Machine Learning, 51, 263–281, 2003

Estimating the Optimal Margins of Embeddings in Euclidean Half Spaces

JÜRGEN FORSTER forster@lmi.ruhr-uni-bochum.de
NIELS SCHMITT nschmitt@lmi.ruhr-uni-bochum.de
HANS ULRICH SIMON simon@lmi.ruhr-uni-bochum.de
THORSTEN SUTTORP suttorp@lmi.ruhr-uni-bochum.de
Lehrstuhl Mathematik & Informatik, Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany

Editor: Claudio Gentile

Abstract. Concept classes can canonically be represented by matrices with entries 1 and -1. We use the singular value decomposition of this matrix to determine the optimal margins of embeddings of the concept classes of singletons and of half intervals in homogeneous Euclidean half spaces. For these concept classes the singular value decomposition can be used to construct optimal embeddings and also to prove the corresponding best possible upper bounds on the margin. We show that the optimal margin for embedding n singletons is $\frac{n}{3n-4}$ and that the optimal margin for half intervals over $\{1, \ldots, n\}$ is $\frac{\pi}{2 \ln n} + \Theta(\frac{1}{(\ln n)^2})$. For the upper bounds on the margins we generalize a bound by Forster (2001). We also determine the optimal margin of some concept classes defined by circulant matrices up to a small constant factor, and we discuss the concept classes of monomials to point out limitations of our approach.

Keywords: maximal margins, Euclidean half spaces, singular value decomposition

1. Introduction

Recently there has been a lot of interest in maximal margin classifiers. Learning algorithms that calculate the hyperplane with the largest margin on a sample and use this hyperplane to classify new instances have shown excellent empirical performance (see Cristianini & Shawe-Taylor, 2000). Often the instances are mapped (implicitly when a kernel function is used) to some possibly high dimensional space before the hyperplane with maximal margin is calculated. If the norms of the instances are bounded and a hyperplane with large margin can be found, a bound on the VC-dimension can be applied (Vapnik, 1998; Cristianini & Shawe-Taylor, Theorem 4.16). A small VC-dimension means that a concept class can be learned with a small sample (Vapnik & Chervonenkis, 1971; Blumer et al., 1989; Kearns & Vazirani, 1994, Theorem 3.3). The success of maximal margin classifiers raises the question which concept classes can be embedded in half spaces with a large margin.

Another motivation for studying the margins of embeddings of concept classes is discussed in Forster et al. (2001). There a close connection between margins and the bounded error model of probabilistic communication complexity is shown.

What do we mean by an embedding of a concept class in half spaces? Let $\mathcal{C}$ be a concept class over an instance space X. A k-dimensional Euclidean half space consists of the points

$p \in \mathbb{R}^k$ that lie on one side of some k-dimensional hyperplane. A collection of k-dimensional half spaces is called an embedding of the concept class $\mathcal{C}$ if for all instances $x \in X$ there are points p_x in the unit ball of $\mathbb{R}^k$ such that for each concept $c \in \mathcal{C}$ there is a half space H_c in the collection such that for all instances $x \in X$ the point p_x lies in the half space H_c if and only if $x \in c$. The minimum of the distances of the points p_x to the boundaries of the half spaces is called the margin γ of the embedding.

For technical reasons we will only consider homogeneous half spaces in this paper, i.e. half spaces whose boundaries contain the null vector. Note that it is not really a restriction to assume that the half spaces are homogeneous: There is a standard way to transform an embedding with inhomogeneous half spaces into an embedding with homogeneous half spaces that has at least half the old margin.

For every concept class there is a trivial embedding into half spaces. Ben-David, Eiron, and Simon (2001) show that most concept classes cannot be embedded with a margin that is much larger than the margin of this trivial embedding. They use counting arguments that do not give upper bounds on the margins of particular concept classes. Vapnik (1998) also showed an upper bound on the margin in terms of the VC-dimension. A stronger result was shown by Forster (2001): First note that a concept class $\mathcal{C}$ over an instance space X can be represented by any matrix $M \in \mathbb{R}^{X \times \mathcal{C}}$ for which the entry M_{xc} is positive if $x \in c$ and is negative otherwise. An embedding of such a matrix in homogeneous half spaces is defined analogously to an embedding of a concept class (we require that $p_x \in H_c$ if and only if $M_{xc} > 0$). Forster showed that a matrix $M \in \{-1, 1\}^{X \times \mathcal{C}}$ can only be embedded in homogeneous half spaces with margin at most

$$\gamma \leq \frac{\|M\|}{\sqrt{|X|\,|\mathcal{C}|}}, \tag{1}$$

where $\|M\|$ denotes the operator norm of M.

In this paper we give a straightforward generalization of Forster's result to the case where the entries of the matrix M can be arbitrary real numbers. We introduce a new tool from functional analysis, the singular value decomposition, to estimate the optimal margins of concept classes. The singular value decomposition can not only be used to construct embeddings, but also to show upper bounds on the margins of all embeddings. We show that for two types of concept classes, namely singletons and half intervals, our techniques can be used to calculate the best possible margins exactly. For some concept classes that lie between singletons and half intervals our bounds are tight up to a small constant factor. However, we also show that our singular value decomposition techniques fail for concept classes of monomials.

The paper is organized as follows: In Section 2 we fix some notation for the rest of the paper. In Section 3 we show that a matrix $M \in \mathbb{R}^{X \times Y}$ can only be embedded in homogeneous half spaces with margin at most

$$\frac{\|M\|\sqrt{|X|}}{\sqrt{\sum_{y \in Y} \left(\sum_{x \in X} |M_{xy}| \right)^2}}.$$

Next we show in Section 4 how the singular value decomposition can be used to apply the upper bound from Section 3. In Sections 5 and 6 we use the results from the previous sections to calculate the optimal margins for the concept classes of singletons and of half intervals. In Section 7 we study a family of concept classes that generalizes singletons and half intervals. Finally, we discuss the concept class of monomials in Section 8.

2. Notation and preliminaries

For a finite set X, $\mathbb{R}^X$ is the vector space of real functions ("vectors") on X, and $\mathbb{C}^X$ is the space of complex valued functions on X. The Euclidean norm of a vector $u \in \mathbb{C}^X$ is $\|u\|_2 := \sqrt{\sum_{x\in X} |u_x|^2}$, the supremum norm is $\|u\|_\infty := \max_{x\in X} |u_x|$. As usual we write $\mathbb{R}^n = \mathbb{R}^{\{1,\dots,n\}}$ and $\mathbb{C}^n = \mathbb{C}^{\{1,\dots,n\}}$. The vectors $u \in \mathbb{C}^X$ are column vectors. For two finite sets X, Y we write $\mathbb{C}^{X\times Y}$ for the set of complex matrices with rows indexed by the elements of X and columns indexed by the elements of Y. The transpose of a matrix $A \in \mathbb{C}^{X\times Y}$ is denoted by $A^\top \in \mathbb{C}^{Y\times X}$ and the complex conjugate transpose by $A^* \in \mathbb{C}^{Y\times X}$. We define $e_x \in \mathbb{C}^X$, $x \in X$, to be the canonical base of $\mathbb{C}^X$ for which

$$(e_x)_y = \begin{cases} 1, & x = y, \\ 0, & x \neq y, \end{cases}$$

for $x, y \in X$. For a complex vector $u \in \mathbb{C}^X$, $\bar{u}$ is the complex conjugate and u^* is the complex conjugate transpose of u. I_X is the identity matrix. A nonsingular matrix $A \in \mathbb{R}^{X\times X}$ is called orthogonal if $A^{-1} = A^\top$.

