Theoretical Analysis of Support Vector Machine with Cardinality Constraint

Evelin Heringer Manoel Krulikovski[1] & Mael Sachine[2] & Ademir Alves Ribeiro[3]

Supported by CAPES.

evelin.hmk@gmail.com[1]; mael@ufpr.br[2]; ademir.ribeiro@ufpr.br[3]

Abstract

The general objective of this work was to perform a theoretical study about SVM, which includes reporting justifications for the use of such technique and showing its geometric interpretation and analytical perspective. In order to apply the technique to classification problems, we seek to base its use mathematically, since it involves a quadratic, convex and constrained programming problem [1, 2]. For the analysis of the technique, we use the theory of Lagrangian duality, to facilitate the calculations and the analysis of the solutions. We worked with the *Kernel* function to solve the problem when it is not possible to find a decision function in the input space[2]. We propose an Algorithm, which search a dual solution as sparse as possible[1]. Based on this algorithm, we consider adding a cardinality constraint to the problem. So we started working with *Mathematical Programs with Cardinality Constraints*, which are very difficult mathematical programs. We show that under the hypothesis that the cardinality constraint is active the Mangasarian-Fromovitz condition is violated. Here we introduce a mixed integer formulation that standard relaxation still has the same solutions (global minimal) with the problem limited by cardinality[3]. We then propose a new definition of W-stationarity for this class of problems with base on [4]. We show the equivalence of this definition with M-stationarity based on [3], also that at a feasible point $(\bar{x}, \bar{y})$ of the relaxed problem above mentioned, this conditions are exactly the KKT conditions of the *Tightened Nonlinear Program* TNLP$(\bar{x}, \bar{y})$ of this problem.

Introduction

In this work, we carried out a theoretical study of a supervised machine learning technique: the Support Vector Machine (SVM) [2]. This technique focuses attention on the following quadratic, convex and constrained programming problem

$$\begin{aligned} &\min_{w,b} \ f(w,b) \ \text{ with } w \in \mathbb{R}^n \text{ and } b \in \mathbb{R} \\ &s.a \ \ g(w,b) \leq 0, \end{aligned} \tag{1}$$

where the functions $f : \mathbb{R}^{n+1} \to \mathbb{R}$ and $g : \mathbb{R}^{n+1} \to \mathbb{R}$ are continuously differentiable.

Methodology

In this work, we use SVM for classification. We consider the data set

$$\{(x^1, y_1), \ldots, (x^m, y_m), x^i \in \mathbb{R}^n, y_i \in \{-1, 1\}\}.$$

In the case Linear SVM, the problem of finding the optimal hyperplane can be formulated as

$$\begin{aligned} &\min_{w,b} \tfrac{1}{2}\|w\|^2 \qquad\qquad w \in \mathbb{R}^n, \ b \in \mathbb{R} \\ &s.a \ \ y_i(w^T x^i + b) \geq 1, \ i = 1, \ldots, m. \end{aligned} \tag{2}$$

In the case *CSVM*, we introduce m slack variables to penalize the objective function and give the constraints a slack. In the case *SVM - Non linear* we used a mapping function $\phi : \mathbb{R}^n \to \mathbb{R}^N$, $N > n$. The dual problem of the problem CSVM is

$$\begin{aligned} &\max_{\alpha} \ -\frac{1}{2}\sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j \phi(x^i)^T \phi(x^j) + \sum_{i=1}^{m} \alpha_i \\ &s.a \ \sum_{i=1}^{m} \alpha_i y_i = 0, \\ &\qquad 0 \leq \alpha_i \leq C, \ i = 1, \ldots, m. \end{aligned}$$

Then, we use the Kernel's trick.

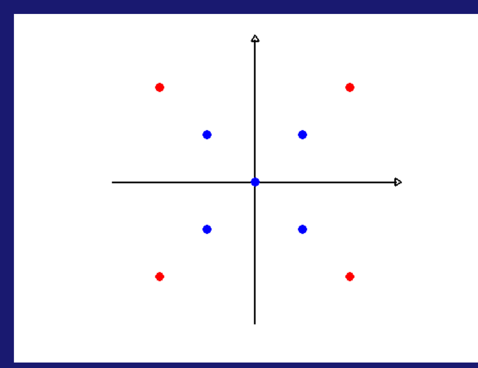

Figure 4: Data in $\mathcal{X}$.

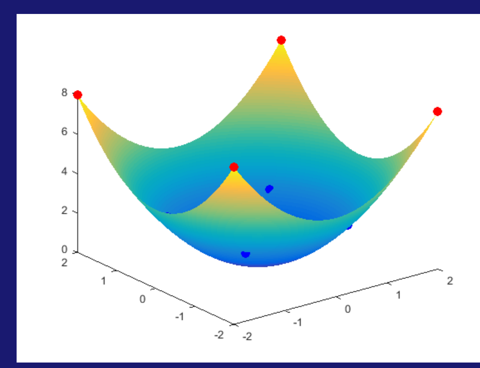

Figure 5: Surface S in $\mathbb{R}^N$.

