), \; (a \in A \setminus A''), \\ (f_{k-1}(b) < 0) \vee (f_k(b) < 0), \; (b \in B \setminus B''); \end{cases}$$

   (the existence of such a pair is guaranteed by Property 1; for technical details, see Khachai and Rybin 1998);
6:    assign $Z = Z \setminus Z''$;
7: **end while**
8: assign to K the affine committee $(f_0, f_1, \ldots, f_k)$.

Actually, the proof of Theorem 4 presents the above Algorithm 1 constructing an affine separating committee for the given sets A and B (an approximate solution of the MASC-GP(n) problem). This algorithm appears to be a simple generalization of the algorithm proposed by Vl. Mazurov much earlier (Mazurov 1971). So, we call Algorithm 1 Mazurov's, as well.

Remark 2 Algorithm 1 has the approximation guarantee of

$$2 \left\lceil \frac{\lfloor (m-n)/2 \rfloor}{n} \right\rceil + 1 = O(m/n) \tag{6}$$

and running time of

$$O(m/n \times T_n), \tag{7}$$

where T_n is the time-complexity bound of solving a Kramer system of linear equations in the n-dimensional Euclidean space.

Remark 3 There exists a non-trivial subclass of sets A and B, for which bounds (4) and (5) are proved to be tight (Khachai 2006). These sets define a polynomial-time solvable subclass of the MASC-GP(n) problem. Given some pair of such sets, Algorithm 1 finds a minimum affine separating committee in time $O(m/n \times T_n)$ as shown in (7). These sets were invented by Gale (1956) and named *uniformly distributed sets*. The formal definition adapted to our purposes is as follows.

Definition 3 A pair of subsets $A, B \subset \mathbb{R}^n$ such that $A \cap B = \varnothing$ and $|A \cup B| = 2k + n$ for some $k \in \mathbb{N}$ is called *uniformly distributed (by Gale)* if, for any non-trivial hyperplane $H = \{x \in \mathbb{R}^n : c^T x - d = 0\}$, there exist $A' \subset A$ and $B' \subset B$, for which

$$c^T a - d > 0, \ (a \in A'), \quad c^T b - d < 0, \ (b \in B'),$$

and $|A' \cup B'| \geq k$.

In (Gale 1956), it is proved that, for any natural numbers n and k, there exists a pair $A, B \subset \mathbb{R}^n$ that is uniformly distributed by Gale.

Hence, in terms of machine learning, we can say that if a training sample induces uniformly distributed subsets A and B, then Algorithm 1 provides a perfect affine committee classifier of minimum length (i.e., making no misclassifications on this sample).

In (Khachay and Poberii 2009), the *GreedyCommittee* approximation algorithm is proposed. This algorithm has the best known approximation ratio and a rather huge but polynomial running time. For brevity, we skip its formal description but recall the main properties of the algorithm in Theorem 5.

In sequel, we call an instance of the MASC-GP(n) problem *nice* if there exists a minimum committee $K = (f_0, f_1, \ldots, f_{q-1})$ such that, for any $t = 1, \ldots, (q-1)/2$, the following conditions

$$(f_t(a) > 0) \vee (f_{t+1}(a) > 0), \quad (a \in A),$$
$$(f_t(b) < 0) \vee (f_{t+1}(b) < 0), \quad (b \in B),$$

are valid.

Theorem 5 *The GreedyCommittee is an approximation algorithm for the MASC-GP(n) problem with running time of $O(m^{3n})$ and approximation ratio of $O(\ln m)$ if the instance to be solved is nice and of $O(m/n)$, otherwise.*

In Sect. 6, we propose the new approximation algorithm *BoostedGreedyCommittee* with a better upper bound of the approximation ratio and the same (by order of magnitude) time-complexity bound.

4.2 Game-theoretic approach

We start with a version of the Multiplicative Weights Update method (Arora et al. 2012) sourcing from (Littlestone and Warmuth 1994) and adapted to Freund's *boosting by sampling* technique (Freund 1995). Recall that we have a finite sample $\xi = ((x_1, y_1), \ldots, (x_m, y_m))$, where y_i are true values of an unknown function, which should be approximated with majority committees defined on some class $\mathcal{F} : [\mathbb{R}^n \to \{-1, 1\}]$ of *weak classifiers* (in our case, $\mathcal{F} = \{\text{sign}(c^T x - d): c \in \mathbb{R}^n, d \in \mathbb{R}\}$).

Without any additional prior information on a classifier to be constructed (besides the training sample ξ), it is convenient to factorize the set $\mathcal{F}$ by this sample as follows. We assume that two weak classifiers $f_1, f_2 \in \mathcal{F}$ are ξ-equivalent if $f_1(x_i) = f_2(x_i)$ for each $i = 1, \ldots, m$. It can be easily verified that the introduced binary relation is a true equivalence. Denote by $\Phi = \mathcal{F}/\xi$ the factor-set induced by this relation. By construction, $\Phi = \{\varphi_1, \ldots, \varphi_k\}$ is a finite set, for which $k \leq 2^m$. For any pair $((x_i, y_i), \varphi_j)$, we introduce *a payoff* g_{ij} of the classifier φ_j on the pattern (x_i, y_i) by the formula

$$g_{ij} = \begin{cases} 1, & \text{if } \varphi_j(x_i) = y_i, \\ -1, & \text{otherwise.} \end{cases} \tag{8}$$

Following (Arora et al. 2012), we consider the iterative process (Algorithm 2).

