Screening Rules for the Support Vector Machine and the Minimum Enclosing Ball

Bachelor Thesis

Author(s):
Wysling, Julia

Publication date:
2015

Permanent link:
https://doi.org/10.3929/ethz-a-010482844

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Screening Rules for the Support Vector Machine and the Minimum Enclosing Ball

Bachelor Thesis

Julia Wysling

Sunday 15th February, 2015

Advisors: Prof. Dr. Bernd Gärtner, Dr. Martin Jaggi

D-MATH, ETH Zürich

Abstract

The support vector machine and the minimum enclosing ball are both extensively studied tools in classification of multidimensional data. In this thesis I study the connection between the two tools. Furthermore I propose screening rules for the SVM problem that accelerate the running time of existing algorithms which solve the problems.

Contents

Chapter 1

Introduction

The support vector machine (SVM) is one of the most popular and widely applied machine learning tools in classification of multidimensional data. Given labeled training data (the label indicates to which class a datum belongs) the task is to compute a classifier such that new data can be categorised, i.e. such that a label can be attached to it. This tool has been successfully applied to a variety of data mining applications such as text categorization ([5]), bioinformatics (for example protein structure prediction [6]) and image processing (for example for face recognition in images [9]).

The minimum enclosing ball (MEB) is also an extensively studied tool in classification of multidimensional data. Given a finite set of input points the task is to find the smallest closed ball that contains all of the input data.

The emergence of big-data analysis poses new challenges for classification tools: With data that consists of tens of millions of high-dimensional entries the need for screening rules which reduce the input size in order to accelerate the search for a solution increased.

In this thesis I develop new screening rules using the similarity between the minimum enclosing ball problem and the support vector machine problem as well as using basic geometric considerations.

In Chapter 2 I introduce the support vector machine problem and the minimum enclosing ball problem and show how these two problems are connected.

In Chapter 3 I introduce the concept of screening and give some insights into its effects and the procedure. Furthermore I establish how I will assess the quality of a screening rule in the further chapters of my thesis.

In Chapter 4 I give some insight into existing screening rules for the MEB-problem and present how they contribute to the acceleration of solving the problem.

In Chapter 5 I establish a screening rule for the SVM-problem based on a screening rule for the MEB-problem. I show how this screening rule connects with the screening rule for the MEB-problem and assess its quality.

In Chapter 6.3 I establish the main result of this thesis: A screening rule based on a geometric idea. Again I assess it's quality based on the criterion defined in Chapter 3.

In the last chapter of this thesis (chapter 7) I give an Outlook how this work could be extended.

Chapter 2

Support Vector Machines and Minimum Enclosing Balls

In this chapter I will first give two definitions which are preliminaries for the rest of the definitions in this chapter. Then I will formally introduce the concepts of the support vector machine and the minimum enclosing ball and show how these two problems are connected.

2.1 Preliminaries

The following two definitions are preliminaries for the formal definition of the support vector machine problem and the minimum enclosing ball problem:

Definition 2.1 (Unit simplex [4]) *The unit simplex* $\triangle$ *is the set of all non-negative vectors in* $\mathbb{R}^n$ *whose entries sum up to one, i.e.* $\triangle = \{\mathbf{x} \in \mathbb{R}^n \,|\, \sum_{j=1}^{n} x_j = 1, \mathbf{x} \geq 0\}$.

Definition 2.2 (Feasible region / Feasible solution) *The feasible region of an optimization problem is the set of all points that satisfy the constraints of the problem.*

A feasible solution of an optimization problem is an element of the feasible region.

Example 2.3 (Feasible region) *The convex optimization problem*

$$\begin{aligned} \min \quad & \mathbf{x}^2 \\ s.t. \quad & \mathbf{x} \in \triangle \end{aligned}$$

has the unit simplex $\triangle$ *as feasible region.*

2.2 Support Vector Machines (SVM)

The basic idea of the *one-class* support vector machine problem is to find a hyperplane that, given a set of input points $\mathbf{a}_i$, has a maximal distance from the origin and has the property to separate the points in the input set from the origin.

The basic idea of the *two-class* support vector machine problem is to find a (classifying) hyperplane that, given a set of input points $\mathbf{b}_i$ which each have a binary label $\mathbf{y}_i \in \{\pm 1\}$, has a maximal distance to all input points and has the property to separate the points with a positive label from the points with a negative label. Here we only consider hyperplanes that pass through the origin [1].

Since the assumption of linearly separable data sets is not realistic for most applications we often consider the soft-margin SVM: A soft-margin SVM allows points to lie on the wrong side of the classifier if perfect linear separation is impossible. These outliers are punished in the objective function by their squared distance from the classification hyperplane (ℓ_2-loss).

In the following chapters I will consider those support vector machine variants whose dual optimization problem has the following form:

Definition 2.4 (Support vector machine problem (SVM-problem) [4]) *Given a set of n input points $\mathcal{A} := \{\mathbf{a}_1, ..., \mathbf{a}_n\} \subset \mathbb{R}^d$ we can denote the support vector machine problem (SVM-problem) as the following dual convex optimization problem:*

$$\begin{array}{ll} \min & \|A\mathbf{x}\|^2 \\ s.t. & \mathbf{x} \in \triangle \end{array}$$

where the j-th column of the matrix $A \in \mathbb{R}^{dxn}$ is formed by the d coordinates of the point $\mathbf{a}_j \in \mathbb{R}^d$ of the input set $\mathcal{A}$.

Observation 2.5 ([4]) *We can reduce the two-class SVM-problem to a one-class SVM-problem which takes the form of Definition 2.4 using the following procedure: Given a set of n input points $\mathcal{B} := \{\mathbf{b}_1, ..., \mathbf{b}_n\} \subset \mathbb{R}^d$ together with their binary labels $\mathbf{y}_i \in \{\pm 1\} \quad \forall i \in \{1, ..., n\}$ we can reflect all points with label $\mathbf{y}_i = -1$ at the origin (illustrated in Figure 2.1). The two-class SVM-problem reduces to solving the one-class SVM-problem with points $\mathbf{a}_i = \mathbf{y}_i \mathbf{b}_i \quad \forall i \in \{1, ..., n\}$.*

[1] There is a well-known trick to emulate the SVM-problem with classifiying hyperplanes that do not necessarily pass through the origin by the SVM-problem only considering hyperplanes that pass through the origin. The idea is to simply increase the dimensionality of the input points and the classifier by one [4].

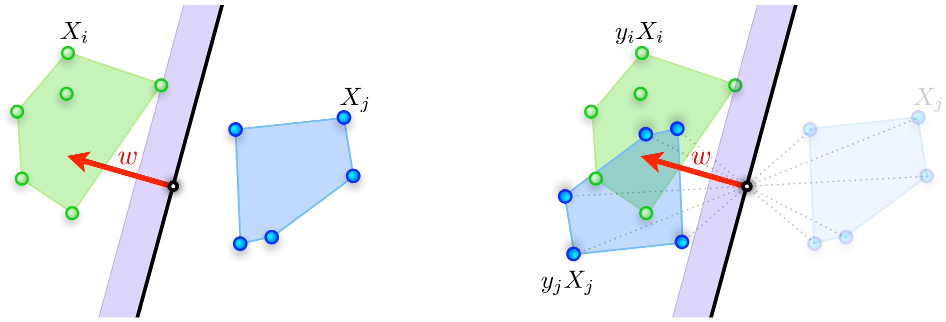


Figure 2.1: Conversion of a two-class SVM-problem into a one-class SVM-problem (source of figure: [4])

Due to this observation I will only consider the one-class SVM-problem from now on. Of course all results could be translated to the two-class SVM-problem.

The convex optimization problem formulation in Definition 2.4 includes the commonly used soft-margin SVM with ℓ_2-loss and (as already explained) both the one- and the two-class SVM variant [4].

Definition 2.6 (Candidate classifier [4]) *Any feasible solution* $\mathbf{x}$ *of the SVM-problem gives us a candidate classifier* $\mathbf{w} = A\mathbf{x}$.

Observation 2.7 *The candidate classifier* $\mathbf{w}$ *is a convex combination of the input points since* $\mathbf{x} \in \mathbb{R}^n$ *lies in the unit simplex* $\triangle$.

So the SVM-problem from Definition 2.4 can be rewritten to be:

$$\begin{aligned} \min \quad & \|\mathbf{w}\|^2 \\ s.t. \quad & \mathbf{w} \in conv(\mathcal{A}) \end{aligned}$$

which allows the following geometric interpretation: The goal of the SVM-problem is to find the element of the convex hull of the input points $\mathcal{A}$ that is closest to the origin.

Theorem 2.8 *For the SVM-problem from Definition 2.4 it holds that*

a) the problem has an optimal solution $\mathbf{x}^*$

and

b) there exists a unique $\mathbf{w}^*$ *such that* $\mathbf{w}^* = A\mathbf{x}^*$ *holds for every optimal solution* $\mathbf{x}^*$

Proof a) The feasible region of our optimization problem is a compact set (the unit simplex) and we want to minimize a continuous function over it
$\Rightarrow$ There exists an optimal solution $\mathbf{x}^*$.

b) We prove the uniqueness of $\mathbf{w}^*$ by contradiction.

Consider two optimal solutions $\mathbf{y}^*$ and $\mathbf{z}^*$ to the SVM-problem with their associated classifiers $\mathbf{u}^*$ and $\mathbf{v}^*$, i.e. $\|\mathbf{u}^*\|^2 = \|A\mathbf{y}^*\|^2$ and $\|\mathbf{v}^*\|^2 = \|A\mathbf{z}^*\|^2$.

It holds that $\|\mathbf{u}^*\|^2 = \|A\mathbf{y}^*\|^2 = \|A\mathbf{z}^*\|^2 = \|\mathbf{v}^*\|^2 := \tau$ since $\|A\mathbf{y}^*\|^2$ and $\|A\mathbf{z}^*\|^2$ are both minimal and therefore take the same value.

We know that $\mathbf{u}^*$ and $\mathbf{v}^*$ are both elements of the convex hull of $\mathcal{A}$ and therefore $\lambda\mathbf{u}^* + (1-\lambda)\mathbf{v}^* \in conv(\mathcal{A}) \quad \forall 0 \leq \lambda \leq 1$.