We use the following formal definition of the margin γ of an embedding of a matrix in homogeneous half spaces in the rest of the paper:

Definition 2.1. For finite sets X, Y we say that a matrix $M \in \mathbb{R}^{X\times Y}$ can be embedded in homogeneous half spaces with margin γ if there are vectors u_x, $x \in X$, and v_y, $y \in Y$, that lie in the unit ball of some $\mathbb{R}^k$ (where k can be arbitrarily large) such that M_{xy} and $\langle u_x, v_y\rangle$ have the same sign and $|\langle u_x, v_y\rangle| \geq \gamma$ for all $x \in X$, $y \in Y$.

A vector v_y can be interpreted as a normal vector of the boundary of the homogeneous half space $\{z \in \mathbb{R}^k \mid \langle v_y, z\rangle \geq 0\}$. Then $\langle u_x, v_y\rangle > 0$ means that the vector u_x lies in the interior of this half space. The sign of M_{xy} determines whether u_x must lie in the half space or not. The requirement $|\langle u_x, v_y\rangle| \geq \gamma$ means that the point u_x has distance at least γ from the boundary of the half space. Analogously we can interpret the vectors u_x as normal vectors of half spaces and the vectors v_y as points.

It is crucial that we require the vectors to lie in a unit ball (or that they are bounded) because otherwise we could increase the margin by simply stretching all vectors.

For every matrix $M \in \mathbb{R}^{X\times Y}$ there exists an optimal embedding: We can assume without loss of generality that the vectors of any embedding lie in the unit ball of $\mathbb{R}^X$. (Because the linear span of the vectors u_x has dimension at most $|X|$ and we can project the vectors v_y to this span without changing the scalar products $\langle u_x, v_y\rangle$.) The margin of the embedding is continuous in the vectors u_x, v_y, and the unit ball of $\mathbb{R}^X$ is compact. Thus the maximal margin is attained.

Ben-David, Eiron, and Simon (2001) observed that a matrix $M \in \{-1,+1\}^{X\times Y}$ can always be represented by an arrangement of homogeneous half spaces with margin

$$\gamma = \frac{1}{\sqrt{\min(|X|,|Y|)}}. \tag{2}$$

For the case $|X| \leq |Y|$ this *trivial arrangement* consists of canonical unit vectors and of normalized columns of M: For $x \in X$, we define u_x to be the vector e_x from the canonical base of $\mathbb{R}^X$, and we set $v_y := |X|^{-\frac{1}{2}}(M_{xy})_{x\in X} \in \mathbb{R}^X$ for $y \in Y$. This leads to an arrangement of homogeneous half spaces representing M with margin $|X|^{-\frac{1}{2}}$ (because $\|u_x\|_2 = \|v_y\|_2 = 1$, $\operatorname{sign}\langle u_x, v_y\rangle = M_{xy}$ and $|\langle u_x, v_y\rangle| = |X|^{-\frac{1}{2}}$ for all $x \in X$, $y \in Y$).

In the upper bound (1) on the margin of matrices M with entries ± 1 the operator norm of M appears. The operator norm of an arbitrary matrix $A \in \mathbb{R}^{X\times Y}$ is

$$\|A\| = \sup_{\substack{u\in\mathbb{R}^Y \\ \|u\|\leq 1}} \|Au\| = \max_{\substack{u\in\mathbb{R}^Y \\ \|u\|\leq 1}} \|Au\|.$$

The supremum is attained because $\|Au\|$ is a continuous function of $u \in \mathbb{R}^X$ and the unit ball $\{u \in \mathbb{R}^X \mid \|u\| \leq 1\}$ is compact. It is well known that $\|A\|^2 = \|A^\top A\| = \|AA^\top\|$ for any matrix $A \in \mathbb{R}^{X\times Y}$, and it is not hard to see that for every matrix $M \in \{-1,1\}^{X\times Y}$:

$$\max(\sqrt{|X|}, \sqrt{|Y|}) \leq \|M\| \leq \sqrt{|X|\,|Y|}$$

(see for example Krause, 1996, Lemma 1.1.) The equality $\|M\| = \sqrt{|Y|}$ holds if and only if the rows of M are orthogonal, $\|M\| = \sqrt{|X|}$ holds if and only if the columns of M are orthogonal, and $\|M\| = \sqrt{|X|\,|Y|}$ holds if and only if $\operatorname{rank}(M) = 1$. The Hadamard matrices $H_n \in \mathbb{R}^{2^n\times 2^n}$ are examples of matrices with orthogonal rows and columns. They are recursively defined by

$$H_0 = 1, \quad H_{n+1} = \begin{pmatrix} H_n & H_n \\ H_n & -H_n \end{pmatrix}.$$

From the upper bound (1) on the margin (with $\mathcal{C} = Y$) it follows easily that for matrices $M \in \{-1,1\}^{X\times Y}$ with orthogonal rows or orthogonal columns the trivial embedding has the optimal margin

$$\gamma = \frac{1}{\sqrt{\min(|X|,|Y|)}}. \tag{3}$$

The signum function sign: $\mathbb{R} \to \mathbb{R}$ is given by

$$\operatorname{sign}(x) = \begin{cases} 1, & x > 0, \\ 0, & x = 0, \\ -1, & x < 0. \end{cases}$$

3. An upper bound on the margins of embeddings

It was previously known that the margin of any embedding of a matrix $M \in \{-1, 1\}^{X \times Y}$ in homogeneous Euclidean half spaces is at most $\frac{\|M\|}{\sqrt{|X||Y|}}$ (see Forster, 2001, Theorem 3). In the following theorem we prove a straightforward generalization of this result to matrices $M \in \mathbb{R}^{X \times Y}$ with arbitrary entries. This improved result is needed in the following sections to show new optimal bounds on the margin in some cases.

Theorem 3.1. *Let $M \in \mathbb{R}^{X \times Y}$ be a matrix. Any embedding* (*in the sense described in Section* 1) *of M in homogeneous Euclidean half spaces has margin γ at most*

$$\gamma \le \frac{\sqrt{|X|}\, \|M\|}{\sqrt{\sum_{y \in Y} \left(\sum_{x \in X} |M_{xy}|\right)^2}}.$$

Proof: Let an embedding u_x, v_y of M with margin γ be given. For every $y \in Y$ we have that

$$\gamma \sum_{x \in X} |M_{xy}| \le \sum_{x \in X} M_{xy} \langle u_x, v_y \rangle = \left\langle \sum_{x \in X} M_{xy} u_x, v_y \right\rangle \overset{\|v_y\| \le 1}{\le} \left\| \sum_{x \in X} M_{xy} u_x \right\|, \tag{4}$$

where we used the Cauchy-Schwartz Inequality. We square the above inequality and sum over $y \in Y$:

$$\begin{aligned} \gamma^2 \sum_{y \in Y} \left(\sum_{x \in X} |M_{xy}| \right)^2 &\overset{(4)}{\le} \sum_{y \in Y} \left\langle \sum_{x \in X} M_{xy} u_x, \sum_{\tilde{x} \in X} M_{\tilde{x}y} u_{\tilde{x}} \right\rangle \\ &= \sum_{x, \tilde{x} \in X} (MM^\top)_{x\tilde{x}} \langle u_x, u_{\tilde{x}} \rangle \overset{(*)}{\le} \sum_{x, \tilde{x} \in X} (\|M\|^2 I_X)_{x\tilde{x}} \langle u_x, u_{\tilde{x}} \rangle \\ &= \|M\|^2 \sum_{x \in X} \|u_x\|^2 \le |X|\, \|M\|^2. \end{aligned}$$