Algorithm 2 Multiplicative Weights Update algorithm

Input: parameters $0 < \gamma < 1/2$ and $q \in \mathbb{N}$.
Output: a sequence $(\varphi_{j_1}, \ldots, \varphi_{j_q})$.

1: initiate a weight vector $w_0 = [w_{01}, \ldots, w_{0m}]^T$ by the equation $w_{0i} = 1$;
2: **for all** $t = 1, \ldots, q$ **do**
3:    take any $j_t \in \{1, \ldots, k\}$ such that

$$\sum_{i=1}^{m} w_{(t-1)i} \, g_{ij_t} \geq 2\gamma \sum_{i=1}^{m} w_{(t-1)i}; \tag{9}$$

4:    define a new weight vector w_t recurrently by the formula

$$w_{ti} = w_{(t-1)i} \, (1 - \gamma \cdot g_{ij_t})$$

5: **end for**

It is easy to show that, given an arbitrary $\gamma \in (0, 1/2)$, an accuracy level $\varepsilon > 0$, and the appropriate $q = q(\gamma, \varepsilon)$, Algorithm 2 produces a committee with the classification error on the sample ξ less than ε. Indeed, consider the classifier

$$h(x) = \text{sign} \sum_{t=1}^{q} \varphi_{j_t}(x)$$

(of length q). Let $I_h' = \{i \in \mathbb{N}_m : h(x_i) \neq y_i\}$ be the error-set of h. By the majority rule, for any $i \in I_h'$, at least a half of the weak classifiers φ_{j_t} determining h make a mistake at the i-th pattern; i.e., $\varphi_{j_t}(x_i) \neq y_i$ and therefore,

$$w_{qi} \geq (1 + \gamma)^{q/2}(1 - \gamma)^{q/2} = (1 - \gamma^2)^{q/2} \geq e^{-(\gamma^2 + \gamma^4)q/2}. \tag{10}$$

The last bound follows from the inequality $1 - z \geq e^{-(z+z^2)}t$, which is valid for any $0 \leq z \leq 1/2$.

On the other hand, by construction,

$$\sum_{i=1}^{m} w_{qi} = \sum_{i=1}^{m} w_{(q-1)i} \left(1 - \gamma \cdot g_{ij_t}\right) = \sum_{i=1}^{m} w_{(q-1)i} - \gamma \sum_{i=1}^{m} w_{(q-1)i} \, g_{ij_t} \leq$$

$$\leq \left(1 - 2\gamma^2\right) \sum_{i=1}^{m} w_{(q-1)i} \leq \left(1 - 2\gamma^2\right)^q \sum_{i=1}^{m} w_{0i} \leq m e^{-2\gamma^2 q}. \tag{11}$$

Hence,

$$|I_h'| e^{-\gamma^2 q} \leq |I_h'| e^{-(\gamma^2 + \gamma^4)q/2} \leq \sum_{i \in I_h'} w_{qi} \leq \sum_{i=1}^{m} w_{qi} \leq m e^{-2\gamma^2 q}$$

and

$$\frac{|I_h'|}{m} \leq e^{-\gamma^2 q}.$$

To provide the required accuracy level, it is sufficient to take

$$q = q(\gamma, \varepsilon) > \left\lceil \frac{\ln(1/\varepsilon)}{\gamma^2} \right\rceil. \tag{12}$$

Bound (12) is entirely equivalent to the claim of Freund's Boosting theorem (Freund 1995).

Remark 4 Algorithm 2 can be applied to constructing a perfect committee classifier for any non-contradictory sample ξ, for which $A \cap B = \varnothing$. Indeed, to construct a perfect committee classifier w.r.t. the m-sample ξ, we can take any $\varepsilon < 1/m$ (e.g., $\varepsilon = 1/(m+1)$) and any admissible value of $\gamma \in (0, 1/2)$ (in Sect. 6, we discuss this in detail); then, we obtain a committee of length

$$q = \left\lceil \frac{\ln(m+1)}{\gamma^2} \right\rceil$$

in time of $O(mq \times T')$, where T' is the running time of step 3 of Algorithm 2.

Unfortunately, as it follows from Remark 3, the optimal value of the MASC-GP(n) problem can be equal to the right-hand side of (4) and, simultaneously,

$$\gamma \leq \frac{m+n}{2m} - \frac{1}{2} = \frac{n}{2m}.$$

Therefore, the approximation ratio r of Algorithm 2 can be underestimated by

$$r \geq \left\lceil \frac{4m^2 \ln(m+1)}{n^2} \right\rceil \bigg/ \left(2 \left\lceil \frac{\lfloor (m-n)/2 \rfloor}{n} \right\rceil + 1\right) = \Omega \left(\frac{m \ln m}{n}\right),$$

and the running time is of $\Omega(m^3 n^{-2} \ln m \times T')$.

Thus, for the MASC-GP(n) problem, Algorithm 2 is dominated by Algorithm 1, both in the approximation ratio and running time.

