Theoretical Analysis of Support Vector Machine

Evelin Heringer Manoel Krulikovski & Mael Sachine & Ademir Alves Ribeiro

Universidade Federal do Paraná

evelin.hmk@gmail.com

UFPR UNIVERSIDADE FEDERAL DO PARANÁ

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. 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

Introduction

In this work, we carried out a theoretical study of a supervised machine learning technique: the Support Vector Machine (SVM) [4, 5]. This technique focuses attention on the following quadratic, convex and constrained programming problem

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

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

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

The objective is to obtain a decision function, to separate the data in two subsets $X_1 = \{x^i \mid y_i = 1\}$ and $X_2 = \{x^i \mid y_i = -1\}$. Figure 1 represents a classification problem that can be modeled by the problem (1), which we are interested in solving. From the solution of this problem we construct a hyperplane defined by equation $w^T x + b = 0$, which will be used for classification of new data.

We deal with details of the SVM model with rigid margin and its generalizations, which are flexible and non-linear margin (Figures 1-3).

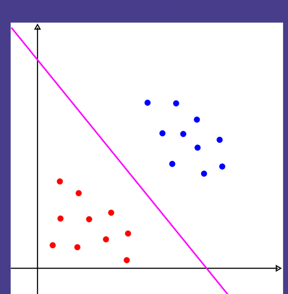

Figure 1 Linear.

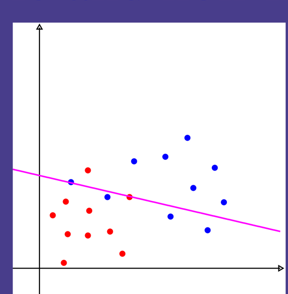

Figure 2 Flexible.

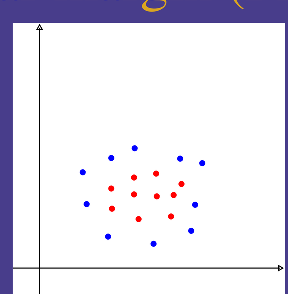

Figure 3 Non linear.

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

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

For the dual, we can guarantee the existence of a solution for the SVM with rigid margin.

In the case *CSVM*, we introduce m slack variables ($\xi_i \ge 0$) to penalize the objective function and give the constraints a slack. In this way, we have the following problem

$$\begin{aligned} \min_{w,b,\xi}\ & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \xi_i \\ s.a\ & y_i(w^T x^i + b) \ge 1 - \xi_i,\ i = 1, \dots, m, \\ & \xi_i \ge 0,\ i = 1, \dots, m \end{aligned}$$

with $C > 0$ a penalty constant, defined by the user.

For the SVM problem with flexible margin, only under certain conditions can we guarantee the uniqueness of the solution [2].

In the case *SVM - Non linear* we used a mapping function $\phi : \mathbb{R}^n \to \mathbb{R}^N$, $N > n$. Thus, the dual problem of the problem CSVM is

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

Then, we use the Kernel's trick.

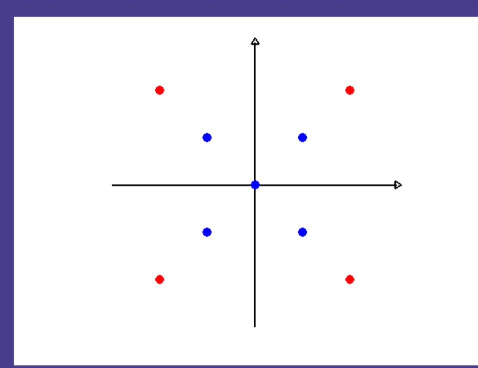

Figure 4: Data in $\mathcal{X}$.

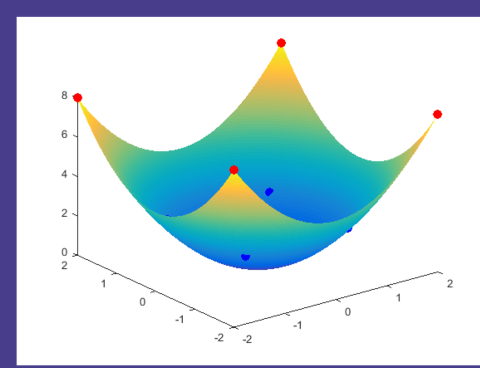

Figure 5: Surface S in $\mathbb{R}^N$.

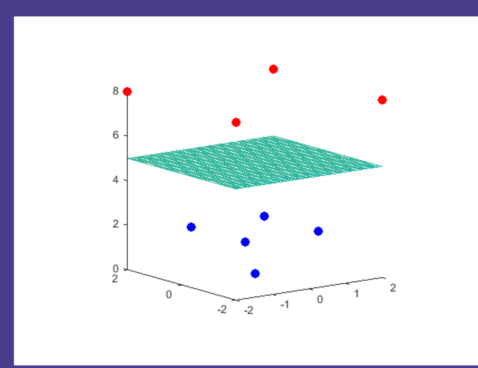

Figure 6: Hyperplane H in $\mathbb{R}^N$.

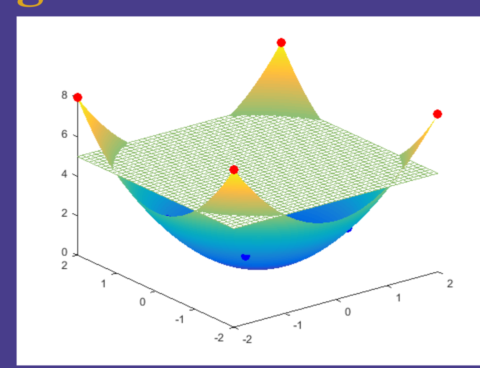

Figure 7: H and S.

Conclusions

The main contributions of this work are:

- We theoretically base the existence and uniqueness of the solutions of primal and dual problems, obtained in SVM. We discuss the definitions of support vectors, present in the literature [1, 3].
- We present changes in the results on the uniqueness of the primal solution, for SVM with flexible margin, found in [2].
- More information in [7].

References

[1] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006.

[2] C. J. C. Burges et al, *Uniqueness of the SVM Solution.* NIPS. v. 99, pp. 223-229, 1999.

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

[4] V. N. Vapnik and I. M. Guyon, *A training algorithm for optimal margin classifiers*. ACM. pp. 144-152, 1992.

[5] V. N. Vapnik and C. Cortes , *Support-vector networks*, Machine learning. Springer. v. 20. n. 3, pp. 273-297, 1995.

[6] P. Wolfe and M. Frank, *An Algorithm for Quadratic Programming*. Naval Research Logistics Quarterly. v. 3. pp. 95-110, 1956.

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

Acknowledgements

My teachers Mael Sachine and Ademir Alves Ribeiro. Supported by CAPES (Brazil).