Here inequality $(*)$ holds because $A := \|M\|^2 I_X - MM^\top$ and $B := (\langle u_x, u_{\tilde{x}} \rangle)_{x, \tilde{x} \in X}$ are positive semi-definite, thus $\sum_{x, \tilde{x} \in X} A_{x\tilde{x}} B_{x\tilde{x}} \ge 0$ because of Fejer's Theorem (see Horn & Johnson, 1985, Corollary 7.5.4). □

If we apply Theorem 3.1 to a matrix $M \in \{-1, 1\}^{X \times Y}$ with entries -1 and 1 we get the upper bound $\frac{\|M\|}{\sqrt{|X|\,|Y|}}$ from Forster (2001), Theorem 3. For an arbitrary matrix $M \in \mathbb{R}^{X \times Y}$

we also get the upper bound

$$\gamma \le \frac{\sqrt{|Y|}\,\|M\|}{\sqrt{\sum_{x\in X}\left(\sum_{y\in Y}|M_{xy}|\right)^2}}$$

on the margin γ if we apply Theorem 3.1 to $M^\top$.

4. The singular value decomposition

The problem of embedding a matrix $M \in \mathbb{R}^{X\times Y}$ in Euclidean half spaces with a large margin can be stated as follows: We are looking for two matrices B, C with rows of norm 1 such that the signs of the entries of $BC^\top$ are equal to the signs of the entries of M, and such that the smallest absolute value of the entries of $BC^\top$ is as large as possible.

One possibility of writing M as a product of matrices is the *singular value decomposition* of M: Let r be the rank of M. Then there always exist matrices $U \in \mathbb{R}^{X\times r}$ and $V \in \mathbb{R}^{Y\times r}$ with orthonormal columns and nonnegative numbers $s_1, \ldots, s_r$, called the *singular values* of M, such that $M = U\mathrm{diag}(s_1, \ldots, s_r)V^\top$ (see Horn & Johnson, 1985). Obviously the matrices

$$B = U\mathrm{diag}(\sqrt{s_1}, \ldots, \sqrt{s_r}), \quad C = V\mathrm{diag}(\sqrt{s_1}, \ldots, \sqrt{s_r})$$

satisfy $M = BC^\top$. If we normalize the rows of B and C we get an embedding of M. Surprisingly, we can show that for the concept classes of singletons and of half intervals this embedding has the best possible margin. For both of these concept classes we can also use the singular value decomposition to show the optimal upper bound on the margin: We can simply apply Theorem 3.1 to the matrix $UV^\top$. Trying this matrix can be a good idea because it is orthogonal, which means that all its singular values are equal, they are "optimally balanced". Of course we have to check that the entries of $UV^\top$ have correct signs, since this is not true for all matrices M.

Theorem 4.1. *Let $M \in \{-1, 1\}^{X\times Y}$ be a matrix with singular value decomposition $U\mathrm{diag}(s_1, \ldots, s_r)V^\top$. Let $\hat{u}$, $\hat{v}$ be the vectors whose entries are the squared Euclidean norms of the rows of the two matrices $U\mathrm{diag}(\sqrt{s_1}, \ldots, \sqrt{s_r})$ and $V\mathrm{diag}(\sqrt{s_1}, \ldots, \sqrt{s_r})$, i.e.*

$$\hat{u} := \left(\sum_{j=1}^{r} s_j U_{xj}^2\right)_{x\in X} \in \mathbb{R}^X, \quad \hat{v} := \left(\sum_{j=1}^{r} s_j V_{yj}^2\right)_{y\in Y} \in \mathbb{R}^Y.$$

Then the embedding

$$u_x := \frac{1}{\sqrt{\hat{u}_x}}(\sqrt{s_j}U_{xj})_{j=1,\ldots,r} \in \mathbb{R}^r, \quad x \in X, \tag{5}$$

$$v_y := \frac{1}{\sqrt{\hat{v}_y}}(\sqrt{s_j}V_{yj})_{j=1,\ldots,r} \in \mathbb{R}^r, \quad y \in Y, \tag{6}$$

of the matrix M has margin γ

$$\gamma = \frac{1}{\sqrt{\|\hat{u}\|_\infty \|\hat{v}\|_\infty}} \le \frac{\sqrt{|X|\,|Y|}}{\sum_{j=1}^r s_j}. \tag{7}$$

If the entries of M and of $UV^\top$ have the same signs, then every embedding of M in homogeneous Euclidean half spaces has margin γ at most

$$\gamma \le \min\left(\frac{\sqrt{|X|}}{\|\hat{v}\|_2}, \frac{\sqrt{|Y|}}{\|\hat{u}\|_2}\right) \le \frac{\sqrt{|X||Y|}}{\sum_{j=1}^r s_j}. \tag{8}$$

If additionally all norms of the rows of $U\operatorname{diag}(\sqrt{s_1}, \dots, \sqrt{s_r})$ and all norms of the rows of $V\operatorname{diag}(\sqrt{s_1}, \dots, \sqrt{s_r})$ are equal, then the embedding (5), (6) *of M has margin γ*

$$\gamma = \frac{1}{\sqrt{\|\hat{u}\|_\infty \|\hat{v}\|_\infty}} = \frac{\sqrt{|X|\,|Y|}}{\sum_{j=1}^r s_j},$$

and this margin is optimal.

Proof: Obviously $\|u_x\| = 1 = \|v_y\|$ holds, and from $\langle u_x, v_y\rangle = \frac{M_{xy}}{\sqrt{\hat{u}_x \hat{v}_y}}$ it follows that the margin of the embedding is $1/\sqrt{\|\hat{u}\|_\infty \|\hat{v}\|_\infty}$. The upper bound $\sqrt{|X|}/\|\hat{v}\|_2$ on the margin follows if we apply Theorem 3.1 to the matrix $UV^\top$, because $\|UV^\top\| \le \|U\|\|V\| = 1$ and because of

$$\begin{aligned}\sum_{x\in X} |(UV^\top)_{xy}| &= \sum_{x\in X} M_{xy}(UV^\top)_{xy} \\ &= \sum_{x\in X}\sum_{j=1}^r s_j U_{xj} V_{yj} \sum_{k=1}^r U_{xk} V_{yk} = \sum_{j=1}^r s_j V_{yj}^2 = \hat{v}_y.\end{aligned}$$

The first equality holds because the entries of M and $UV^\top$ have the same signs, for the second equality we used that $U\operatorname{diag}(s_1, \dots, s_r)V^\top$ is the singular value decomposition of M, and for the third equality we used that the rows of U are orthonormal. Both the sum of the components of $\hat{u}$ and the sum of those of $\hat{v}$ are equal to $\sum_{j=1}^r s_j$. From this it follows that

$$\begin{aligned}\|\hat{u}\|_2 &\ge \frac{\sum_{j=1}^r s_j}{\sqrt{|X|}}, \quad \|\hat{v}\|_2 \ge \frac{\sum_{j=1}^r s_j}{\sqrt{|Y|}}, \\ \|\hat{u}\|_\infty &\ge \frac{\sum_{j=1}^r s_j}{|X|}, \quad \|\hat{v}\|_\infty \ge \frac{\sum_{j=1}^r s_j}{|Y|},\end{aligned}$$

and this implies the inequalities in (7) and (8). If all the components of $\hat{u}$ are equal, and all the components of $\hat{v}$ are equal, then equality holds in (7) and (8). □

It is easy to see that Theorem 4.1 gives an embedding with optimal margin for matrices that have orthogonal rows or orthogonal columns. Note that it was already observed by Forster (2001) that the margin of the trivial embedding is optimal in this case. In the following two sections we show that Theorem 4.1 can also be used to construct optimal embeddings for the concept classes of singletons and of half intervals.