In Sect. 6, we combine these algorithms and propose a new approximation algorithm with an improved approximation ratio

$$O\left(\left(\frac{m \ln m}{n}\right)^{1/2}\right)$$

at the expense of an increase in the total running time. We call this algorithm *BoostingCommittee*.

5 Multi-stage zero-sum two-player game

Here, we deal with the following simple observation on the scheme of Algorithm 2. To ensure that Algorithm 2 succeeds, we should choose a value of γ such that, for each $t \in \{1, \ldots, q\}$, there exists j_t satisfying inequality (9). To make such a choice, Y. Freund proposes (Freund 1995) to examine a candidate value of γ w.r.t the set of all probability measures χ, which are defined on the finite measurable space $(I, 2^I)$, where $I = \{1, \ldots, m\}$ is the index set of the training sample $\xi = ((x_1, y_1), \ldots, (x_m, y_m))$.

Namely, in (Freund 1995), it is proposed to choose a value of γ such that, for this value and for any measure $\chi = [\chi_1, \ldots, \chi_m]$, there exists $j = j(\chi) \in \{1, \ldots, n\}$ satisfying the equation

$$\sum_{i=1}^{m} \chi_i \, g_{ij} \geq 2\gamma. \tag{13}$$

Obviously, the above condition is sufficient for γ to be admissible for Algorithm 2, since every weight vector w_t induces an appropriate measure χ_t defined by values $\chi_{ti} = w_{ti}/\sum_{l=1}^{m} w_{tl}$. But this condition is not necessary, as Algorithm 2 might not examine overall possible measures during its runtime. Therefore, Algorithm 2 can succeed[6] even for γ violating (13) for some χ.

The main idea is as follows. We try to run Algorithm 2 with some value of γ taken in advance. There are two alternatives, either the run is successful or Algorithm 2 fails at step 3. If Algorithm 2 succeeds, we obtain a perfect classifier, which can be passed on for further processing.

There exists a way (in numerical optimization, this approach is known as *backtracking*, see, e.g., Boyd and Vandenberghe 2009) to overcome the failure of Algorithm 2 induced by an overestimated value of γ. According to this approach, similarly to AdaBoost (Schapire and Freund 2012), we can decrease the current value of γ each time equation (9) is violated; after that, the computations are resumed until the desired accuracy is achieved. But, as proved in (Arora et al. 2012), this approach could not improve the approximation guarantee of Algorithm 2.

Relying on Lemma 1 and the following game, we propose another approach for constructing an algorithm with a better approximation ratio.

Lemma 1 *Suppose, for a sample ξ, there exists a perfect committee classifier of length q. Then, $\gamma_0 = 1/(2q)$ satisfies Eq. (13) for this sample and for any discrete probability measure χ.*

We prove Lemma 1 in "Appendix 1". It is easily seen that, each time Algorithm 2 fails at step 3, Lemma 1 yields us a new lower bound for the minimum length of any perfect committee classifier for the sample ξ.

Below, we define a multi-round zero-sum game (Game 1) generalizing the scheme of Algorithm 2. As above, we use the notation $I = \{1, \ldots, m\}$ and $J = \{1, \ldots, k\}$, and denote by $\mathcal{X}$ the set of all discrete probability measures defined on $(I, 2^I)$. To any measure $\chi \in \mathcal{X}$, we assign a subset $J[\chi] \subset J$ containing all numbers j, for which condition (13) is valid for

---

[6] If we are lucky.

the given $\gamma \in (0, 1/2)$. Similarly to Algorithm 2, while playing the game, two sequences $\sigma_I = (\chi_0, \chi_1, \ldots)$ and $\sigma_{II} = (j_1, j_2, \ldots)$ are constructed. We call σ_I and σ_{II} strategies of the 1st and 2nd players, respectively. The payoff function $\mathcal{L} : (\sigma_I, \sigma_{II}) \mapsto [0, 1]$ expresses the loss of the 1st player (the gain of the 2nd one), its value is calculated at the end of the game.

Game 1 Committee construction game

Input: an initial distribution χ_0 and a trial value $\gamma \in (0, 1/2)$.
Output: strategies σ_I and σ_{II} of the players and the payoff $\mathcal{L}(\sigma_I, \sigma_{II})$.

1: **for** $t = 1, 2, \ldots$ **do**
2:     **if** $J[\chi_{t-1}] \neq \varnothing$ **then**
3:         2nd player takes any $j_t \in J[\chi_{t-1}]$ and the turn passes to the 1st player;
4:         1st player either breaks the loop and proceeds to Step 9 or chooses a new measure χ_t; then, the loop
         continues with the next value of t;
5:     **else**
6:         the 2nd player breaks the game with the gain $\mathcal{L}(\sigma_I, \sigma_{II}) = 1$;
7:     **end if**
8: **end for**
9: the resulting classifier h is defined by the formula

$$h(x) = \text{sign} \sum_{t=1}^{q} \varphi_{j_t}(x);$$

10: in turn, the value of the payoff function $\mathcal{L}(\sigma_I, \sigma_{II})$ is defined by the equation

$$\mathcal{L}(\sigma_I, \sigma_{II}) = \chi_0(\{i \in I : h(x_i) \neq y_i\}).$$

We summarize the properties of Game 1 in Theorem 6.