$\|\lambda\mathbf{u}^* + (1-\lambda)\mathbf{v}^*\| < \lambda\|\mathbf{u}^*\| + (1-\lambda)\|\mathbf{v}^*\| = \tau$ since $\|\cdot\|$ is strictly convex. It follows that $\mathbf{y}^*$ and $\mathbf{z}^*$ can't be optimal and therefore we have contradiction. □

In the process of finding an optimal solution to the SVM-problem we make use of some additional definitions:

Definition 2.9 (Margin of separation) *Given a candidate classifier* $\mathbf{w}$ *for our SVM-problem we call a maximal* σ *with the property that* $\langle \mathbf{a}_i, \mathbf{w}\rangle \geq \sigma \quad \forall \mathbf{a}_i \in \mathcal{A}$ *margin of separation.*

Definition 2.10 ((1 − δ)-approximation) *Given* $\delta \geq 0$ *a candidate classifier* $\mathbf{w}$ *is said to be a* $(1-\delta)$*-approximation to an optimal solution* $\mathbf{w}^*$ *of the SVM-problem if* $\sigma \geq (1-\delta)\|\mathbf{w}\|$ *where* σ *is the margin of separation of* $\mathbf{w}$ *(see Figure 2.2).*

Special case: *In the case that* $\delta > 1$ *this definition still holds (see Figure 2.3).*

Properties 2.11 *It holds that* $\|\mathbf{w}^*\| \geq (1-\delta)\|\mathbf{w}\|$

For the support vector machine problem there are points of our input set which contribute to the optimal solution (so called support vectors) and points that don't (so called non support vectors). These points can be defined based on our convex optimization problem formulation:

Definition 2.12 (Support vector (SV) [4]) *The datapoints corresponding to non-zero entries in an optimal solution* $\mathbf{x}^*$ *of the SVM-problem are called support vectors. All other datapoints are called non support vectors.*

Properties 2.13

i) $\langle \mathbf{a}_i, \mathbf{w}^*\rangle = \|\mathbf{w}^*\|^2 = \|A\mathbf{x}^*\|^2$ *for a support vector* $\mathbf{a}_i$

ii) $\|\mathbf{a}_i - \mathbf{w}^*\|^2 = \|\mathbf{a}_i\|^2 - \|\mathbf{w}^*\|^2 = \|\mathbf{a}_i\|^2 - \|A\mathbf{x}^*\|^2$ *for a support vector* $\mathbf{a}_i$

iii) $\|\mathbf{a}_j - \mathbf{w}^*\|^2 < \|\mathbf{a}_j\|^2 - \|\mathbf{w}^*\|^2$ *for a non support vector* $\mathbf{a}_j$

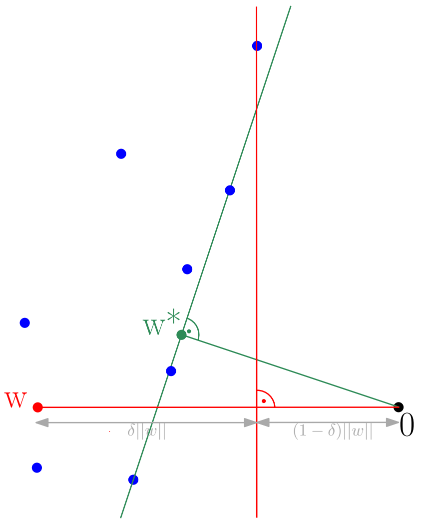


Figure 2.2: (1-δ)-approximation for case $\delta \leq 1$

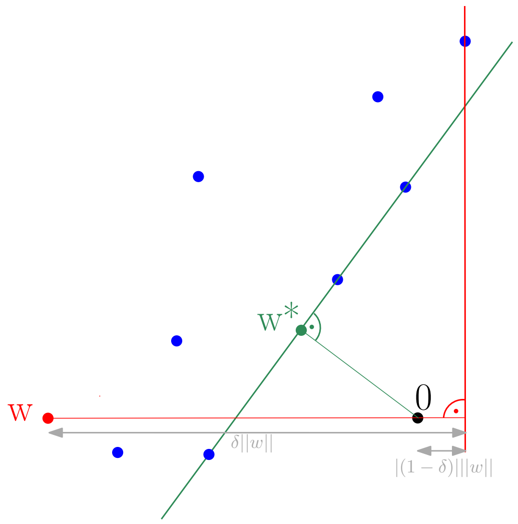


Figure 2.3: (1-δ)-approximation for case $\delta > 1$

Proof

i) $\langle \mathbf{a}_i, \mathbf{w}^* \rangle = \|\mathbf{a}_i\| \|\mathbf{w}^*\| cos(\sphericalangle(\mathbf{a}_i, \mathbf{w}^*)) \stackrel{1.}{=} \|\mathbf{w}^*\|^2 \stackrel{2.}{=} \|A\mathbf{x}^*\|^2$

Where we used:

1. $\mathbf{a}_i$ is support vector and therefore $cos(\sphericalangle(\mathbf{a}_i, \mathbf{w}^*))\|\mathbf{a}_i\| = \|\mathbf{w}^*\|$

2. Theorem 2.8 ($\|\mathbf{w}^*\|^2 = \|A\mathbf{x}^*\|^2$)

ii) $\|\mathbf{a}_i - \mathbf{w}^*\|^2 \stackrel{1.}{=} \|\mathbf{a}_i\|^2 - \|\mathbf{w}^*\|^2 \stackrel{2.}{=} \|\mathbf{a}_i\|^2 - \|A\mathbf{x}^*\|^2$

Where we used:

1. Pythagorean Theorem

2. Theorem 2.8 ($\|\mathbf{w}^*\|^2 = \|A\mathbf{x}^*\|^2$)

iii) For all non support vectors $\mathbf{a}_j$ we have that $\langle \mathbf{a}_j, \mathbf{w}^* \rangle > \|\mathbf{w}^*\|^2$ (otherwise $\mathbf{a}_j$ would be a support vector or would lie on the wrong side of the classifier). Denote the distance between $\mathbf{a}_j$ and $\langle \mathbf{a}_j, \mathbf{w}^* \rangle \mathbf{w}^*$ by d. Applying the Pythagorean Theorem and the Triangle Inequality it follows that: $\|\mathbf{a}_j\|^2 - \|\mathbf{w}^*\|^2 > d > \|\mathbf{a}_j - \mathbf{w}^*\|^2$. □

Based on the idea of the Hemisphere Lemma for the minimum enclosing ball which I will state later on (Proposition 2.21) I propose and prove the Hemisphere Lemma for the SVM:

Lemma 2.14 (Hemisphere Lemma for SVM) *Given a set of n input points* $\mathcal{A} := \{\mathbf{a}_1, ..., \mathbf{a}_n\} \subset \mathbb{R}^d$ *with an optimal classifier* $\mathbf{w}^*$*. Then any closed halfspace that contains* $\mathbf{w}^*$ *also contains at least one support vector* $\mathbf{a}_j \in \mathcal{A}$*.*

Proof I prove the Hemisphere Lemma for SVM by contradiction: Suppose there exists a halfspace $\mathcal{H}$ that doesn't contain a support vector $\mathbf{a}_j \in \mathcal{A}$.

We know that $\mathbf{w}^*$ is in the convex hull of the support vectors, i.e. $\mathbf{w}^* \in conv(\mathcal{A}^*)$ where $\mathcal{A}^* \subset \mathcal{A}$ is the set of all support vectors of an optimal solution to the SVM-problem. Since $\mathcal{H}$ does not contain any support vectors $\mathcal{A}^* \cap \mathcal{H} = \emptyset$ which implies that $conv(\mathcal{A}^*) \cap \mathcal{H} = \emptyset$. Since $\mathbf{w}^* \in conv(\mathcal{A}^*) \cap \mathcal{H}$ we have contradiction and therefore there does not exist a halfspace $\mathcal{H}$ that doesn't contain a support vector. □

2.3 Minimum Enclosing Balls (MEB)

In this section I will introduce the Minimum Enclosing Ball problem. The problem setting is the following: Given n input points in $\mathbb{R}^d$ we want to find a ball of smallest radius that contains all the points of the input set ([8]).

Definition 2.15 (MEB-problem [8]) *Given a set of n input points* $\mathcal{A} := \{\mathbf{a}_1, ..., \mathbf{a}_n\} \subset \mathbb{R}^d$ *we can denote the minimum enclosing ball problem (MEB-problem) as the following convex optimization problem:*

$$\begin{aligned} \min \quad & \mathbf{x}^T A^T A \mathbf{x} - \sum_{j=1}^{n} x_j \mathbf{a}_j^T \mathbf{a}_j \\ s.t. \quad & \mathbf{x} \in \triangle \end{aligned}$$

where the j-th column of the matrix $A \in \mathbb{R}^{dxn}$ *is formed by the d coordinates of the point* $\mathbf{a}_j \in \mathbb{R}^d$ *of the input set* $\mathcal{A}$*.*

Theorem 2.16 ([8]) *For the MEB-problem from Definition 2.15 it holds that*

a) the problem has an optimal solution $\mathbf{x}^*$

and

b) there exists a point $\mathbf{c}^*$ *such that* $\mathbf{c}^* = A\mathbf{x}^*$ *holds for every optimal solution* $\mathbf{x}^*$*. Moreover, the ball with center* $\mathbf{c}^*$ *and squared radius* $-\mathbf{x}^{*T} A^T A \mathbf{x}^* + \sum_{j=1}^{n} x_j^* \mathbf{a}_j^T \mathbf{a}_j$ *is the unique ball of smallest radius containing* $\mathcal{A}$*.*

This theorem is proved in [8].

In the process of finding an optimal solution to the MEB-problem we make use of some additional definitions:

Definition 2.17 (Ball) *We define a ball in* $\mathbb{R}^d$ *to be the set* $\mathcal{B}_{\mathbf{c},\rho}$ $:= \{\mathbf{x} \in \mathbb{R}^d \mid \|\mathbf{x} - \mathbf{c}\| \leq \rho\}$.