5. The optimal margins of singleton concept classes

One of the most fundamental concept classes are singletons. For a parameter $n \in \mathbb{N}$, the singleton concept class over an instance space of size n can be represented by the matrix

$$\text{SINGLETONS}_n = \begin{pmatrix} 1 & -1 & \cdots & -1 \\ -1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -1 \\ -1 & \cdots & -1 & 1 \end{pmatrix} \in \{-1, 1\}^{n \times n}.$$

It is obvious that this matrix can be embedded with constant margin: We can get an embedding in inhomogeneous half spaces if we choose the points to be the canonical unit vectors $e_1, \ldots, e_n$ of $\mathbb{R}^n$ and choose the half spaces $\{z \in \mathbb{R}^n \mid \langle e_i, z\rangle \geq \frac{1}{2}\}$ with the canonical unit vectors as normal vectors and thresholds $\frac{1}{2}$. This leads to a margin of $\frac{1}{2}$. Ben-David, Eiron, and Simon (2001) observed that the optimal margin that can be achieved with inhomogeneous half spaces is $\frac{1}{2} + \frac{1}{2(n-1)}$.

We show that Theorem 4.1 can be used to calculate the optimal margin for embeddings with homogeneous half spaces. The matrix SINGLETONS_n is symmetric and has the eigenvalue 2 with eigenspace $\text{null}(M - 2I_n) = \{x \in \mathbb{R}^n \mid \sum_{k=1}^n x_k = 0\}$ and the eigenvalue $2 - n$ with eigenvector $(1, \ldots, 1)^\top$. The eigenvectors

$$a_j = \frac{1}{\sqrt{j^2 + j}}(\underbrace{1, \ldots, 1}_{j \text{ times}}, -j, 0, \ldots, 0)^\top, \quad j = 1, \ldots, n-1,$$

$$a_n = \frac{1}{\sqrt{n}}(1, \ldots, 1)^\top,$$

of M form an orthonormal basis of $\mathbb{R}^n$. From this it follows that a singular value decomposition of M is given by

$$\underbrace{(a_1 \ \cdots \ a_n)}_{U} \underbrace{\text{diag}(2, \ldots, 2, n-2)}_{\text{diag}(s_1, \ldots, s_n)} \underbrace{(a_1 \ \cdots \ a_{n-1} \quad -a_n)^\top}_{V}$$

(to check this we can apply the above to the vectors $a_1, \ldots, a_n$.) For $n \geq 3$ the entries of

$$UV^\top = \frac{1}{2}M + \left(\frac{n}{2} - 2\right)a_n a_n^\top = \begin{pmatrix} 1 - \frac{2}{n} & -\frac{2}{n} & \cdots & -\frac{2}{n} \\ -\frac{2}{n} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\frac{2}{n} \\ -\frac{2}{n} & \cdots & -\frac{2}{n} & 1 - \frac{2}{n} \end{pmatrix}$$

have the same signs as the corresponding entries of SINGLETONS_n and we can apply Theorem 4.1. From

$$\sum_{j=1}^{r} s_j U_{kj}^2 = \sum_{j=1}^{r} s_j V_{kj}^2 = 2 - \frac{2}{k} + 2\underbrace{\sum_{j=k}^{n-1} \overbrace{\frac{1}{j(j+1)}}^{=\frac{1}{j}-\frac{1}{j+1}}}_{=\frac{1}{k}-\frac{1}{n}} + 1 - \frac{2}{n} = 3 - \frac{4}{n} = \frac{3n-4}{n}$$

for $k = 1, \ldots, n$ it follows that

Theorem 5.1. *For $n \geq 3$ the maximal margin of an embedding of the matrix*

$$\text{SINGLETONS}_n = \begin{pmatrix} 1 & -1 & \cdots & -1 \\ -1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -1 \\ -1 & \cdots & -1 & 1 \end{pmatrix} \in \{-1, 1\}^{n \times n}$$

with homogeneous Euclidean half spaces is

$$\frac{n}{3n-4} = \frac{1}{3} + \frac{4}{9n-12} = \frac{1}{3} + \Theta\left(\frac{1}{n}\right).$$

6. The optimal margins of half interval concept classes

Let $X = \{1, \ldots, n\}$ be an instance space of size n. The concept class of half intervals over X consists of the concepts $c = \{1, \ldots, k\}$ for $k = 1, \ldots, n$. This concept class can be represented by the following matrix:

$$\text{HALF-INTERVALS}_n = \begin{pmatrix} 1 & -1 & \cdots & -1 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & -1 \\ 1 & \cdots & \cdots & 1 \end{pmatrix} \in \{-1, 1\}^{n \times n}.$$

As observed by Ben-David (2000), we can use Novikoff's Theorem (Novikoff, 1962) to get an upper bound on the margins of embeddings of this matrix in half spaces: If there exists an embedding with margin γ, then it follows from Novikoff's Theorem that the concept class can be learned with at most γ^{-2} EQ-queries. The learning complexity of HALF-INTERVALS_n with arbitrary EQ-queries is $\lfloor \log_2 n \rfloor$ (see Maass & Turan, 1992, Proposition 4.2.) This shows that $\gamma^{-2} \geq \lfloor \log_2 n \rfloor$, or equivalently $\gamma \leq 1/\sqrt{\lfloor \log_2 n \rfloor}$.

We can show a much stronger result: From Theorem 4.1 we get an exact formula for the optimal margin. In the following we consider only the case that n is even, but the case of n being odd is very similar.