Theorem 6 1. *If γ satisfies Eq. (13) for any probability measure χ, then, for each $\varepsilon \in (0, 1)$, the 1st player has a minmax strategy $\sigma_I^* = \sigma_I^*(\varepsilon)$ such that $\mathcal{L}(\sigma_I^*, \sigma_{II}) < \varepsilon$ for any feasible strategy σ_{II} of the 2nd player.*
2. *If Game 1 stops at Step 5, then, for the sample ξ, there are no perfect committee classifiers with length $q \leq 1/(2\gamma)$.*

The proof of Theorem 6 is closely related to the above reasoning concerning Algorithm 2 and partially follows from Lemma 1. We give the proof in "Appendix 2".

Actually, to construct approximation algorithms for the MASC-GP(n) problem in Sect. 6, we need the following straightforward Corollary 1.

Corollary 1 *For any sample $\xi = ((x_1, y_1), \ldots, (x_m, y_m))$ and any $\gamma \in (0, 1/2)$, there are two mutually exclusive alternatives:*

(i) *there exists a perfect committee classifier of length*

$$q = \left\lceil \frac{\ln(m + 1)}{\gamma^2} \right\rceil,$$

which can be constructed while playing Game 1;

(ii) *every perfect committee classifier for the sample ξ (if any) has a length $q > 1/(2\gamma)$.*

To prove Corollary 1, it is sufficient to apply the claim of Theorem 6 for $\chi_0 = [1/m, \dots, 1/m]$ and recall that the inequality $\chi_0(I') < 1/m$ implies $I' = \varnothing$ for any subset $I' \subset I = \{1, \dots, m\}$.

Remark 5 According to Remark 4, the time-complexity of Game 1 can be overestimated by the running time of Algorithm 2, which is of $O(mq \times T')$.

6 Approximation algorithms

6.1 General scheme

We start with the description of Meta-Algorithm 1 providing a general scheme for all the algorithms proposed below. As an outer parameter, Meta-Algorithm 1 uses some learning algorithm $\mathcal{A}$, which is required to construct *a baseline* perfect committee classifier h_b (of length q_b) for the given sample ξ.

Meta-Algorithm 1 Approximation meta-algorithm

Input: a baseline committee construction algorithm $\mathcal{A}$ with upper bounds $r_{\mathcal{A}}$ and $T_{\mathcal{A}}$ of the approximation ratio and running time, respectively; a training sample ξ;
Output: a perfect classifier h for the sample ξ;
1: construct a baseline perfect classifier h_b of length $q_b = 2p - 1$ using the learning algorithm $\mathcal{A}$;
2: find a minimum $1 \le s \le p$ such that alternative (i) of Corollary 1 takes place for $\gamma_s = \gamma = 1/(4s - 2)$;
3: **if** $(4s - 2)^2 \ln(m + 1) < q_b$ **then**
4:     return the perfect committee constructed while playing Game 1 for γ_s;
5: **else**
6:     return h_b.
7: **end if**

W.l.o.g., we assume that $q_b = 2p - 1$ is an odd number (since, otherwise, we can construct another perfect classifier of length $q_b - 1$ by removing any constituent weak classifier). As it follows from Corollary 1, there is a perfect committee classifier of length

$$q_0 = \left\lceil 4q_b^2 \ln(m + 1) \right\rceil,$$

which can be constructed while playing Game 1 for $\gamma = 1/(2q_b)$. The main idea is to find a minimum s, $1 \le s \le p$, such that the same assertion is valid for a committee of length $\left\lceil 4(2s - 1)^2 \ln(m + 1) \right\rceil$. Theorem 7 summarizes the properties of Meta-Algorithm 1.

Theorem 7 *Suppose, $\mathcal{A}$ outputs a baseline committee h_b of length $q_b = 2p - 1$. Also, assume that s is the value taken at Step 2 of Meta-Algorithm 1. Then, Meta-Algorithm 1 constructs an r-approximation of a minimum perfect committee (for the sample ξ) for*

$$r \le \min \left\{ 2 \left(q_b \ln(m + 1) \right)^{1/2}, r_{\mathcal{A}} \right\} \tag{14}$$

in time of

$$O \left(q_b^3 m \ln(m + 1) T'' + T_{\mathcal{A}} \right), \tag{15}$$

where T'' is the time of checking the condition at Step 2 in Game 1.

We prove Theorem 7 in "Appendix 3".

Remark 6 As follows from Theorem 7, the approximation ratio provided by Meta-Algorithm 1 is not always better than the approximation guarantee of the original algorithm $\mathcal{A}$. For instance, if $r_{\mathcal{A}} = O(1)$ then the algorithm $\mathcal{A}$ has the same accuracy bound and appears to be much more preferable w.r.t. the running time, since

$$2\left(q_b \ln(m+1)\right)^{1/2} = \Omega(\sqrt{\ln m}).$$

But, for any algorithm $\mathcal{A}$, whose approximation ratio is highly overestimated ($r_{\mathcal{A}}$ is far from its tight bound), Meta-Algorithm 1 brings much more benefits. Indeed, since $q_b \leq r_{\mathcal{A}} \cdot OPT$ (where OPT is an optimal value of the problem in question), substituting $r_{\mathcal{A}} \cdot OPT$ for q_b in (14), we derive

$$2\left(q_b \ln(m+1)\right)^{1/2} < r_{\mathcal{A}} \iff r_{\mathcal{A}} > 4 \cdot OPT \cdot \ln(m+1).$$

In particular, this is the case for the MASC and the MASC-GP(n) problems, for which, as shown above, the best known approximation ratio is $O(m)$, whereas the lower bound for OPT is 1.