Definition 2.18 (Minimum Enclosing Ball (MEB) [1]) *Given a set of input points* $\mathcal{A} := \{\mathbf{a}_1, ..., \mathbf{a}_n\} \subset \mathbb{R}^d$ *we denote the unique minimum enclosing ball of* $\mathcal{A}$ *by* $MEB(\mathcal{A})$*, i.e.* $MEB(\mathcal{A}) = \mathcal{B}_{\mathbf{c}^*,\rho^*} = \{\mathbf{x} \in \mathbb{R}^d \mid \|\mathbf{x} - \mathbf{c}^*\| \leq \rho^*\}$ *where* $\mathbf{c}^* \in \mathbb{R}^d$ *is the optimal center and* $\rho^* \in \mathbb{R}$ *is the optimal radius.*

Lemma 2.19 ([3]) *Given a set of n input points* $\mathcal{A} := \{\mathbf{a}_1, ..., \mathbf{a}_n\} \subset \mathbb{R}^d$ *with* $n \geq d+1$*. Then there exists a set* $\mathcal{A}' \subset \mathcal{A}$ *with* $|\mathcal{A}'| = d+1$ *such that* $MEB(\mathcal{A}') = MEB(\mathcal{A})$*.*

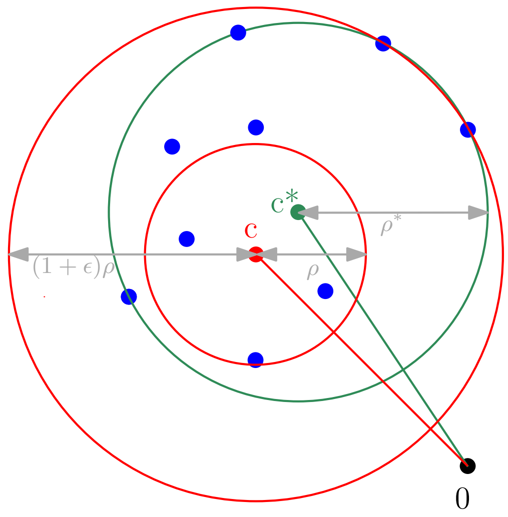
Proof ([3]) Consider the $MEB(\mathcal{A})$ and denote its center by $\mathbf{c}^*$ and its radius by ρ^*. For some $\mathbf{a}_i \in \mathcal{A}$ consider the set $\{\mathbf{x} \in \mathbb{R}^d \mid \|\mathbf{x} - \mathbf{a}_i\| \leq \rho^*\}$ which is the set of all centers which admit a ball of radius ρ^* that encloses $\mathbf{a}_i$. The intersection of all these sets $I = \bigcap_{\mathbf{a}_i \in \mathcal{A}} \{\mathbf{x} \in \mathbb{R}^d \mid \|\mathbf{x} - \mathbf{a}_i\| \leq \rho^*\}$ contains exactly one point: $\mathbf{c}^*$. It follows that the interior of I is empty, i.e. $int(I) = \bigcap_{\mathbf{a}_i \in \mathcal{A}} \{\mathbf{x} \in \mathbb{R}^d \mid \|\mathbf{x} - \mathbf{a}_i\| < \rho^*\} = \emptyset$. By Hellys Theorem ([2]) it follows that there exists $\mathcal{A}' \subset \mathcal{A}$ with $|\mathcal{A}'| = d+1$ and $\bigcap_{\mathbf{a}_i \in \mathcal{A}'} \{\mathbf{x} \in \mathbb{R}^d \mid \|\mathbf{x} - \mathbf{a}_i\| < \rho^*\} = \emptyset$. So there doesn't exist a ball of radius smaller than ρ^* that encloses the elements in $\mathcal{A}'$ and therefore $MEB(\mathcal{A})$ and $MEB(\mathcal{A}')$ have the same radius ρ^*. By the uniqueness of the minimum enclosing ball it follows that $MEB(\mathcal{A}') = MEB(\mathcal{A})$. □

This lemma pushes towards the concept of a support point, which I will introduce formally:

Definition 2.20 (Support point) *Given a set of input points* $\mathcal{A} := \{\mathbf{a}_1, ..., \mathbf{a}_n\} \subset \mathbb{R}^d$ *a support point of the* $MEB(\mathcal{A})$ *is a point on the boundary of* $MEB(\mathcal{A})$*.*

Lemma 2.21 (Hemisphere Lemma for MEB ([11]) *Given a set of n input points* $\mathcal{A} := \{\mathbf{a}_1, ..., \mathbf{a}_n\} \subset \mathbb{R}^d$ *with* $MEB(\mathcal{A}) = \mathcal{B}_{\mathbf{c}^*,\rho^*}$*. Then any closed halfspace that contains* $\mathbf{c}^*$ *also contains at least one support point* $\mathbf{a}_j \in \mathcal{A}$*.*

Definition 2.22 ($(1+\varepsilon)$-approximation [1]) *Given* $\varepsilon > 0$ *a ball* $\mathcal{B}_{\mathbf{c},\rho}$ *is said to be a* $(1+\varepsilon)$*-approximation to a* $MEB(\mathcal{A})$ *if* $\rho \leq \rho^*$ *and* $\mathcal{A} \subset \mathcal{B}_{\mathbf{c},(1+\varepsilon)\rho}$



2.4 Connection between the SVM and the MEB

In this section I will explain the connection between the SVM and the MEB. I will make use of this connection throughout this thesis.

2.4.1 Connection between the convex optimization problems

Proposition 2.23 *If we reduce the SVM-problem in Definition 2.4 and the MEB-problem in Definition 2.15 to the case where* $\|\mathbf{a}_i\| = 1 \quad \forall i \in \{1, ...n\}$ *then*

a) an optimal solution $\mathbf{x}^*$ *to the SVM-problem is also an optimal solution to the MEB-problem and conversely (and therefore the optimal classifier* $\mathbf{w}^*$ *and the center of the minimum enclosing ball* $\mathbf{c}^*$ *coincide)*

and

b) the support vectors associated with a solution $\mathbf{x}^*$ *of the SVM-problem correspond to the support points associated with the solution* $\mathbf{x}^*$ *to the MEB-problem.*

Proof a) The feasible region of our two convex optimization problems is the same by definition. By the additional restriction of the problems to the case where $\|a_i\| = 1 \quad \forall i \in \{1, ...n\}$ the objective function of the MEB-problem reduces to $\mathbf{x}^T A^T A\mathbf{x} - 1$. This objective function corresponds (up to a constant term) to the objective function of the SVM-problem, so the two convex optimization problems have the same form and therefore will have the same optimal solution(s). It follows by the definition of $\mathbf{w}^*$ and $\mathbf{c}^*$ that the optimal classifier $\mathbf{w}^*$ and the center of our minimum enclosing ball $\mathbf{c}^*$ coincide.

b) To prove this part of our proposition we will first take a closer look at the connection between $\mathbf{w}^*$ and the support vectors and $\mathbf{c}^*$ and the support points.

- By Property 2.13 of support vectors in a SVM-problem it holds that $\|\mathbf{a}_i - \mathbf{w}^*\|^2 = \|\mathbf{a}_i\|^2 - \|A\mathbf{x}^*\|^2$. In the case where $\|\mathbf{a}_i\| = 1 \quad \forall i \in \{1, ...n\}$ this reduces to $\|\mathbf{a}_i - \mathbf{w}^*\|^2 = 1 - \|A\mathbf{x}^*\|^2$.

- By Theorem 2.16 it holds that $\|\mathbf{a}_j - \mathbf{c}^*\|^2 = -\mathbf{x}^{*T} A^T A\mathbf{x}^* + \sum_{j=1}^n x_j^* \mathbf{a}_j^T \mathbf{a}_j$ for all support points $\mathbf{a}_j$ in a MEB-problem. In the case where $\|\mathbf{a}_i\| = 1 \quad \forall i \in \{1, ...n\}$ this reduces to $\|\mathbf{a}_j - \mathbf{c}^*\|^2 = -\mathbf{x}^{*T} A^T A\mathbf{x}^* + 1$.

Using the facts that $\mathbf{w}^*$ and $\mathbf{c}^*$ coincide and the conditions on the distances of support vectors and support point to them are the same in the SVM- and the MEB-problem it follows that the support vectors associated with the solution $\mathbf{x}^*$ of our SVM-problem correspond to the support points associated with the solution $\mathbf{x}^*$ of the MEB-problem. □

2.4.2 Connection between $(1-\delta)$-approximation in SVM-problem and $(1+\varepsilon)$-approximation in MEB-problem

As shown in Section 2.4.1 a connection between the SVM-problem and the MEB-problem can be made in the case where we have unit point length, i.e. if $\|\mathbf{a}_i\| = 1 \quad \forall i \in \{1, ..., n\}$. To make a connection between the $(1-\delta)$-approximation in the SVM-problem and the $(1+\varepsilon)$-approximation in the MEB-problem it is additionally needed that these approximations are strict, i.e:

- MEB: Choose ε s.t. for at least one $\mathbf{a}_j \in \mathcal{A} : \|\mathbf{a}_j - \mathbf{c}\| = (1+\varepsilon)\rho$.
- SVM: Choose δ s.t. for at least one $\mathbf{a}_j \in \mathcal{A} : \langle \mathbf{a}_j, \mathbf{w} \rangle = (1-\delta)\|\mathbf{w}\|$

Lemma 2.24 *Consider the convex optimization SVM-problem and the convex optimization MEB-problem reduced to the case that $\|\mathbf{a}_i\| = 1 \quad \forall i \in \{1, ..., n\}$. We have strict approximation in both problems, i.e. ε is s.t. for at least one $\mathbf{a}_j \in \mathcal{A}$: $\|\mathbf{a}_j - \mathbf{c}\| = (1+\varepsilon)\rho$ and δ is s.t. for at least one $\mathbf{a}_j \in \mathcal{A} : \langle \mathbf{a}_j, \mathbf{w} \rangle = (1-\delta)\|\mathbf{w}\|$. It holds that*

$$(1+\varepsilon)^2\rho^2 = 1 + (2\delta - 1)\|\mathbf{w}\|^2 \tag{2.1}$$

Proof

- $\leq$:
  Consider the point $\mathbf{a}_j$ which lies on the boundary of our test-ball $\mathcal{B}_{\mathbf{c},(1+\varepsilon)\rho}$. As always the test-vector $\mathbf{c}$ (center of our test-ball) and the candidate classifier $\mathbf{w}$ correspond to each other.