We start by calculating the complex eigenvalues and eigenspaces of M = HALF-INTERVALS$_n$. Let μ be an nth complex root of -1, i.e. $\mu \in \mathbb{C}$, $\mu^n = -1$. Then the vector $x_\mu := (\mu^{k-1})_{k=1,\ldots,n} \in \mathbb{C}^n$ is an eigenvector of M for the eigenvalue $\frac{2\mu}{\mu-1}$:

$$\begin{aligned} Mx_\mu &= \left(\sum_{j=1}^{k} \mu^{j-1} - \sum_{j=k+1}^{n} \mu^{j-1}\right)_{k=1,\ldots,n} = \left(\frac{1-\mu^k}{1-\mu} - \frac{\mu^k+1}{1-\mu}\right)_{k=1,\ldots,n} \\ &= \left(\frac{-2\mu^k}{1-\mu}\right)_{k=1,\ldots,n} = \frac{2\mu}{\mu-1} x_\mu. \end{aligned}$$

Because the eigenvectors x_μ are pairwise orthogonal and because of $\|x_\mu\|^2 = n$ this means that we can write M as

$$M = \sum_{\mu\in\mathbb{C}:\mu^n=-1} \frac{2\mu}{n(\mu-1)} x_\mu x_\mu^*.$$

(To check this we can apply the above to the vectors x_μ for the n complex roots μ of -1.) Now the entries of M can be written as

$$\begin{aligned} M_{jk} &= \frac{2}{n} \sum_{\mu\in\mathbb{C}:\mu^n=-1} \frac{\mu}{\mu-1} \mu^{j-1}\mu^{1-k} \\ &= \frac{2}{n} \sum_{\mu\in\mathbb{C}:\mu^n=-1} \frac{\mu^{j-k+1}}{\mu-1} \overset{\mu=e^{i\pi(2l-1)/n}}{=} \frac{2}{n} \sum_{l=1}^{n} \frac{e^{i\pi(2l-1)(j-k+1)/n}}{e^{i\pi(2l-1)/n}-1} \\ &= \frac{2}{n} \sum_{l=1}^{n/2} \left(\frac{e^{i\pi(2l-1)(j-k+1)/n}}{e^{i\pi(2l-1)/n}-1} + \frac{e^{-i\pi(2l-1)(j-k+1)/n}}{e^{-i\pi(2l-1)/n}-1} \right) \\ &= \frac{2}{n} \sum_{l=1}^{n/2} \frac{e^{i\pi(2l-1)(2j-2k+1)/2n} - e^{-i\pi(2l-1)(2j-2k+1)/2n}}{e^{i\pi(2l-1)/2n} - e^{-i\pi(2l-1)/2n}} \\ &= \frac{2}{n} \sum_{l=1}^{n/2} \frac{\sin \frac{\pi(2l-1)(2j-2k+1)}{2n}}{\sin \frac{\pi(2l-1)}{2n}} \\ &= \frac{2}{n} \sum_{l=1}^{n/2} \left(\sin \frac{\pi(2l-1)}{2n}\right)^{-1} \left(\sin \frac{\pi(2l-1)j}{n} \cos \frac{\pi(2l-1)(2k-1)}{2n} \right. \\ &\qquad \left. - \cos \frac{\pi(2l-1)j}{n} \sin \frac{\pi(2l-1)(2k-1)}{2n} \right). \end{aligned}$$

Thus we can write $M = UDV^\top$, where $U, D, V \in \mathbb{R}^{n\times n}$ are given by

$$U = \left(\sqrt{\frac{2}{n}} \sin \frac{\pi(2l-1)j}{n}; \sqrt{\frac{2}{n}} \cos \frac{\pi(2l-1)j}{n} \right)_{\substack{j=1,\ldots,n \\ l=1,\ldots,n/2}}$$

$$D = \operatorname{diag}\left(\left(\sin\frac{\pi(2l-1)}{2n}\right)^{-1},\ \left(\sin\frac{\pi(2l-1)}{2n}\right)^{-1}\right)_{l=1,\dots,n/2}$$

$$V = \left(\sqrt{\frac{2}{n}}\cos\frac{\pi(2l-1)(2k-1)}{2n};\ -\sqrt{\frac{2}{n}}\sin\frac{\pi(2l-1)(2k-1)}{2n}\right)_{\substack{k=1,\dots,n\\ l=1,\dots,n/2}}.$$

It is not hard to check that the rows of U and V are orthonormal, i.e. $UDV^\top$ is a singular value decomposition of M. The entries of $UV^\top$ have the same signs as those of HALF-INTERVALS$_n$, because for all $j, k \in \{1, \dots, n\}$ (for shortness let $\alpha := \frac{\pi(2j-2k+1)}{2n}$)

$$\begin{aligned}
(UV^\top)_{jk} &= \frac{2}{n}\sum_{l=1}^{n/2}\left(\sin\frac{\pi(2l-1)j}{n}\cos\frac{\pi(2l-1)(2k-1)}{2n}\right.\\
&\quad\left.-\cos\frac{\pi(2l-1)j}{n}\sin\frac{\pi(2l-1)(2k-1)}{2n}\right)\\
&= \frac{2}{n}\sum_{l=1}^{n/2}\underbrace{\sin\frac{\pi(2l-1)(2j-2k+1)}{2n}}_{=\operatorname{Im}(\exp(i(2l-1)\alpha))} = \frac{2}{n}\operatorname{Im}\left(e^{i\alpha}\sum_{l=0}^{n/2-1}(e^{2i\alpha})^l\right)\\
&= \frac{2}{n}\operatorname{Im}\left(\underbrace{\frac{e^{i\alpha}}{1-e^{2i\alpha}}}_{=\frac{1}{e^{-i\alpha}-e^{i\alpha}}=\frac{i}{2\sin\alpha}}(1-\underbrace{e^{in\alpha}}_{=\pm i})\right) = \frac{1}{n\sin\alpha}
\end{aligned}$$

is positive if and only if $j \geq k$. Now we can apply Theorem 4.1, and because the sums

$$\sum_{l=1}^{n} D_{ll}U_{jl}^2 = \sum_{l=1}^{n} D_{ll}V_{kl}^2 = \frac{1}{n}\sum_{l=1}^{n} D_{ll}$$

are equal for all j, k (the above equalities follow immediately from the special structure of the matrices U, D, V) we get

Theorem 6.1. *The maximal margin of an embedding of the concept class of half intervals with matrix*

$$\text{HALF-INTERVALS}_n = \begin{pmatrix} 1 & -1 & \cdots & -1\\ \vdots & \ddots & \ddots & \vdots\\ \vdots & & \ddots & -1\\ 1 & \cdots & \cdots & 1 \end{pmatrix} \in \{-1, 1\}^{n\times n}$$

in Euclidean homogeneous half spaces is

$$\gamma_{\max}(\text{HALF-INTERVALS}_n) = n\left(\sum_{l=1}^{n}\left(\sin\frac{\pi(2l-1)}{2n}\right)^{-1}\right)^{-1}$$
$$= \frac{\pi}{2\ln n} + \Theta\left(\frac{1}{(\ln n)^2}\right).$$

Proof: We still have to show that the optimal margin is asymptotically $\frac{\pi}{2\ln n}$. This is done by proving the following two claims:

Claim 1:

$$\gamma_{\max}(\text{HALF-INTERVALS}_n) \geq \left(\frac{2\ln n}{\pi} + \frac{2\ln 2}{\pi} + 2\right)^{-1} \tag{9}$$

This holds because we can upper bound the sum of the singular values as follows:

$$2\sum_{l=1}^{n/2}\left(\sin\frac{\pi(2l-1)}{2n}\right)^{-1} \leq \underbrace{\frac{2}{\sin\frac{\pi}{2n}}}_{\leq 2n} + 2\int_1^{n/2}\left(\sin\frac{\pi(2x-1)}{2n}\right)^{-1} dx$$
$$\overset{y=\frac{\pi(2x-1)}{2n}}{\leq} \frac{2n}{\pi}\int_{\frac{\pi}{2n}}^{\frac{\pi}{2}-\frac{\pi}{2n}} \frac{1}{\sin y}\, dy + 2n = \frac{2n}{\pi}\left[\ln\frac{\sin y}{1+\cos y}\right]_{\frac{\pi}{2n}}^{\frac{\pi}{2}-\frac{\pi}{2n}} + 2n$$
$$= \frac{2n}{\pi}\left(\ln\underbrace{\frac{\cos\frac{\pi}{2n}}{1+\sin\frac{\pi}{2n}}}_{\leq 1} - \ln\underbrace{\frac{\sin\frac{\pi}{2n}}{1+\cos\frac{\pi}{2n}}}_{\geq \frac{1}{2n}}\right) + 2n \leq \frac{2n}{\pi}\ln(2n) + 2n.$$

We used that $\sin\frac{\pi}{2n} \geq \frac{1}{n}$ for all positive integers n. This follows from the concavity of the sine function on $[0, \pi/2]$.