6.2 Approximation algorithms for the MASC-GP(n) problem

In Sect. 6.2, we construct new polynomial-time approximation algorithms for the MASC-GP(n) problem based on Meta-Algorithm 1.

To use the framework provided by Meta-Algorithm 1, for a given sample ξ, at the first stage, we describe a factor-set $\Phi = \{\varphi_1, \ldots, \varphi_k\}$ induced by the sample ξ on the class of affine classifiers. Then, we define a payoff-matrix G with entries (8), which sets up Game 1 in our case. As was proved in Khachay and Poberii (2009), these calculations can be conducted in the time

$$T_{\Phi} = O(|\Phi|) = O\left(\binom{m}{n}\right) = O(m^n). \tag{16}$$

We proceed with the specialization of Meta-Algorithm 1, taking Algorithm 1 as a value of the formal parameter $\mathcal{A}$.

Algorithm 3 BoostedCommittee

Input: a training sample ξ
Output: an affine separating committee h
1: construct the set Φ and the matrix G defined by the sample ξ;
2: apply Meta-Algorithm 1, using Algorithm 1 as a value of the parameter $\mathcal{A}$;
3: assign to h the resulting committee of Meta-Algorithm 1.

Upper bounds for the approximation ratio and running time of Algorithm 3 are given in Theorem 8.

Theorem 8 *Algorithm 3 is an approximation algorithm for the MASC-GP(n) problem with the approximation ratio of*

$$O\left(\left(\frac{m \ln m}{n}\right)^{1/2}\right)$$

and running time of $O(m^{n+4}/n^3 \ln m)$.

We prove Theorem 8 in "Appendix 4".

Remark 7 Comparing Algorithms 1 and 3, we can say that the latter outperforms the former w.r.t. approximation ratio at the expense of a significant increase in running time.

Now, we propose a new approximation algorithm for the MASC-GP(n) problem by combining our Meta-Algorithm 1 and the GreedyCommittee algorithm (mentioned in Sect. 4).

Algorithm 4 BoostedGreedyCommittee

Input: a training sample ξ
Output: an affine separating committee h
1: construct the set Φ and the matrix G defined by the sample ξ;
2: apply Meta-Algorithm 1, using the GreedyCommittee algorithm as a value of the parameter $\mathcal{A}$.
3: assign to h the resulting committee of Meta-Algorithm 1.

As for Algorithm 3, we formulate the properties of Algorithm 4 in Theorem 9.

Theorem 9 *Algorithm 4 has the approximation ratio of*

$$O\left(\left(\frac{m \ln m}{n}\right)^{1/2}\right)$$

in general case and of $O(\ln(m))$ for the nice instances. Its time-complexity is of $O(m^{3n})$ for any $n > 2$.

We prove Theorem 9 in "Appendix 5".

Remark 8 Comparing GreedyCommittee and Algorithm 4, we reveal that the latter improves the former w.r.t. approximation ratio without any increase in time-complexity, at least for $n > 2$.

7 Conclusions and future work

We have considered the approximability issues of the MASC-GP(n) NP-hard combinatorial optimization problem formalizing the machine learning strategy based on the structural risk minimization principle in the class of majority voting piecewise linear classifiers. The desire to improve the approximation ratios of the known approximation algorithms has led us to the new game-theoretic framework (Game 1) extending the Multiplicative Weights Update method and 'boosting by sampling' technique.

We have used the proposed framework to construct the new approximation algorithms BoostedCommittee (Algorithm 3) and BoostedGreedyCommittee (Algorithm 4) with the best known (state-of-the-art) approximation guarantees. Table 1 summarizes the properties of both known and proposed algorithms.

Along with their contribution to computation complexity theory, the results obtained provides a new approach to efficient learning for ensembles of affine classifiers having high generalization ability.

However, some questions still remain open.

(i) As follows from Table 1, any time we succeed in improving an approximation ratio, this implies a significant increase in time-complexity (except, maybe, the pair Greedy-Committee and BoostedGreedyCommittee). The main reason for such an increase is

Table 1 Approximation algorithms for the MASC-GP(n) problem: summary

Name	Approximation ratio		Time-complexity
	General case	Nice instances	
previously known			
Mazurov's algorithm (Algorithm 1)	$O(m/n)$	$O(m/n)$	$O(m/n \times T_n)$
GreedyCommittee algorithm	$O(m/n)$	$O(\ln m)$	$O(m^{3n})$
Multiple Weight Update algorithm (Algorithm 2)	$\Omega\left(\frac{m \ln m}{n}\right)$	?	$\Omega(m^3 n^{-2} \ln m \times T')$
proposed in this paper			
BoostedCommittee (Algorithm 3)	$O\left(\left(\frac{m \ln m}{n}\right)^{1/2}\right)$	$O\left(\left(\frac{m \ln m}{n}\right)^{1/2}\right)$	$O(m^{n+4}/n^3 \ln m)$
BoostedGreedyCommittee (Algorithm 4) for $n > 2$	$O\left(\left(\frac{m \ln m}{n}\right)^{1/2}\right)$	$O(\ln m)$	$O(m^{3n})$

the exhaustive search used at Step 2 of Game 1 to find admissible alternatives of the 2nd player. Therefore, the construction of more efficient search techniques for such alternatives seems to be relevant.