It holds that

$$\begin{aligned}
(1+\varepsilon)^2\rho^2 &= \|\mathbf{a}_j - \mathbf{c}\|^2 && (2.2)\\
&= \|\mathbf{a}_j - \mathbf{w}\|^2 && (2.3)\\
&\overset{(*)}{\leq} (\delta\|\mathbf{w}\|)^2 + (1-(1-\delta)^2\|\mathbf{w}\|^2) && (2.4)\\
&= \delta^2\|\mathbf{w}\|^2 + 1 - \|\mathbf{w}\|^2 + 2\delta\|\mathbf{w}\|^2 - \delta^2\|\mathbf{w}\|^2 && (2.5)\\
&= 1 + (2\delta - 1)\|\mathbf{w}\|^2 && (2.6)
\end{aligned}$$

where we used the Pythagorean Theorem and the fact that $\langle \mathbf{a}_j, \mathbf{w}\rangle \geq (1-\delta)\|\mathbf{w}\|$ in $(*)$.

- $\geq$:
  Consider the point $\mathbf{a}_i$ which defines the margin of separation $\sigma = (1-\delta)\|\mathbf{w}\|$ of a candidate classifier $\mathbf{w}$. As always the candidate classifier $\mathbf{w}$ and the test-vector $\mathbf{c}$ (center of our test-ball) correspond to each other. It holds that

$$\begin{aligned}
1 + (2\delta - 1)\|\mathbf{w}\|^2 &= \delta^2\|\mathbf{w}\|^2 + 1 - \|\mathbf{w}\|^2 + 2\delta\|\mathbf{w}\|^2 - \delta^2\|\mathbf{w}\|^2 && (2.7)\\
&= (\delta\|\mathbf{w}\|)^2 + (1-(1-\delta)^2\|\mathbf{w}\|^2) && (2.8)\\
&= \|\mathbf{a}_i - \mathbf{w}\|^2 && (2.9)\\
&= \|\mathbf{a}_i - \mathbf{c}\|^2 && (2.10)\\
&\overset{(*)}{\leq} (1+\varepsilon)^2\rho^2 && (2.11)
\end{aligned}$$

where we used the fact that $\mathbf{a}_i \in \mathcal{B}_{\mathbf{c},(1+\varepsilon)\rho} \quad \forall i \in \{1, ..., n\}$. □

Example 2.25 (Lemma 2.24 for the case that $\varepsilon = 0$ and $\delta = 0$) *When $\varepsilon = 0$ we have that $\rho = \rho^*$. Analogously, when $\delta = 0$ we have that $\mathbf{w} = \mathbf{w}^*$. If we consider Lemma 2.24 we find a fact that can also be proved by the Pythagorean Theorem: $\rho^{*2} = 1 - \|\mathbf{w}^*\|^2$.*

Chapter 3

Screening

In this Chapter I introduce the concept of screening and I give some insights into its effects and show how it is applied. Furthermore I will establish how I will assess the quality of a screening rule in further chapters of this thesis.

3.1 Idea behind screening

The idea behind screening (also called pruning) is to pre-process input data of an algorithm, for example input points $\mathbf{a}_i$, in order to find and eliminate elements of the input which do not contribute to the optimal solution. Hereby the problem size is reduced which hopefully results in a reduction of the total running time of the algorithm.

3.2 Effects of screening

In general the running time of algorithms that solve convex optimization problems such as the SVM-problem or the MEB-problem depends on the amount of input points n and the number of dimensions of the space (d). By screening we reduce the number of input points (n gets smaller) and therefore the running time decreases.

Example 3.1 *A very simple algorithm to compute a minimum enclosing ball is based on Lemma 2.19 (the minimum enclosing ball is determined by d+1 points of the input set). The idea of the algorithm is to initially take a subset S of size $d+1$ of our input set $\mathcal{A}$ and to find an initial minimum enclosing ball of S (MEB(S)). Then, a point that lies outside the boundary of the current minimum enclosing ball is added to the set S and a new minimum enclosing ball of the new set S is computed. After this step a point that doesn't lie on the boundary of the newly found minimum enclosing ball is eliminated and the procedure is restarted. This way the minimum enclosing ball grows iteratively until it includes all points of $\mathcal{A}$.*

```
1 Take $S \subset \mathcal{A}$, $|S| = d+1$
2 compute $MEB(S)$
3 while($\mathcal{A} \not\subset MEB(S)$)
4    $\mathbf{a}_i$:=farthest point of $\mathcal{A}$ from center $\mathbf{c}^*$ of $MEB(S)$
5    $S$:=$S \cup \{\mathbf{a}_i\}$
6    compute $MEB(S)$
7    find $\mathbf{a}_j \in S$ which isn't support point of $MEB(S)$
8    $S := S \setminus \{\mathbf{a}_j\}$
```

The running time of this simple algorithm is highly dependent on the amount of input points: In line 4, the search for the farthest point from the center of our candidate ball touches every point of the input set.

3.3 Screening procedure

Our usual algorithms for finding a solution to convex optimization problems such as the SVM- and the MEB-problem work by starting with a test-solution and working towards the optimal solution by iteration. The screening procedure is often implemented in the iterative part of the algorithm.

Example 3.2 *In our algorithm from Example 3.1 the screening would be a part of the while-loop:*

```
1 Take $S \subset \mathcal{A}$, $|S| = d+1$
2 compute $MEB(S)$
3 while($\mathcal{A} \not\subset MEB(S)$)
4    $\mathbf{a}_i$:=farthest point of $\mathcal{A}$ from center $\mathbf{c}^*$ of $MEB(S)$
5    screen($\mathcal{A}$)
6    $S$:=$S \cup \{\mathbf{a}_i\}$
7    compute $MEB(S)$
8    find $\mathbf{a}_j \in S$ which isn't support point of $MEB(S)$
9    $S := S \setminus \{\mathbf{a}_j\}$
```

The screening-step in line 5 removes points of the input set $\mathcal{A}$ based on a previously defined screening rule. In further iterations the search for the farthest point in line 4 does not require n computational steps as before but only n-(amount of screened points) computational steps.

3.4 Quality of screening rules

Depending on what we want to achieve by screening there are different ways of assessing the quality of a screening rule. In the further chapters I will assess screening rules based on one quality a screening rule could have: Effectiveness.

The main goal in applying screening rules is to reduce the amount of input points as much as possible to make the computation of the optimal solution faster. A screening rule should therefore have the property that it screens more points the closer the candidate solution gets to an optimal solution.

Definition 3.3 (Effective screening rule) *Let* $\mathbf{t}^*$ *be the optimal solution of a convex optimization problem and let* $(\mathbf{t}_1, ..., \mathbf{t}_k)$ *be a sequence of feasible solutions with* $\|\mathbf{t}_i - \mathbf{t}^*\| = \theta_i$ *and* $\theta_i = \|\mathbf{t}_i - \mathbf{t}^*\| \to 0$ *as* $i \to k$*. We call a screening rule effective if it has the property to screen all input points that correspond to zero-entries in the optimal solution* $\mathbf{t}^*$ *of the convex optimization problem as* $\theta_i \to 0$*.*

Similar requirements for the quality of a screening rule were made in connection with other convex optimization problems, for example for a screening rule for the logistic regression problem. In [10] it was emphasised that "The more accurate a solution is, the more inactive features can be detected [by the proposed screening rule]."

Chapter 4

Existing results for the minimum enclosing ball problem

In the past years screening rules for different convex optimization problems (such as minimum enclosing ball and the support vector machine) were proposed and their impact was tested. These tests usually assessed one of two (or both) aspects to make a statement about the impact the screening rule has on the computation:

- The speed-up of the running time
- The amount of input points which are excluded for further computations by the screening rule

For this thesis I had a closer look at two articles that proposed screening rules for the minimum enclosing ball problem: [1] and [7]. Since the screening rule proposed in [1] is tested (again) in [7] I will only consider the results of article [7].

4.1 Screening rules

In the two articles ([1] and [7]) a total of three screening rules for the minimum enclosing ball problem in the form of Definition 2.15 are proposed. In this thesis I will only give an insight into the screening rule of [1] and the first screening rule of [7]. This choice is based on the fact that the second screening rule of [7] follows similar ideas like the first screening rule of [7] but doesn't improve the results.

Both screening rules considered are based on the idea of starting with a $(1+\varepsilon)$-approximation of a minimum enclosing ball (Definition 2.22) with an associated center $\mathbf{c}$ and then screening points inside a ball that has its center in $\mathbf{c}$ and a radius such that it only includes points which are guaranteed not to be support points. For the screening rule proposed in [7] we have the additional requirement that the $(1+\varepsilon)$-approximation $\mathcal{B}_{\mathbf{c},\rho}$ is viable:

Definition 4.1 (Viable Ball ([7])) *A ball $\mathcal{B}_{\mathbf{c},\rho}$ is viable if it satisfies* $\rho^2 + \|\mathbf{c} - \mathbf{c}^*\|^2 \leq \rho^{*2}$.

Geometrically this means that at least one diameter of $B(\mathbf{c},\rho)$ is enclosed in the minimum enclosing ball ([7]).

The screening rules proposed in [1] and [7] have the following form:

Screening Rule 4.2 (AY [1]) *Remove any input point $\mathbf{a}_k \in \mathcal{A}$ with* $\|\mathbf{a}_k - \mathbf{c}\| < \rho(1 - \sqrt{2\varepsilon + \varepsilon^2})$.

Screening Rule 4.3 (KL1 ([7])) *Remove any input point $\mathbf{a}_k \in \mathcal{A}$ with* $\|\mathbf{a}_k - \mathbf{c}\| < \rho(\sqrt{1 + \varepsilon + \frac{\varepsilon^2}{2}} - \sqrt{\varepsilon + \frac{\varepsilon^2}{2}})$.

4.2 Experimental setting

In [7] the screening rules AY and KL1 were implemented into three different MEB-algorithms. The resulting algorithms were tested on input sets in 3 to 200 dimensions, each with $n = 10^6$ input points. The input points were generated using different types of distribution: standard normal distribution, uniform distribution inside a cube and uniform distribution inside a ball.