Claim 2:

$$\gamma_{\max}(\text{HALF-INTERVALS}_n) \leq \left(\frac{2\ln n}{\pi} + \frac{2\ln\frac{2}{\pi}}{\pi}\right)^{-1} \tag{10}$$

This holds because we can lower bound the sum of the singular values as follows:

$$2\sum_{l=1}^{n/2}\left(\sin\frac{\pi(2l-1)}{2n}\right)^{-1} \geq 2\int_1^{(n+1)/2}\left(\sin\frac{\pi(2x-1)}{2n}\right)^{-1} dx$$
$$\overset{y=\frac{\pi(2x-1)}{2n}}{=} \frac{2n}{\pi}\int_{\frac{\pi}{2n}}^{\frac{\pi}{2}} \frac{1}{\sin y}\, dy = \frac{2n}{\pi}\left[\ln\frac{\sin y}{1+\cos y}\right]_{\frac{\pi}{2n}}^{\frac{\pi}{2}}$$
$$= -\frac{2n}{\pi}\ln\underbrace{\frac{\sin\frac{\pi}{2n}}{1+\cos\frac{\pi}{2n}}}_{\leq \frac{\pi}{2n}} \geq \frac{2n}{\pi}\ln\frac{2n}{\pi}.$$

$\square$

7. Circulant matrices

In this section we consider concept classes defined by *circulant matrices*. Circulant matrices are matrices of the form

$$M = \begin{pmatrix} m_1 & m_2 & \cdots & \cdots & m_n \\ m_n & m_1 & m_2 & & \vdots \\ \vdots & m_n & m_1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & m_2 \\ m_2 & \cdots & \cdots & m_n & m_1 \end{pmatrix} \in \mathbb{R}^{n\times n}.$$

The singleton concept classes are simple examples of circulant matrices. Furthermore, a half interval matrix $M = \text{HALF-INTERVALS}_n$ has the same maximal margin as the circulant matrix $\left(\begin{smallmatrix} M & -M \\ -M & M \end{smallmatrix}\right)$.

There is a convenient way to calculate the eigenvectors and eigenvalues of a circulant matrix: Any $n \times n$ circulant matrix M can be written as

$$M = \sum_{l=0}^{n-1} m_{l+1} C^l, \tag{11}$$

where C is the *basic circulant permutation* matrix

$$C = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & & & \ddots & 1 \\ 1 & 0 & \dots & \dots & 0 \end{pmatrix} \in \{0,1\}^{n\times n}.$$

It is easy to check that the eigenvalues of C are $\mu_k = e^{\frac{k}{n}2\pi i}$, $k = 0, \dots, n-1$, and the eigenvectors are $(e^{\frac{jk}{n}2\pi i})_{j=0,\dots,n-1} \in \mathbb{R}^n$, $k = 0, \dots, n-1$. From (11) we see that M has the same eigenvectors and it has eigenvalues

$$\nu_k = \sum_{l=0}^{n-1} m_{l+1} \mu_k^l. \tag{12}$$

Because M has an orthonormal set of n eigenvectors, it follows that M is unitarily diagonalizable, i.e. M is a normal matrix (see Horn & Johnson, 1985, Theorem 2.5.4).

The following theorem gives a lower bound on the best possible margin for normal matrices:

Theorem 7.1. *Let $M \in \{-1, 1\}^{n\times n}$ be a normal matrix with eigenvalues $\nu_1, \dots, \nu_n$. Then there is a unitary matrix $U \in \mathbb{C}^{n\times n}$ such that $M = U \operatorname{diag}(\nu_1, \dots, \nu_n) U^*$. Let $\hat{u} \in \mathbb{R}^n$ be the vector with entries*

$$\hat{u}_j = \sum_{k=1}^{n} |U_{jk}|^2 |\nu_k|,$$

$j = 1, \dots, n$. Then M can be embedded with margin

$$\gamma = \frac{1}{\|\hat{u}\|_\infty} \le \frac{n}{\sum_{k=1}^{n} |\nu_k|}. \tag{13}$$

Equality in (13) *holds if and only if all entries of $\hat{u}$ are equal.*

The proof of Theorem 7.1 is omitted here because it is very similar to the proof of Theorem 4.1.

It is easy to see that for circulant matrices equality holds in (13):

Corollary 7.1. *Let $M \in \{-1, 1\}^{n\times n}$ be a circulant matrix with complex eigenvalues ν_k. Then the maximal margin $\gamma_{\max}(M)$ for embedding M is at least*

$$\gamma_{\max}(M) \ge \frac{n}{\sum_k |\nu_k|}.$$

Proof: We have already seen that we can use the matrix

$$U = \left(\frac{1}{\sqrt{n}} \exp\left(\frac{jk}{n} 2\pi i\right)\right)_{j,k=0,\dots,n-1}$$

in Theorem 7.1. We get immediately that the entries of $\hat{u}$ are equal:

$$\hat{u}_j = \sum_k \left|\frac{1}{\sqrt{n}} e^{\frac{jk}{n} 2\pi i}\right|^2 |\nu_k| = \frac{1}{n} \sum_k |\nu_k|.$$ □

For the following class of circulant matrices we can determine the optimal margin up to a small factor: For positive integers m, n satisfying $m \le \frac{n}{2}$ consider the matrix $Z(m, n) = \sum_{l=0}^{n-1} m_{l+1} C^l$ where

$$m_l = \begin{cases} 1, & l = 1, \dots, m, \\ -1, & l = m+1, \dots, n. \end{cases}$$

For simplicity we only consider the case where m and n are even and m divides n. The following theorem holds for the matrices $Z(m, n)$:

Theorem 7.2. *For $m \le \frac{n}{2}$ the maximal margin for embedding $Z(m, n)$ in Euclidean half spaces is of the order $\gamma_{\max}(Z(m, n)) = \Theta(1/\ln m)$.*

Proof: For the upper bound note that HALF-INTERVALS$_m$ is a submatrix of $Z(m, n)$. It follows that

$$\gamma_{\max}(Z(m,n)) \le \gamma_{\max}(\text{HALF-INTERVALS}_m) \overset{(10)}{\le} \frac{1}{\frac{2}{\pi}\ln m + \frac{2}{\pi}\ln\frac{2}{\pi}}$$

(for the last inequality see the proof of Theorem 6.1).