(ii) Game 1 is entirely based on the scheme of the classic Multiplicative Weights Update algorithm; i.e., at any round of the game, it is assumed that the 2nd player chooses a pure strategy in an appropriate matrix game. To improve the overall running time, it seems interesting to investigate an extension of Game 1 providing a wider subset of mixed strategies for the 2nd player.

(iii) Game 1 seems to be extendable in order to describe of approximation algorithms for a wider family of combinatorial optimization problems related to machine learning. We plan to investigate this issue in future work.

(iv) Eventually, the issue marked in Table 1 by the question sign should be investigated, as well.

Appendix 1: Proof of Lemma 1

Consider the $(m \times k)$-matrix G, whose entries g_{ij} are defined in Sect. 4.2 for the given sample ξ. This matrix induces the following pair of mutually dual linear programs

$$
\begin{array}{ll}
P : \min \alpha_1 + \ldots + \alpha_k & D : \max w_1 + \ldots + w_m \\
\quad s.t. & \quad s.t. \\
\quad\quad G\alpha \geq e, \ \alpha \geq 0, & \quad\quad G^T w \leq f, \ w \geq 0,
\end{array} \tag{17}
$$

where $e = [1, \ldots, 1]^T \in \mathbb{R}^m$ and $f = [1, \ldots, 1]^T \in \mathbb{R}^k$. Let a committee $h = \operatorname{sign} \sum_{j=1}^{k} \alpha_{0j} \varphi_j$ with $\sum \alpha_{0j} = q$ be perfect for the sample ξ. Then, the vector $\alpha_0 = [\alpha_{01}, \ldots, \alpha_{0k}]$ is a feasible solution of the program P. Therefore, by the duality theorem, both programs P and D have optimal solutions and the same optimum value $0 < OPT \leq q$. Using the transformation

$$
\gamma = \sum_{i=1}^{m} w_i, \ \chi = \frac{1}{\gamma} w,
$$

to program D, we assign the new equivalent linear program

$$
\begin{aligned}
D' : &\min \gamma \\
&\ s.t. \\
&\quad\quad G^T \chi \leq 2\gamma f, \\
&\quad\quad \sum_{i=1}^{m} \chi_i = 1, \\
&\quad\quad \chi \geq 0, \ \gamma > 0,
\end{aligned}
$$

with the optimum value $OPT' = 1/(2 \cdot OPT) \geq 1/(2q)$. Therefore, for $\gamma_0 = 1/(2q)$, there exists a number $j_0 \in \{1, \ldots, k\}$ such that

$$
\sum_{i=1}^{m} \chi_i \, g_{ij_0} \geq 2\gamma_0.
$$

Lemma 1 is proved.

Appendix 2: Proof of Theorem 6

1. Let the given value of parameter γ satisfy Eq. (13) for any probability measure χ, which is defined on $(I, 2^I)$. In this case, $J[\chi] \neq \varnothing$ and Game 1 never stops at Step 5. W.l.o.g., we can assume that, for each t, the measure $\chi_t = [w_{t1}/W_t, \ldots, w_{tm}/W_t]^T$, where $W_t = \sum_{i=1}^m w_{ti}$.

Let h be the committee classifier constructed at Step 9, and let

$$I_h' = \{i \in I : h(x_i) \neq y_i\}.$$

Consider the 1st-player strategy $\sigma^* = [\chi_0, \chi_1, \ldots, \chi_{q-1}]^T$, for which (as in Algorithm 2) the measure χ_t is defined recurrently by the following equations

$$w_{ti} = \left(1 - \gamma \cdot g_{ij_t}\right) w_{(t-1)i}, \ (i \in I),$$
$$W_t = \sum_{i \in I} w_{ti},$$

and

$$\chi_{ti} = w_{ti}/W_t, \ (i \in I).$$

Similarly to the case of Eqs. (10) and (11), we can find a lower and an upper bounds for $\sum_{i \in I_h'} w_{ti}$. Indeed, for any $i \in I_h'$, the inequality $\varphi_{j_t}(x_i) \neq y_i$ holds at least for a half of $t \in \{1, \ldots, q\}$. Therefore,

$$\sum_{i \in I_h'} w_{qi} \geq \sum_{i \in I_h'} w_{0i}(1 - \gamma)^{q/2}(1 + \gamma)^{q/2} \geq e^{-\gamma^2 q} \sum_{i \in I_h'} w_{0i},$$

since $0 < \gamma < 1/2$.