4.3 Results

The two screening rules were tested on their speed-up factor and the amount of work saved. The amount of work saved was found by dividing the number of point distance computations in the algorithm with the screening rule by the number of point distance computations in the non-screening version of the algorithm. The amount of work saved therefore depends on the amount of input points which are eliminated in each step: For each point which is eliminated one point distance computation less is needed in each iteration of the algorithm.

The impact of the screening depended on the dimension of the input set, the chosen distribution of the input set and the screening rule:

- **Dimension**

  There was a large difference in the results of the tests depending on the amount of dimensions in the input (the more dimensions the less impact by the screening rule): In 3 dimensions a factor of up to 140 was found for the speed-up and up to 99% of the amount of work was saved. In 200 dimensions in the worst case no speed-up could be achieved and 0% of the amount of work could be saved.

This interconnection was explained the following way: The volume of a ball with radius r in dimension d is proportional to r^d. Thus, the volume ratio between any pruning ball and the minimum enclosing ball will approach 0 as d approaches infinity.

- **Distribution**

  The effect of the screening was the largest on normally distributed data: For normally distributed data a factor of up to 140 was found for the speed-up and up to 99% of the amount of work was saved. For data that was uniformly distributed inside a ball in the worst case no speed-up could be achieved and 0% of the amount of work could be saved.

  This phenomenon was explained the following way: The points in normally distributed sets are concentrated near the center of the optimal minimum enclosing ball and are likely to be screened. The smallest effect of the screening was observed for input sets with uniform distribution inside a ball because these point sets are distributed in the entire volume of B^* which reduces the number of points covered by each screening ball.

- **Screening rule**

  KL1 was more successfull concerning the speed-up and the amount of work saved than AY.

  KL1 performed best on normally distributed data and in lower dimensions (3 to 10) with the following results: A factor of up to 140 was found for the speed-up (in dimension 3) and up to 99% of the amount of work was saved (in dimensions 5 and 10). AY also performed well on normally distributed data and in lower dimensions (3 to 10) with the following results: A factor of up to 100 was found for the speed-up (in dimension 3) and up to 99% of the amount of work was saved (in dimension 5).

Chapter 5

Screening rules for the SVM based on a screening rule for the MEB

5.1 Idea

In this chapter we try to use the idea of a screening rule for MEB-problems formulated by Ahipasaoglu and Yildirim in 'Identification and elimination of interior points for the minimum enclosing ball problem' ([1]) to find and prove a screening rule for the SVM-problem from Definition 2.4.

5.2 Proposition and Proof

Proposition 5.1 *Given an input set* $\mathcal{A} := \{\mathbf{a}_1, ..., \mathbf{a}_n\} \subset \mathbb{R}^d$ *let* $\mathbf{w} \in conv(\mathcal{A})$ *be a* $(1-\delta)$*-approximation of the solution* $\mathbf{w}^*$ *of the SVM-problem from Definition 2.4.*
Then:

a) $\|\mathbf{w} - \mathbf{w}^*\| \leq \sqrt{2\delta}\|\mathbf{w}\|$

b) each point $\mathbf{a}_i \in \mathcal{A}$ *which is a support point satisfies*

$$\|\mathbf{a}_i - \mathbf{w}\| \geq \sqrt{\|\mathbf{a}_i\|^2 - \|\mathbf{w}^*\|^2} - \sqrt{2\delta}\|\mathbf{w}\| \tag{5.1}$$

Proof First consider the simple case that $\mathbf{w} = \mathbf{w}^*$:

a) In the case that $\mathbf{w} = \mathbf{w}^*$ the δ of the $(1-\delta)$-approximation is 0. So both sides of the inequality are zero and the first part of the proposition holds with equality: $0 = \|\mathbf{w} - \mathbf{w}^*\| = \sqrt{2\delta}\|\mathbf{w}\| = 0$

b) To prove the second part of the proposition we use the fact that by the Pythagorean Theorem it holds that

$$\|\mathbf{a}_j - \mathbf{w}^*\| = \sqrt{\|\mathbf{a}_j\|^2 - \|\mathbf{w}^*\|^2}$$

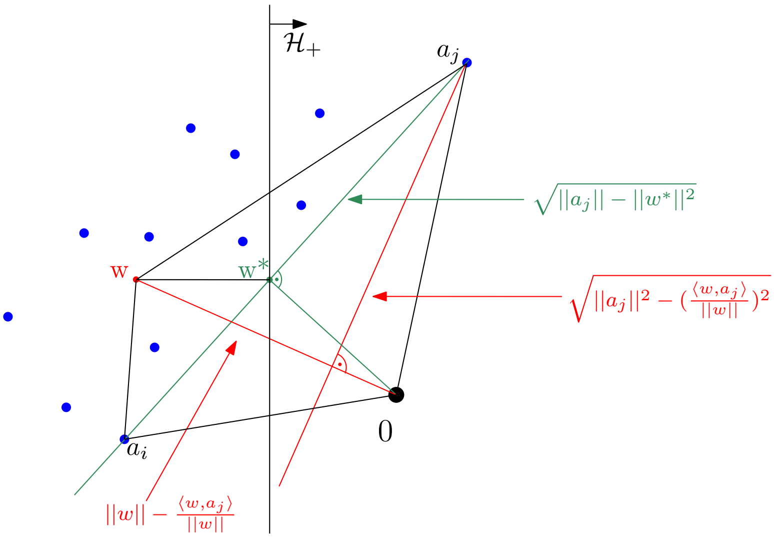


Figure 5.1: Geometric display of proof of proposition 5.1

By again using the fact that $\delta = 0$ we find that

$$\|\mathbf{a}_i - \mathbf{w}\| = \|\mathbf{a}_i - \mathbf{w}^*\| = \sqrt{\|\mathbf{a}_i\|^2 - \|\mathbf{w}^*\|^2} = \sqrt{\|\mathbf{a}_i\|^2 - \|\mathbf{w}^*\|^2} - \sqrt{2\delta}\|\mathbf{w}\| \tag{5.2}$$

and therefore the second part of the proposition also holds with equality.

Now consider the more difficult case that $\mathbf{w} \neq \mathbf{w}^*$ (consider Figure 5.1).

a) Consider the hyperplane $\mathcal{H}$ passing through $\mathbf{w}^*$ perpendicular to $(\mathbf{w}^* - \mathbf{w})$.
Let $\mathcal{H}_+$ denote the closed halfspace bounded by $\mathcal{H}$ and not containing $\mathbf{w}$. By the Hemisphere Lemma for SVM (Lemma 2.14) there exists a support vector $\mathbf{a}_j \in \mathcal{H}_+$ such that

$$\|\mathbf{a}_j - \mathbf{w}^*\| = \sqrt{\|\mathbf{a}_j\|^2 - \|\mathbf{w}^*\|^2}$$

By the law of cosines and the fact, that we have an obtuse angle at $\mathbf{w}^*$ it holds that

$$\|\mathbf{w} - \mathbf{a}_j\|^2 \geq \|\mathbf{w} - \mathbf{w}^*\|^2 + \|\mathbf{w}^* - \mathbf{a}_j\|^2$$

We can also find that:

$$\|\mathbf{w} - \mathbf{a}_j\|^2 = \|\mathbf{w}\|^2 - 2\langle \mathbf{w}, \mathbf{a}_j \rangle + \|\mathbf{a}_j\|^2 \tag{5.3}$$

$$\leq \|\mathbf{w}\|^2 - 2(1-\delta)\|\mathbf{w}\|^2 + \|\mathbf{a}_j\|^2 \tag{5.4}$$

$$= (2\delta - 1)\|\mathbf{w}\|^2 + \|\mathbf{a}_j\|^2 \tag{5.5}$$

Using these two inequalities we can establish the first part of the proposition:

$$\begin{aligned}
\|\mathbf{w}-\mathbf{w}^*\|^2 &\leq \|\mathbf{w}-\mathbf{a}_j\|^2 - \|\mathbf{w}^*-\mathbf{a}_j\|^2 &(5.6)\\
&\leq (2\delta-1)\|\mathbf{w}\|^2 + \|\mathbf{a}_j\|^2 - (\|\mathbf{a}_j\|^2 - \|\mathbf{w}^*\|^2) &(5.7)\\
&= (2\delta-1)\|\mathbf{w}\|^2 + \|\mathbf{w}^*\|^2 &(5.8)\\
&\leq (2\delta-1)\|\mathbf{w}\|^2 + \|\mathbf{w}\|^2 &(5.9)\\
&= 2\delta\|\mathbf{w}\|^2 &(5.10)
\end{aligned}$$

b) Let $\mathbf{a}_i$ be any support vector.
We know that

$$\|\mathbf{a}_i-\mathbf{w}^*\| = \sqrt{\|\mathbf{a}_i\|^2 - \|\mathbf{w}^*\|^2} \tag{5.11}$$

By the triangle-inequality we know

$$\|\mathbf{a}_i-\mathbf{w}^*\| \leq \|\mathbf{a}_i-\mathbf{w}\| + \|\mathbf{w}-\mathbf{w}^*\|$$

It follows:

$$\begin{aligned}
\|\mathbf{a}_i-\mathbf{w}\| &\geq \|\mathbf{a}_i-\mathbf{w}^*\| - \|\mathbf{w}-\mathbf{w}^*\| &(5.12)\\
&\geq \sqrt{\|\mathbf{a}_i\|^2 - \|\mathbf{w}^*\|^2} - \sqrt{2\delta}\|\mathbf{w}\| &(5.13)
\end{aligned}$$

□

5.3 Screening rules

We can derive three screening rules from Proposition 5.1.