For the lower bound we use Corollary 7.1. The eigenvalues ν_k of $Z(m, n)$ can be calculated using (12):

$$\nu_k = \sum_{l=0}^{m-1} e^{l\frac{k}{n}2\pi i} - \sum_{l=m}^{n-1} e^{l\frac{k}{n}2\pi i} = 2\frac{1 - e^{m\frac{k}{n}2\pi i}}{1 - e^{\frac{k}{n}2\pi i}}$$

for $0 < k < n$. Some further elementary algebraic transformations lead to $|\nu_k| = 2|\frac{\sin(m\frac{k}{n}\pi)}{\sin(\frac{k}{n}\pi)}|$, $0 < k < n$. For $k = 0$ we get $\nu_0 = 2m - n$.

This means that $Z(m, n)$ can be embedded with margin

$$\gamma_{\max}(Z(m,n)) = \frac{n}{|2m-n| + 4\sum_{k=1}^{\frac{n}{2}-1}\left|\frac{\sin(m\frac{k}{n}\pi)}{\sin(\frac{k}{n}\pi)}\right|},$$

where we used that $|\nu_k| = |\nu_{n-k}|$.

The sum in the denominator can be bounded as follows:

$$\begin{aligned}\sum_{k=1}^{\frac{n}{2}-1}\underbrace{\left|\frac{\sin\left(m\frac{k}{n}\pi\right)}{\sin\left(\frac{k}{n}\pi\right)}\right|}_{\le m} &\le \frac{n}{m}m + \sum_{k=\frac{n}{m}+1}^{\frac{n}{2}} \frac{1}{\sin\left(\frac{k}{n}\pi\right)} \\ &\le n + \int_{\frac{n}{m}}^{\frac{n}{2}} \frac{1}{\sin\left(\frac{k}{n}\pi\right)}\,dl \overset{(*)}{<} n + \frac{n}{\pi}\ln\frac{2}{\frac{2}{m}} = n + \frac{n}{\pi}\ln m.\end{aligned}$$

For $(*)$ we use a calculation similar to that in the proof of Theorem 6.1. It follows that

$$\gamma_{\max}(Z(m,n)) > \frac{n}{|2m-n| + 4\left(n + \frac{n}{\pi}\ln m\right)} \ge \frac{1}{\frac{1}{\pi}\ln m + 4 + \frac{1}{n}}.$$

□

8. Monomials

In this section we discuss the concept class of monomials. We show that there is an embedding of this class with a margin of $\Omega(\frac{1}{n})$, and we prove an upper bound of $O(\frac{1}{\sqrt{n}})$ on the best possible margin. Furthermore, we show that the margin of the embedding from Theorem 4.1 decreases exponentially in n, i.e. the use of the singular value decomposition leads to a very poor margin in the case of monomials.

For the concept class of monomials over n Boolean variables $x_1, \ldots, x_n$, the instances are all possible assignments to these variables. The concepts are all conjunctions of literals $x_1, \neg x_1, \ldots, x_n, \neg x_n$. The matrix MONOMIALS$_n$ = M_n representing this concept class is recursively given by

$$M_0 = (1), \quad M_n = \begin{pmatrix} M_{n-1} & -1 \\ M_{n-1} & M_{n-1} \\ -1 & M_{n-1} \end{pmatrix} \in \{-1, 1\}^{3^n \times 2^n}.$$

For $n = 0$ we have only the empty monomial which is always true. For $n \geq 1$ the first 2^{n-1} columns of M_n correspond to the assignments of the variables for which x_n is true, and the last 2^{n-1} columns correspond to the assignments for which x_n is false. The first 3^{n-1} rows correspond to the monomials m containing the literal x_n, the next 3^{n-1} rows to the monomials containing neither x_n nor $\neg x_n$, and the last rows to the monomials containing $\neg x_n$.

There is an embedding of MONOMIALS$_n$ in inhomogeneous half spaces with margin $1/n$: We map each monomial m to the half space $\{z \in \mathbb{R}^n \mid \langle u_m, z\rangle \geq t_m\}$ given by a normal vector u_m and threshold t_m. The j-th component of $u_m \in \mathbb{R}^n$ is 1 if m contains the positive literal x_j, -1 if m contains $\neg x_j$, and 0 otherwise. If l_m is the number of literals contained in m, we define the threshold t_m as $l_m - 1 \in \mathbb{R}$. For each assignment a of the variables $x_1, \ldots, x_n$ we define a vector $v_a \in \mathbb{R}^n$ by $(v_a)_j = 1$ if a assigns true to x_j and $(v_y)_j = -1$ otherwise. Given a monomial m, the scalar product $\langle u_m, v_a\rangle$ attains its maximal value $l_m = t_m + 1$ for the assignments a that fulfill m. For all other assignments a we have $\langle u_m, v_a\rangle \leq l_m - 2 = t_m - 1$. This shows that $\langle u_m, v_a\rangle > t_m$ if and only if a fulfills m, and that $|\langle u_m, v_a\rangle - t_m| \geq 1$. After dividing the vectors by $\sqrt{n}$ and dividing the thresholds by n we have an embedding with margin $1/n$.

Now we show that the margin of any embedding of MONOMIALS$_n$ with homogeneous half spaces is at most $1/\sqrt{n}$. This can be seen as follows. The matrix MONOMIALS$_n$ has n orthogonal rows: We consider only the monomials that consist of a single positive literal. The corresponding rows of MONOMIALS$_n$ are orthogonal because two distinct literals differ on exactly half of all of the assignments to the variables $x_1, \ldots, x_n$. As noted in Section 1 (see (3)) we cannot embed this $n \times 2^n$-submatrix of MONOMIALS$_n$ with a margin larger than $1/\sqrt{n}$.

A slightly stronger upper bound of $1/\sqrt{n+1}$ on the margin was pointed out to us by one of the anonymous referees. To show this bound we apply Novikoff's Theorem in the same way as described at the beginning of the section on half interval concept classes (Section 6). We use the well known fact that the learning complexity of MONOMIALS with arbitrary EQ-queries is at least $n + 1$.

We want to argue now that the embedding of Theorem 4.1 does not give good margins for the concept classes of monomials. For this we first calculate the sum of the singular values of M_n. Consider the matrix

$$A_n = \frac{1}{2}(M_n + 1_{3^n \times 2^n}) \in \{0, 1\}^{3^n \times 2^n}$$

which results from M_n if we replace the -1 entries of M_n by zeros. To find the singular values of A_n we look at the matrices

$$A_0^\top A_0 = (1), \quad A_n^\top A_n = \begin{pmatrix} 2A_{n-1}^\top A_{n-1} & A_{n-1}^\top A_{n-1} \\ A_{n-1}^\top A_{n-1} & 2A_{n-1}^\top A_{n-1} \end{pmatrix} \in \mathbb{R}^{2^n \times 2^n}.$$

Obviously $x_1^{(0)} = (1)$ is an eigenvector of $A_0^\top A_0$ for the eigenvalue $\lambda_1^{(0)} = 1$. It is easy to see that if $x_j^{(n-1)}$ is an eigenvector of $A_{n-1}^\top A_{n-1}$ for the eigenvalue $\lambda_j^{(n-1)}$ then

$$x_{2j-1}^{(n)} = \begin{pmatrix} x_j^{(n-1)} \\ -x_j^{(n-1)} \end{pmatrix}, \quad x_{2j}^{(n)} = \begin{pmatrix} x_j^{(n-1)} \\ x_j^{(n-1)} \end{pmatrix}$$

are linearly independent eigenvectors of the matrix $A_n^\top A_n$ for the eigenvalues $\lambda_{2j-1}^{(n)} = \lambda_j^{(n-1)}$ and $\lambda_{2j}^{(n)} = 3\lambda_j^{(n-1)}$.