On the other hand,

$$\sum_{i \in I_h'} w_{qi} \leq \sum_{i=1}^m w_{qi} = \sum_{i=1}^m w_{(q-1)i}\left(1 - \gamma \cdot g_{ij_q}\right) = W_{q-1}\left(1 - \gamma \sum_{i=1}^m \chi_{(q-1)i}\, g_{ij_q}\right) \leq$$
$$\leq W_{q-1}(1 - 2\gamma^2) \leq W_0\left(1 - 2\gamma^2\right)^q \leq W_0 e^{-2\gamma^2 q}.$$

Thus,

$$\chi_0(I_h') = \sum_{i \in I_h'} w_{0i}/W_0 \leq e^{-\gamma^2 q},$$

and, for any

$$q > \frac{\ln(1/\varepsilon)}{\gamma^2},$$

the finite sequence $\sigma_I^* = (\chi_0, \ldots, \chi_{q-1})$ is the required minmax strategy, since $\mathcal{L}(\sigma_I^*, \sigma_{II}) < \varepsilon$ for any feasible strategy σ_{II} of the 2nd-player. The claim is proved.

2. This claim is a simple corollary of Lemma 1. Indeed, as it follows from Lemma 1, for any $q \leq 1/(2\gamma)$, the existence of a perfect committee classifier $h = \text{sign} \sum_{t=1}^q \varphi_{j_t}$ for the sample ξ implies the validity of Eq. (13) for γ and for any measure χ; this contradicts the stop condition of Game 1 at Step 5. Theorem 6 is proved.

Appendix 3: Proof of Theorem 7

Indeed, let OPT be the length of a minimum perfect committee for the sample ξ and let s be the value chosen at Step 2 of Meta-Algorithm 1. By Corollary 1, $OPT \geq 2s - 1$ and there exists a perfect committee classifier h_s of length

$$q_s = \lceil (4s - 2)^2 \ln(m + 1) \rceil.$$

Suppose, $(4s - 2)^2 \ln(m + 1) < q_b$ and Meta-Algorithm 1 returns h_s. Then,

$$\frac{(4s - 2)^2 \ln(m + 1)}{OPT} \leq \frac{4(2s - 1)^2 \ln(m + 1)}{2s - 1} \leq 4(2s - 1) \ln(m + 1) <$$

$$< 2 \left(\frac{q_b}{\ln(m + 1)} \right)^{1/2} \ln(m + 1) = 2 \left(q_b \ln(m + 1) \right)^{1/2}.$$

On the other hand, if $(4s - 2)^2 \ln(m + 1) \geq q_b$ then the resulting committee is h_b of length q_b and

$$\frac{q_b}{OPT} \leq \frac{q_b}{2s - 1} \leq 2 \left(q_b \ln(m + 1) \right)^{1/2}.$$

Thus, Eq. (14) is valid.

Step 2 of Meta-Algorithm 1 involves at most q_b runs of Game 1 for $\gamma_s = 1/(4s - 2)$, each of them has the running time of

$$O \left(m \cdot (4s - 2)^2 \ln(m + 1) \cdot T'' \right) \leq O \left(m \cdot q_b^2 \ln(m + 1) \cdot T'' \right)$$

(see Remark 5). Therefore, the overall running time is of

$$O \left(m \cdot q_b^3 \ln(m + 1) \cdot T'' + T_{\mathcal{A}} \right).$$

Theorem 7 is proved.

Appendix 4: Proof of Theorem 8

The claimed upper bound for the approximation guarantee of Algorithm 3 can be obtained from Theorem 7 by replacing q_b and $r_{\mathcal{A}}$ in Eq. (14) with their bounds defined in Theorem 3 and Remark 2

$$q_b = r_{\mathcal{A}} = 2 \left\lceil \frac{\lfloor (m - n)/2 \rfloor}{n} \right\rceil + 1.$$

Obviously, the running time of Algorithm 3 can be estimated by the sum of the running time of Meta-Algorithm 1 (defined by Eq. 15) and the time T_Φ, which is necessary for constructing the factor-set Φ (see Eq. 16).

Assuming that Step 2 of Game 1 is performed by the exhaustive search against the set Φ, we conclude that T_Φ and T'' have the same upper bound of $O(m^n)$. Therefore, the overall running time of Algorithm 3 is defined by the first term and is equal to $O(m^{n+4}/n^3 \ln m)$. Theorem 8 is proved.

Appendix 5: Proof of Theorem 9

The claim on the approximation ratio follows immediately from Theorems 5 and 7 (Eq. 14).

To prove the upper bound for the running time, it is sufficient to notice that Algorithm 4 differs from Algorithm 3 only in the way of finding a baseline committee classifier. In Algorithm 4, such a committee is found by the GreedyCommittee algorithm, whereas in Algorithm 3, by Mazurov's algorithm (Algorithm 1).

While the time-complexity of Algorithm 1 is vanishingly small compared with the running time of counter-parts of Meta-Algorithm 1, the GreedyCommittee algorithm with the time-complexity of $O(m^{3n})$ (see Theorem 5) prevails over them for any $n > 2$, since, in this case, $m^{n+4}/n^3 \ln m = o(m^{3n})$. Theorem 9 is proved.

References

Ablow, C., & Kaylor, D. (1965). A committee solution of the pattern recognition problem (corresp.). *IEEE Transactions on Information Theory*, *11*, 453–455.

Arora, S., Hazan, E., & Kale, S. (2012). The multiplicative weights update method: A meta-algorithm and applications. *Theory of Computing*, *8*(1), 121–164.

Arora, S., & Safra, S. (1998). Probabilistic checking of proofs: A new characterization of NP. *Journal of the ACM*, *45*, 70–122.