5.3.1 General screening rule

The following screening rule follows immediately from the proposition:

Screening Rule 5.2 *Remove any input point* $\mathbf{a}_k \in \mathcal{A}$ *with*

$\|\mathbf{a}_k-\mathbf{w}\| < \sqrt{\|\mathbf{a}_k\|^2 - \|\mathbf{w}^*\|^2} - \sqrt{2\delta}\|\mathbf{w}\|.$

5.3.2 Adjusted general screening rule

For the case that for an input point $\mathbf{a}_i$ it holds that $\|\mathbf{a}_i\| > \|\mathbf{w}\|$ the general screening rule can be simplified to the following adjusted screening rule:

Screening Rule 5.3 *Remove any input point* $\mathbf{a}_k \in \mathcal{A}$ *with* $\|\mathbf{a}_k\| > \|\mathbf{w}\|$ *and*

$\|\mathbf{a}_k-\mathbf{w}\| < \sqrt{\|\mathbf{a}_k\|^2 - \|\mathbf{w}\|^2} - \sqrt{2\delta}\|\mathbf{w}\|.$

We can not simplify the general screening rule for every point $\mathbf{a}_i \in \mathcal{A}$ using the same idea because in every case where $\|\mathbf{a}_i\| < \|\mathbf{w}\|$ we would get a negative value under the first root-function.

5.3.3 Screening rule for the case with unit point-length

In the case where we reduce our SVM-problem to input points $\mathbf{a}_i$ with $\|\mathbf{a}_i\| = 1 \quad \forall i \in \{1, ..., n\}$ our results of Section 5.2 can be simplified:

$$\|\mathbf{a}_i - \mathbf{w}\| \geq \sqrt{\|\mathbf{a}_i\|^2 - \|\mathbf{w}^*\|^2} - \sqrt{2\delta}\|\mathbf{w}\| \tag{5.14}$$

$$\stackrel{1.}{=} \sqrt{1 - \|\mathbf{w}^*\|^2} - \sqrt{2\delta}\|\mathbf{w}\| \tag{5.15}$$

$$\stackrel{2.}{\geq} \sqrt{1 - \|\mathbf{w}\|^2} - \sqrt{2\delta}\|\mathbf{w}\| \tag{5.16}$$

Where we used:

1. $\|\mathbf{a}_i\| = 1$
2. $\|\mathbf{w}\| \geq \|\mathbf{w}^*\|$, $\|\mathbf{a}_i\| = 1$ and $\mathbf{w} \in conv(\mathcal{A}) \Rightarrow \|\mathbf{w}\| \leq 1$

The following screening rule results:

Screening Rule 5.4 *Remove any input point $\mathbf{a}_k \in \mathcal{A}$ with*

$\|\mathbf{a}_k - \mathbf{w}\| < \sqrt{1 - \|\mathbf{w}\|^2} - \sqrt{2\delta}\|\mathbf{w}\|$

5.4 Connection between the screening rule for the MEB-problem and the screening rule for the SVM-problem

The screening rule we found for the SVM-problem in Section 5.2 is related to the screening rule for the MEB-problem formulated by Ahipasaoglu and Yildirim in 'Identification and elimination of interior points for the minimum enclosing ball problem' ([1]) in the case where we have unit point length. To show this we do not consider the two final screening rules that were found by making some approximations but have a look at the results that were found a few steps earlier in the proofs.

Proposition 5.5 *Given a set of n input points $\mathcal{A} := \{\mathbf{a}_1, ..., \mathbf{a}_n\} \subset \mathbb{R}^d$ with $\|\mathbf{a}_i\| = 1 \quad \forall i = \{1, ...n\}$. The screening rule we defined for the one-class SVM-problem in Section 5.3.1 based on our proposition in Section 5.2 is related to the screening rule for the MEB-problem formulated by Ahipasaoglu and Yildirim [1] up to approximations.*

Proof For the SVM problem we found (for an arbitrary support point $\mathbf{a}_i$ and a support point $\mathbf{a}_j \in \mathcal{H}_+$, $\mathbf{a}_i$ can be the same as $\mathbf{a}_j$):

$$\|\mathbf{a}_i - \mathbf{w}\| \geq \|\mathbf{a}_i - \mathbf{w}^*\| - \|\mathbf{w} - \mathbf{w}^*\| \tag{5.17}$$
$$= \sqrt{1 - \|\mathbf{w}^*\|^2} - \|\mathbf{w} - \mathbf{w}^*\| \tag{5.18}$$
$$\geq \sqrt{1 - \|\mathbf{w}^*\|^2} - \sqrt{\|\mathbf{w} - \mathbf{a}_j\|^2 - \|\mathbf{w}^* - \mathbf{a}_j\|^2} \tag{5.19}$$
$$= \sqrt{1 - \|\mathbf{w}^*\|^2} - \sqrt{((\delta\|\mathbf{w}\|)^2 + (1 - (1-\delta)^2\|\mathbf{w}\|^2)) - (1 - \|\mathbf{w}^*\|^2)} \tag{5.20}$$
$$= \sqrt{1 - \|\mathbf{w}^*\|^2} - \sqrt{(2-\delta)\|\mathbf{w}\|^2 + \|\mathbf{w}^*\|^2} \tag{5.21}$$

For the MEB problem it was found that (for an arbitrary support point $\mathbf{a}_i$ and a support point $\mathbf{a}_j \in \mathcal{H}_+$, $\mathbf{a}_i$ can be the same as $\mathbf{a}_j$) ([1]) :

$$\|\mathbf{a}_i - \mathbf{c}\| \geq \|\mathbf{a}_i - \mathbf{c}^*\| - \|\mathbf{c} - \mathbf{c}^*\| \tag{5.22}$$
$$= \rho^* - \|\mathbf{c} - \mathbf{c}^*\| \tag{5.23}$$
$$\geq \rho^* - \sqrt{\|\mathbf{c} - \mathbf{a}_j\|^2 - \|\mathbf{c}^* - \mathbf{a}_j\|^2} \tag{5.24}$$
$$= \rho^* - \sqrt{(1+\varepsilon)^2\rho^2 - \rho^{*2}} \tag{5.25}$$

These two approximations correspond to each other:

$$\rho^* - \sqrt{(1+\varepsilon)^2\rho^2 - \rho^{*2}} \overset{(*)}{=} \sqrt{1 - \|\mathbf{w}^*\|^2} - \sqrt{(1 + (2\delta - 1)\|\mathbf{w}\|^2 - (1 - \|\mathbf{w}^*\|^2)} \tag{5.26}$$
$$= \sqrt{1 - \|\mathbf{w}^*\|^2} - \sqrt{(2\delta - 1)\|\mathbf{w}\|^2 + \|\mathbf{w}^*\|^2)} \tag{5.27}$$

where we used $\rho^* = \sqrt{1 - \|\mathbf{w}^*\|^2}$ (by the Pythagorean Theorem) and Lemma 2.24 in $(*)$.

So the two screening rules correspond to each other up to this point. By further estimations we derive the screening rule of Ahipasaoglu and Yildirim for the MEB-problem and the screening rule proposed in Section 5.3 for the SVM-problem. □

5.5 Quality Assessment

In this section I will assess the quality of the screening rules proposed in chapter 5.3 by the effectiveness-criterion defined in chapter 3.4.

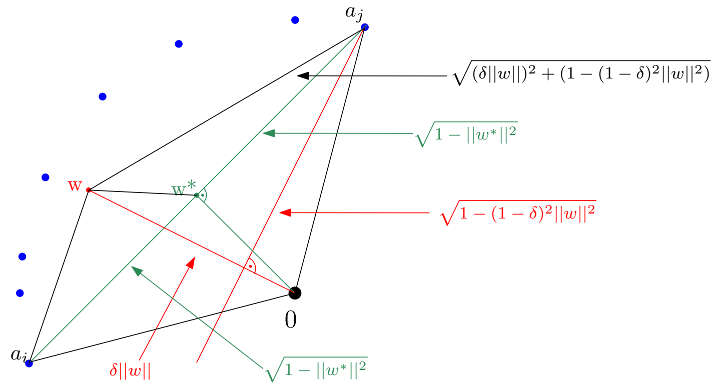


Figure 5.2: Support vector machine problem for the case that $\|\mathbf{a}_i\| = 1 \quad \forall i = \{1, ...n\}$

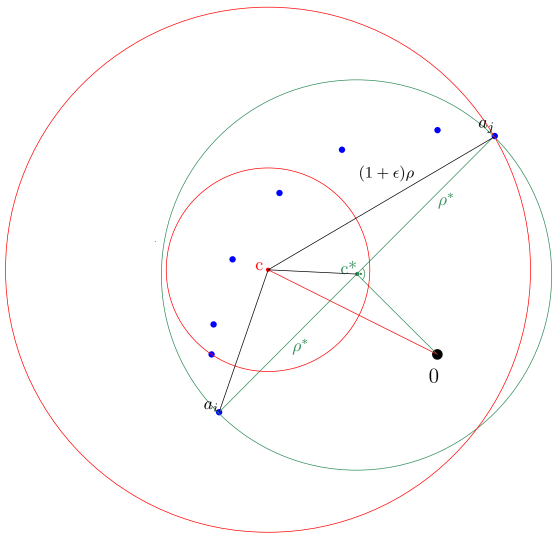


Figure 5.3: Minimum enclosing ball problem for the case that $\|\mathbf{a}_i\| = 1 \quad \forall i = \{1, ...n\}$

5.5.1 Quality of general screening rule

The screening rule proposed in Section 5.3.1 and proved in section 5.2 is applicable to all support vector machine problems and all points of the input set. Since the screening rule depends on already knowing the optimal solution $\mathbf{w}^*$ to the SVM-problem the application in practise is limited.

Our screening rule is based on two factors:

- The length of the point $\mathbf{a}_k$ in comparison to the lenght of the optimal solution $\mathbf{w}^*$. The bigger the difference, the more probable the point gets screened.
- The quality of the approximation $\mathbf{w}$. The better the approximation the more probable the point $\mathbf{a}_k$ is screened.

Therefore our screening rule has the property to screen more points the closer the approximation gets to the optimal solution. Still, a point $\mathbf{a}_k$ which is close to the optimal solution might not be screened if $\|\mathbf{a}_k\|$ is close to $\|\mathbf{w}^*\|$, even if $\mathbf{a}_k$ is not a support vector. The screening rule is therefore not effective in the sense of Definition 3.3 but it still has some kind of effectiveness on certain types of points.