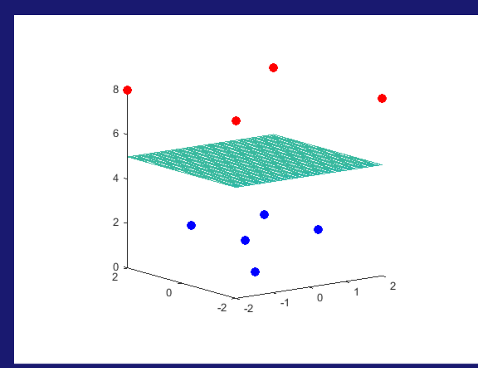

Figure 6: Hyperplane H in $\mathbb{R}^N$.

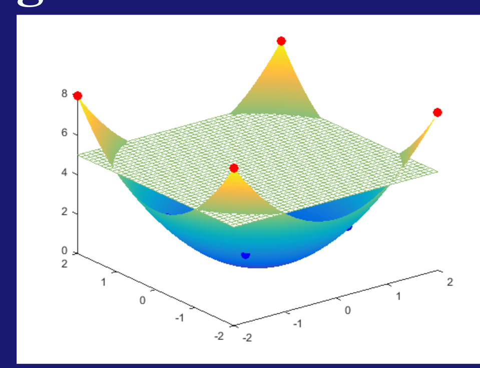

Figure 7: H and S.

Our aim is to study the problem with $x := (w, b)$ and the cardinality constraint $\| x \|_0 \leq \alpha < n$ is a given natural number [3]. We have considered the following standard relaxation, where X is the feasible set of the dual problem SVM, it's not (necessarily) polyhedral convex.

$$\begin{aligned} &\min_{x,y} \ f(x) \\ &s.t. \ x \in X, \\ &\qquad e^T y \geq n - \alpha, \\ &\qquad x_i y_i = 0, \ \text{for all } i = 1, \ldots, n, \\ &\qquad 0 \leq y_i \leq 1, \ \text{for all } i = 1, \ldots, n. \end{aligned} \tag{3}$$

Definition 1. *We say that a feasible point $(\bar{x}, \bar{y})$ of the relaxed problem (3) is weakly stationary (W-stationary) if there is $\lambda := (\lambda^g, \lambda^h, \lambda^\theta, \lambda^{\tilde{H}}, \lambda^G, \lambda^H) \in \mathbb{R}^m_+ \times \mathbb{R}^p \times \mathbb{R}_+ \times \mathbb{R}^n_+ \times \mathbb{R}^{|I_0(\bar{x})|} \times \mathbb{R}^n$ such that $\lambda^g_{\{1,\ldots,m\} \setminus I_g(\bar{x})} = 0$, $\lambda^\theta = 0$ if $\theta(\bar{y}) \neq 0$, $\lambda^{\tilde{H}}_{\{1,\ldots,n\} \setminus I_{01}(\bar{x},\bar{y})} = 0$, $\lambda^H_{I_{0+}(\bar{x},\bar{y}) \cup I_{01}(\bar{x},\bar{y})} = 0$ and $\nabla_{x,y} \mathcal{L}(\bar{x}, \bar{y}, \lambda^g, \lambda^h, \lambda^\theta, \lambda^{\tilde{H}}, \lambda^G, \lambda^H) = 0$, with $\mathcal{L}$ the Lagrangian function of TNLP$(\bar{x}, \bar{y})$ of (3).*

Proposition 1. *Let $(\bar{x}, \bar{y})$ be feasible for the TNLP$(\bar{x}, \bar{y})$ of the relaxed program (3). Then $(\bar{x}, \bar{y})$ is a stationary point of TNLP$(\bar{x}, \bar{y})$, i.e., satisfies the usual KKT conditions, if and only if $(\bar{x}, \bar{y})$ is an W-stationary.*

Conclusions

We have made theoretically the existence and uniqueness of the solutions of problems, obtained in SVM. We discuss the definitions of support vectors. More information on [1]. We achieved similar results to *Mathematical Problems with Complementarity Constraints* for *Mathematical Problems with Cardinality Constraints*. Our goal now is to get a sequential optimality condition for this last class of problems.

References

[1] E. H. M. Krulikovski. *Theoretical Analysis of vector machine support and application to character classification*. Masters dissertation. UFPR. 2017.

[2] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other Kernel-based learning methods*. Cambridge University Press, 2000.

[3] O. P. Burdavok, C. Kanzow and A. Schwartz. *Mathematical Programs with cardinality constraints: Reformulation by Complementarity-type Conditions and a Regularization Method*. SIAM. 2016.

[4] R. Andreani, G. Haeser, L.D. Secchin and P.J.S. Silva. *New sequential optimality conditions for mathematical problems with complementarity constraints and algorithmic consequences*. 2018.