Blum, A., & Rivest, R. (1992). Training a 3-node neural network is NP-complete. *Neural Networks*, *5*, 117–127.

Boyd, S., & Vandenberghe, L. (2009). *Convex optimization*. Cambridge: Cambridge University Press.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140.

Breiman, L. (2001). Using iterated bagging to debias regressions. *Machine Learning*, *45*, 261–277.

Brown, G. W. (1951). Iterative solutions of games by fictitious play. In T. C. Koopmans (Ed.), *Activity Analysis of Production and Allocation* (pp. 374–376). Cowles Commission Monograph No. 13. New York: Wiley.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*, 273–297.

Dinur, I., Regev, O., & Smyth, C. (2002). The hardness of 3-uniform hypergraph coloring. In *Proceedings of the 43rd annual IEEE symposium on founations of computer science*.

Elster, K., Reinhardt, R., Schäuble, M., & Donath, G. (1977). Einführung in die nichtlineare Optimierung. In: *Mathematisch-Naturwissenschaftliche Bibliothek. BSB BG Teubner Verlagsgesellschaft Leipzig*, Vol. 63.

Eremin, I. I. (2002). *Theory of linear optimization*. De Gruyter.

Eremin, I., Mazurov, V., & Astafiev, N. (1983). *Improper linear and convex programming problems*. Moscow: Nauka.

Feige, U. (1998). Threshold of ln *n* for approximating set cover. *Journal of the ACM*, *45*, 634–652.

Freund, Y. (1995). Boosting a weak learning algorithm by majority. *Information and Computation*, *121*, 256–285.

Freund, Y., & Schapire, R. E. (1999). Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, *29*(12), 79–103.

Gale, D. (1956). Neighboring vertices on a convex polyhedron. *Linear inequalities and related system*, *38*, 255–263.

Guruswami, V. (2004). Inapproximability results for set splitting and satisfiability problems with no mixed clauses. *Algorithmica*, *38*, 451–469.

Hofmann, T., Schölkopf, B., & Smola, A. (2008). Kernel methods in machine learning. *The Annals of Statistics*, *36*(3), 1171–1220.

Khachai, M. (2006). Computational complexity of the minimum committee problem and related problems. *Doklady Mathematics*, *73*, 138–141.

Khachai, M. (2008). Computational and approximational complexity of combinatorial problems related to the committee polyhedral separability of finite sets. *Pattern Recognition and Image Analysis*, *18*(2), 236–242.

Khachai, M., Mazurov, V., & Rybin, A. (2002). Committee constructions for solving problems of selection, diagnostics, and prediction. *Proceedings of the Steklov Institute of Mathematics*, *1*, S67–S101.

Khachai, M., & Rybin, A. (1998). A new estimate of the number of members in a minimum committee of a system of linear inequalities. *Pattern Recgnition and Image Analysis*, *8*, 491–496.

Khachay, M. (2007). On the computational complexity of the minimum committee problem. *Journal of Mathematical Modelling and Algorithms*, *6*(4), 547–561.

Khachay, M., & Poberii, M. (2009). Complexity and approximability of committee polyhedral separability of sets in general position. *Informatica*, *20*(2), 217–234.

Kobylkin, K. (2012). Constraint elimination method for the committee problem. *Automation and Remote Control*, *73*, 355–368.

Lin, J., & Vitter, J. (1991). Complexity results on learning by neural nets. *Machine Learning*, *6*, 211–230.

Littlestone, N., & Warmuth, M. (1994). The weighted majority algorithm. *Information and Computation*, *108*, 212–261.

Mazurov, V. (1971). Committees of inequalities systems and the pattern recognition problem. *Kibernetika*, *3*, 140–146.

Mazurov, V. (1990). *Committee method in problems of optimization and classification*. Moscow: Nauka.

Mazurov, V., & Khachai, M. (2004). Committees of systems of linear inequalities. *Automation and Remote Control*, *65*, 193–203.

Mazurov, V., & Khachai, M. (2007). Parallel computations and committee constructions. *Automation and Remote Control*, *68*, 912–921.

Megiddo, N. (1988). On the complexity of polyhedral separability. *Discrete and Computational Geometry*, *3*, 325–337.

Megiddo, N., & Tamir, A. (1982). On the complexity of locating linear facilities in the plane. *Operations Research Letters*, *5*, 194–197.

Nilsson, N. (1965). *Learning machines: Foundations of trainable pattern classifying systems*. New York: McGraw-Hill.

Schapire, R. (1990). The strength of weak learnability. *Machine Learning*, *5*(2), 197–227.

Schapire, R., & Freund, Y. (2012). *Boosting: Foundations and algorithms*. Cambridge: MIT Press.

Schölkopf, B., & Smola, A. (2002). *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. Cambridge: MIT Press.

Syed, U. & Schapire, R. E. (2007). A game-theoretic approach to apprenticeship learning. In *Advances in neural information processing systems 20, proceedings of the twenty-first annual conference on neural information processing systems, Vancouver, British Columbia, Canada* (pp. 1449–1456) 3–6 December 2007.

Wald, A. (1949). Statistical decision functions. *The Annals of Mathematical Statistics*, *20*(2), 165–205.