5.5.2 Quality of the adjusted general screening rule

The screening rule proposed in Section 5.3.2 is only applicable to input points $\mathbf{a}_i$ with $\|\mathbf{a}_i\| > \|\mathbf{w}\|$. This restriction excludes many input points from the screening procedure when the candidate classifier is not close to the optimal solution. As the candidate classifier approaches the optimal solution more points can be checked by the screening rule. All in all, the screening rule is effective:

Lemma 5.6 *The adjusted general screening rule as proposed in Section 5.3.2 is effective in the sense of Definition 3.3.*

Proof Let $(\mathbf{w}_1, ..., \mathbf{w}_k)$ be a sequence of feasible solutions to the SVM-problem with $\|\mathbf{w}_i - \mathbf{w}^*\| = \theta_i \to 0$ as $i \to k$. Each $\mathbf{w}_i$ is a $(1 - \delta_i)$-approximation to the optimal solution $\mathbf{w}^*$ and since $\theta_i \to 0$ we have that $\delta_i \to 0$ as $i \to k$.

Consider the adjusted general screening rule as described in Section 5.3.2:

$$\|\mathbf{a}_k - \mathbf{w}_i\| < \sqrt{\|\mathbf{a}_k\|^2 - \|\mathbf{w}_i\|^2} - \sqrt{2\delta_i}\|\mathbf{w}_i\| \tag{5.28}$$

$$\stackrel{i \to k}{\to} \sqrt{\|\mathbf{a}_k\|^2 - \|\mathbf{w}^*\|^2} \tag{5.29}$$

This condition holds for all points which are not support vectors (compare Property 2.13) so all non support vectors will be screened as $\mathbf{w}^*$ is approached and therefore the screening rule is effective in the sense of Definition 3.3. □

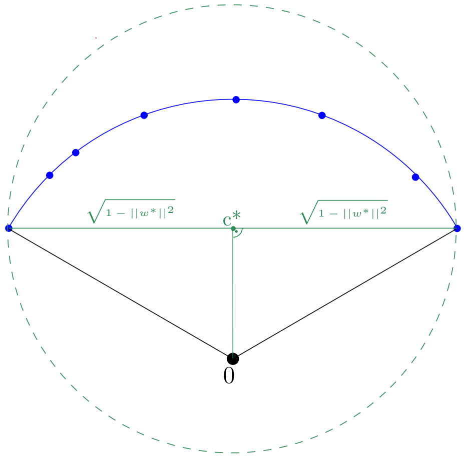
5.5.3 Quality of the screening rule for the case with unit point length

The screening rule in the case that the points $\mathbf{a}_i$ have the property that $\|\mathbf{a}_i\| = 1 \quad \forall i \in \{1, ..., n\}$ as proposed in Section 5.3.3 is only applicable for this certain type of the SVM-problem but to all points $\mathbf{a}_i$ of the input set.

The screening rule is effective in the sense of Definition 3.3 since it has the property to screen more points the closer it gets to the optimal solution $\mathbf{w}^*$. This follows directly from the following Lemma:

Lemma 5.7 *For the SVM-problem with unit point length ($\|\mathbf{a}_i\| = 1 \quad \forall i = \{1, ..., n\}$) the screening rule proposed in Section 5.3.3 will eliminate all points which are not*

Figure 5.4: Screening ball for the case that $\mathbf{w} = \mathbf{w}^*$



support vectors as the candidate classifier $\mathbf{w}$ *approaches the optimal solution* $\mathbf{w}^*$. *The screening rule is therefore effective in the sense of Definition 3.3.*

Proof Let $(\mathbf{w}_1, ..., \mathbf{w}_k)$ be a sequence of feasible solutions to the SVM-problem with $\|\mathbf{w}_i - \mathbf{w}^*\| = \theta_i \to 0$ as $i \to k$. Each $\mathbf{w}_i$ is a $(1 - \delta_i)$-approximation to the optimal solution and since $\theta_i \to 0$ we have that $\delta_i \to 0$ as $i \to k$.

Consider the screening rule for the case where we have unit point length as described in Section 5.3.3:

$$\|\mathbf{a}_k - \mathbf{w}_i\| < \sqrt{1 - \|\mathbf{w}_i\|^2} - \sqrt{2\delta}\|\mathbf{w}_i\| \tag{5.30}$$

$$\overset{i \to k}{\to} \sqrt{1 - \|\mathbf{w}^*\|^2} \tag{5.31}$$

Since $\sqrt{1 - \|\mathbf{w}^*\|^2} \geq \sqrt{1 - \|\mathbf{w}_i\|^2} \quad \forall i \in \{1, ..., k\}$ the radius of our screening ball gets bigger the closer we get to $\mathbf{w}^*$. In the case where we reach $\mathbf{w}^*$ all points $\mathbf{a}_i$ which are not support points are screened, as shown in Figure 5.4. □

Chapter 6

New screening rule for the SVM-problem

In this chapter I present the main result of this thesis: A new screening rule for the support vector machine. This screening rule was suggested by Martin Jaggi and I based the proof of the rule on ideas of Bernd Gärtner.

6.1 Idea

Consider the convex optimization SVM-problem from Definition 2.4. Let $\mathbf{w}$ be a $(1-\delta)$-approximation to the optimal solution $\mathbf{w}^*$.

The basic idea of the screening rule is to screen points $\mathbf{a}_i$ which lie in a cone with apex $\mathbf{w}$ and opening angle $2(\frac{\pi}{2}-\alpha)$, where α is the angle we find by constructing a right-angled triangle, which has $\mathbf{w}$ as hypotenuse and one side with lenght $(1-\delta)\|\mathbf{w}\|$ (see figure 6.1).

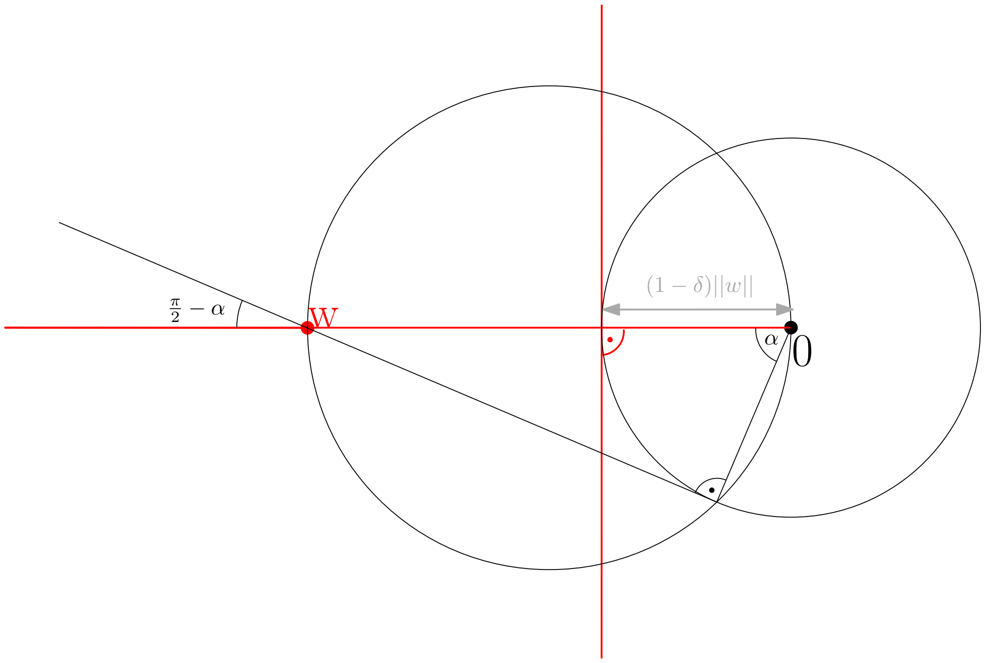
6.2 Proposition and Proof

Proposition 6.1 *Given an input set $\mathcal{A} := \{\mathbf{a}_1, ..., \mathbf{a}_n\} \subset \mathbb{R}^d$ let $\mathbf{w} \in conv(\mathcal{A})$ be a $(1-\delta)$-approximation of the solution $\mathbf{w}^*$ of the SVM-problem with $0 \leq \delta \leq 1$. Then each point $\mathbf{a}_i \in \mathcal{A}$ which is a support vector satisfies*

$$\gamma := \sphericalangle(\mathbf{a}_i - \mathbf{w}, \mathbf{w}) \geq \frac{\pi}{2} - \alpha$$

where α is such that $cos(\alpha) = 1 - \delta$.

Figure 6.1: Geometric idea of the screening rule



Proof In a first step of our proof we show that $\sphericalangle(\mathbf{w}, \mathbf{w}^*) \leq \alpha$:
Let $\mathbf{a}_i$ be a support vector.

$$\cos(\sphericalangle(\mathbf{w}, \mathbf{w}^*)) = \frac{\langle \mathbf{w}, \mathbf{w}^* \rangle}{\|\mathbf{w}\| \|\mathbf{w}^*\|} \tag{6.1}$$

$$\overset{1.}{\geq} \frac{\langle \mathbf{a}_i, \mathbf{w}^* \rangle}{\|\mathbf{w}\| \|\mathbf{w}^*\|} \tag{6.2}$$

$$\overset{2.}{=} \frac{\|\mathbf{w}^*\|^2}{\|\mathbf{w}\| \|\mathbf{w}^*\|} \tag{6.3}$$

$$= \frac{\|\mathbf{w}^*\|}{\|\mathbf{w}\|} \tag{6.4}$$

$$\overset{3.}{\geq} \frac{(1-\delta)\|\mathbf{w}\|}{\|\mathbf{w}\|} \tag{6.5}$$

$$= 1 - \delta \tag{6.6}$$

$$= cos(\alpha) \tag{6.7}$$

where we used

1. $\mathbf{w} \in conv(\mathcal{A})$, $\mathbf{a}_i$ is support vector
2. $\langle \mathbf{a}_i, \mathbf{w}^* \rangle = \|\mathbf{w}^*\|^2 \quad \forall \mathbf{a}_i$ support vector

3. $\mathbf{w}$ is a $(1-\delta)$ approximation of the solution $\mathbf{w}^*$, i.e. $\|\mathbf{w}^*\| \geq (1-\delta)\|\mathbf{w}\|$ (Property 2.11)

It follows that $\sphericalangle(\mathbf{w}, \mathbf{w}^*) \leq \alpha$.