The singular values of A_n are the square roots of the eigenvalues of $A_n^\top A_n$. Thus each singular value $s_j^{(n-1)}$ of A_{n-1} produces two singular values $s_{2j-1}^{(n)} = s_j^{(n-1)}$ and $s_{2j}^{(n)} = \sqrt{3}\, s_j^{(n-1)}$ for A_n. Because of $s_1^{(0)} = 1$ the sum of the singular values of A_n is $\sum_{j=1}^{2^n} s_j^{(n)} = (1+\sqrt{3})^n$.

Because each column of A_n contains exactly 2^n ones, it follows that

$$\begin{aligned} M_n^\top M_n &= (2A_n - 1_{3^n \times 2^n})^\top (2A_n - 1_{3^n \times 2^n}) \\ &= 4A_n^\top A_n - 2A_n^\top 1_{3^n \times 2^n} - 2 \cdot 1_{2^n \times 3^n} A_n + 1_{2^n \times 3^n} 1_{3^n \times 2^n} \\ &= 4A_n^\top A_n + (3^n - 4 \cdot 2^n) \cdot 1_{2^n \times 2^n}. \end{aligned}$$

By construction it follows inductively that each vector $x_j^{(n)}$ for $1 \le j < 2^n$ contains as many 1s as -1s, i.e. $1_{2^n \times 2^n} x_j^{(n)} = 0$. Thus $x_j^{(n)}$ is an eigenvector of $M_n^\top M_n$ for the eigenvalue $4\lambda_j^{(n)}$ for $1 \le j < 2^n$.

The vector $x_{2^n}^{(n)}$ contains only 1s, thus $1_{2^n \times 2^n} x_{2^n}^{(n)} = 2^n x_{2^n}^{(n)}$. Because of $\lambda_{2^n}^{(n)} = 3^n$, this vector is an eigenvector of $M_n^\top M_n$ for the eigenvalue

$$4\lambda_{2^n}^{(n)} + 2^n(3^n - 4 \cdot 2^n) = 4 \cdot 3^n + 6^n - 4^{n+1}.$$

The singular values of M_n are the square roots of the eigenvalues of $M_n^\top M_n$. Thus the sum of the singular values of M_n is almost equal to twice the sum of the singular values of A_n. We just have to add a term that takes care of the special case of the largest eigenvalue of $M_n^\top M_n$. The sum of the singular values of $M_n^\top M_n$ is

$$S := 2((1+\sqrt{3})^n - \sqrt{3^n}) + \sqrt{4 \cdot 3^n + 6^n - 4^{n+1}}.$$

Now it follows from Theorem 4.1 that the margin of the embedding obtained with the singular value decomposition method for the matrix MONOMIALS$_n$ is at most

$$\frac{\sqrt{3^n 2^n}}{S} = O\left(\left(\frac{\sqrt{6}}{1+\sqrt{3}}\right)^n\right).$$

Because of $\sqrt{6}/(1+\sqrt{3}) \approx 0.8966 < 1$ this margin is exponentially small in n. Since we have already seen that there is an embedding of MONOMIALS$_n$ with margin $1/n$, Inequality (8) of Theorem 4.1 does not hold for all embeddings of MONOMIALS$_n$. In particular it follows that for large values of n not all entries of the matrix $UV^\top$ given by a singular value decomposition of $\mathrm{MONOMIALS}_n$ have the correct signs.

9. Conclusion and open problems

We have calculated the optimal margins $\gamma = n/(3n-4)$ for embedding singleton concept classes and $\gamma = \frac{\pi}{2\ln n} + \Theta(\frac{1}{(\ln n)^2})$ for half intervals. We also considered a more general case, namely the matrices $Z(m,n)$ of Section 7. The optimal margin of these was shown to be of the order $(\ln m)^{-1}$. An intuitive interpretation of these results is still missing.

There is a gap between our upper and lower bounds for monomials: We only know that monomials can be embedded with margin $\Omega(\frac{1}{n})$, and that the margin cannot be larger than $O(\frac{1}{\sqrt{n}})$.

For better upper bounds it would be interesting to know if there are other ways of finding matrices M that give stronger bounds in Theorem 3.1.

We have also seen how the singular value decomposition can be used to construct embeddings, i.e. to show lower bounds on the margin. However, these embeddings have very poor margins for monomials. Thus new techniques for finding nontrivial embeddings of concept classes are needed.

Acknowledgments

We thank Dietrich Braess for pointing out how circulant matrices can be used to simplify the presentation of the results of Section 7. Also many thanks to Shai Ben-David for interesting discussions. Jürgen Forster was supported by the Deutsche Forschungsgemeinschaft grant SI 498/4-1. This research was also supported by a grant from the G.I.F., the German-Israeli Foundation for Scientific Research and Development.

References

Ben-David, S. (2000). Personal Communication.

Ben-David, S., Eiron, N., & Simon, H. U. (2001). Limitations of learning via embeddings in Euclidean half-spaces. To appear in the *Journal of Machine Learning Research*. A preliminary version appeared in the *Proceedings of the 14th Annual Conference on Computational Learning Theory* (pp. 385–401). Berlin: Springer.

Blumer, A., Ehrenfeucht, A., Haussler, D., & Warmuth, M. K. (1989). Learnability and the Vapnik-Chervonenkis dimension. *Journal of the ACM, 100*, 157–184.

Cristianini, N., & Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines*. Cambridge, United Kingdom: Cambridge University Press.

Forster, J. (2001). A linear lower bound on the unbounded error probabilistic communication complexity. In *Proceedings of the 16th Annual IEEE Conference on Computational Complexity* (pp. 100–106). IEEE Computer Society Press.

Forster, J., Krause, M., Lokam, S. V., Mubarakzjanov, R., Schmitt, N., & Simon, H. U. (2001). Relations between communication complexity, linear arrangements and computational complexity. In *Proceedings of the 21st Conference on Foundations of Software Technology and Theoretical Computer Science* (pp. 171–182). Berlin: Springer.

Horn, R. A., & Johnson, C. R. (1985). *Matrix Analysis*. Cambridge, United Kingdom: Cambridge University Press.
Kearns, M. J., & Vazirani, U. V. (1994). *An Introduction to Computational Learning Theory*. Cambridge, Massachusetts: Massachusetts Institute of Technology.
Krause, M. (1996). Geometric arguments yield better bounds for threshold circuits and distributed computing. *Theoretical Computer Science, 156*, 99–117.
Maass, W., & Turan, G. (1992). Lower bound methods and separation results for on-line learning models. *Journal of Machine Learning, 9*, 107–145.
Novikoff, A. B. (1962). On convergence proofs on perceptrons. *Symposium on the Mathematical Theory of Automata, 12* (pp. 615–622). Polytechnic Institute of Brooklyn.
Vapnik, V. (1998). *Statistical Learning Theory*. New York: John Wiley & Sons, Inc.
Vapnik, V. N., & Chervonenkis, A. Y. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications, 16*, 264–280.

Received March 8, 2002
Revised September 1, 2002
Accepted October 21, 2002
Final manuscript November 11, 2002