Using this we can finish our proof:

Consider $\lambda\mathbf{w}$ where $\lambda = \frac{\|\mathbf{w}^*\|^2}{\langle \mathbf{w}, \mathbf{w}^* \rangle}$. Notice that $\lambda \leq 1$.

By simple geometric considerations (consider Figure 6.2) we find that

$$\gamma = \sphericalangle(\mathbf{a}_i - \mathbf{w}, \mathbf{w}) \tag{6.8}$$
$$\geq \sphericalangle(\mathbf{a}_i - \lambda\mathbf{w}, \mathbf{w}) \tag{6.9}$$
$$\overset{1.}{\geq} \pi - \sphericalangle(\mathbf{a}_i - \lambda\mathbf{w}, \mathbf{w}) \tag{6.10}$$
$$\overset{2.}{=} \pi - (\frac{\pi}{2} + \sphericalangle(\mathbf{w}, \mathbf{w}^*)) \tag{6.11}$$
$$\overset{3.}{\geq} \frac{\pi}{2} - \alpha \tag{6.12}$$

where we used

1. Use the fact that $(\mathbf{a}_i - \lambda\mathbf{w}, \mathbf{w}) \geq \frac{\pi}{2}$
2. The sum of angles in a triangle is π. Using this we have that $\sphericalangle(\mathbf{a}_i - \lambda\mathbf{w}, \mathbf{w}) = \frac{\pi}{2} + \sphericalangle(\mathbf{w}, \mathbf{w}^*)$
3. $\sphericalangle(\mathbf{w}, \mathbf{w}^*) \leq \alpha$ □

6.3 Screening rule

From the proposition we can derive the following screening rule:

Screening Rule 6.2 *Remove any input point $\mathbf{a}_k \in \mathcal{A}$ with*

$\sphericalangle(\mathbf{a}_k - \mathbf{w}, \mathbf{w}) < \frac{\pi}{2} - \alpha$.

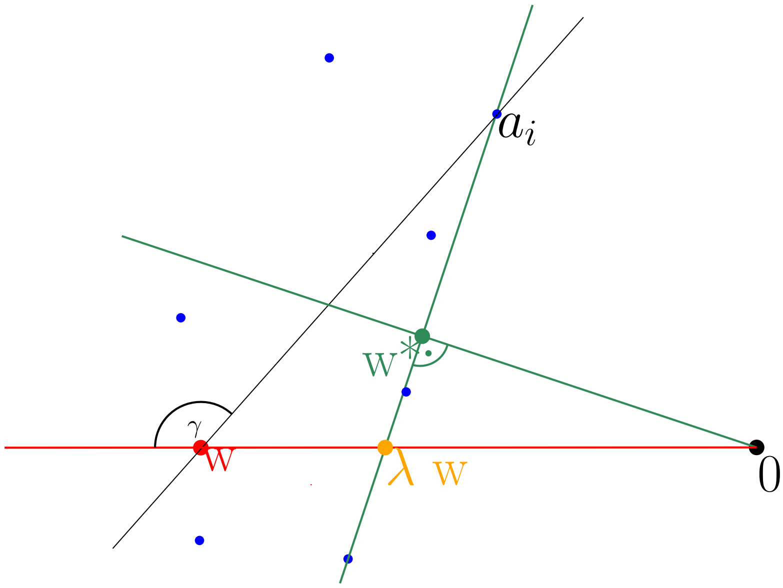
For the application in algorithms we reformulate this screening rule:

Screening Rule 6.3 *Remove any input point $\mathbf{a}_k \in \mathcal{A}$ with*

$$\frac{\langle \mathbf{w}, \mathbf{a}_k - \mathbf{w} \rangle}{\|\mathbf{w}\| \|\mathbf{a}_k - \mathbf{w}\|} > \sqrt{2\delta - \delta^2} \tag{6.13}$$

Proposition 6.4 *Screening Rule 6.2 and 6.3 are equivalent.*

Figure 6.2:



Proof

$$\sphericalangle(\mathbf{a}_k - \mathbf{w}, \mathbf{w}) < \frac{\pi}{2} - \alpha \Leftrightarrow cos(\sphericalangle(\mathbf{a}_k - \mathbf{w}, \mathbf{w})) > cos(\frac{\pi}{2} - \alpha) \quad (6.14)$$

$$\Leftrightarrow \frac{\langle \mathbf{w}, \mathbf{a}_k - \mathbf{w} \rangle}{\|\mathbf{w}\|\|\mathbf{a}_k - \mathbf{w}\|} > sin(\alpha) \quad (6.15)$$

$$\Leftrightarrow \frac{\langle \mathbf{w}, \mathbf{a}_k - \mathbf{w} \rangle}{\|\mathbf{w}\|\|\mathbf{a}_k - \mathbf{w}\|} > \sqrt{1 - cos(\alpha)^2} \quad (6.16)$$

$$\Leftrightarrow \frac{\langle \mathbf{w}, \mathbf{a}_k - \mathbf{w} \rangle}{\|\mathbf{w}\|\|\mathbf{a}_k - \mathbf{w}\|} > \sqrt{1 - (1 - \delta)^2} \quad (6.17)$$

$$\Leftrightarrow \frac{\langle \mathbf{w}, \mathbf{a}_k - \mathbf{w} \rangle}{\|\mathbf{w}\|\|\mathbf{a}_k - \mathbf{w}\|} > \sqrt{1 - (1 - 2\delta + \delta^2)} \quad (6.18)$$

$$\Leftrightarrow \frac{\langle \mathbf{w}, \mathbf{a}_k - \mathbf{w} \rangle}{\|\mathbf{w}\|\|\mathbf{a}_k - \mathbf{w}\|} > \sqrt{2\delta - \delta^2} \quad (6.19)$$

□

6.4 Quality assessment

In this subchapter I will assess the quality of the screening rule proposed in chapter 6.3 by the effectiveness-criterion defined in chapter 3.4

The screening rule is based on knowing a candidate classifier $\mathbf{w}$ which is already quite good ($\mathbf{w}$ is a $(1-\delta)$-approximation with $0 \leq \delta \leq 1$). It is therefore only applicable if this is the case.

To evaluate the effectiveness of our screening rule we consider the following lemma:

Lemma 6.5 *Let $\mathbf{w}$ be a $(1-\delta)$-approximation with $0 \leq \delta \leq 1$.*

Then the screening rule proposed in 6.3 will eliminate all points which are not support vectors as $\mathbf{w}$ approaches the optimal solution $\mathbf{w}^*$.

Proof Let $(\mathbf{w}_1, ..., \mathbf{w}_k)$ be a sequence of feasible solutions to the SVM-problem with $\|\mathbf{w}_i - \mathbf{w}^*\| = \theta_i \to 0$ as $i \to k$. Each $\mathbf{w}_i$ is a $(1-\delta_i)$-approximation to the optimal solution and since $\theta_i \to 0$ we have that $\delta_i \to 0$ as $i \to k$.

Let $i \to k$:

$$\mathbf{w}_i \to \mathbf{w}^* \Rightarrow \delta_i \to 0 \tag{6.20}$$

$$\Rightarrow (1-\delta_i) \to 1 \tag{6.21}$$

$$\Rightarrow cos(\alpha) \to 1 \tag{6.22}$$

$$\Rightarrow \alpha \to 0 \tag{6.23}$$

The proposed screening rule will screen all points inside a cone with vertex $\mathbf{w}_i(\to \mathbf{w}^*)$ and an opening angle $2(\frac{\pi}{2} - \alpha) \to \pi$, so it will eliminate all points which are not support vectors. □

The screening rule is therefore effective in the sense of Definition 3.3.

Chapter 7

Outlook

There is a number of interesting directions for further studies based on this thesis.

One such direction would be to implement the screening rule found in chapter 6.3 and to test its impact. Also, the screening rule for the case with point-length 1 proposed in Chapter 5.3.3 could be tested computationally.

The screening rules found in this thesis are mainly based on geometric ideas. In further works on the topic it would be interesting to make a connection to the solution of the convex optimization problems using KKT-conditions.

Another possible direction would be to translate the screening rule I found in Chapter 6.3 for the one-class SVM-problem to the case where we have a two-class SVM-problem or to the minimum enclosing ball problem with unit point lenght, which should be easily possible since the problems translate to each other.

A similar but probably more laborious direction would be to translate the screening rule I found for the SVM-problem to a screening rule for other, similar convex optimization problems. For example one could make use of the equivalence between the lasso and the support vector machine found in [4].

Bibliography

[1] S. Damla Ahipasaoglu and E. Alper Yildirim. Identification and elimination of interior points from the minimum enclosing ball problem. *SIAM Journal on Optimization*, 2008.

[2] Vasek Chvatal. *Linear Programming*. W.H. Freeman, 1983.

[3] Kaspar Fischer and Bernd Gärtner. The smallest enclosing ball of balls: Combinatorial structure and algorithms. *International Journal of Computational Geometry and Applications (IJCGA)*, 2004.

[4] Martin Jaggi. *An equivalence between the lasso and support vector machines. In Regularization, Optimization, Kernel Methods and Support Vector Machines*. Chapman and Hall/CRC, 2014.

[5] Thorsten Joachims. *Text categorization with support vector machines: Learning with many relevant features*. Springer, 1998.

[6] Hyunsoo Kim and Haesun Park. Prediction of protein relative solvent accessibility with support vector machines and long-range interaction 3d local descriptor. *Proteins: Structure, Function, and Bioinformatics*, 2004.

[7] Linus Källberg and Thomas Larsson. Improved pruning of large data sets for the minimum enclosing ball problem. *Graphical Models*, 2014.

[8] Jiri Matousek and Bernd Gärtner. *Understanding and using linear programming (Universitext)*. Springer-Verlag New York, 2006.

[9] Edgar Osuna, Robert Freund, and Federico Girosi. Support vector machines: Training and applications. *A.I. Memo*, 1602, 1997.

[10] Jie Wang, Jiayu Zhou, Jun Liu, Peter Woka, and Jiepeng Ye. A safe screening rule for sparse logistic regression. *ArXiv e-prints*, 2013.

[11] E. Alper Yildirim. Two algorithms for the minimum enclosing ball problem. *SIAM Journal on Optimization*, 2008.