A Fuzzy BI-Objective Model for SVM With an Interactive Procedure to Identify the Best Compromise Solution

Hager Ali Yahia[1], Mohammed Zakaria Moustafa[2],
Mohammed Rizk Mohammed[1] and Hatem Awad Khater[3]

[1]Department of Communication and Electronics Engineering,
Alexandria University, Alexandria, Egypt
[2]Department of Electrical Engineering (Power and Machines Section)
Alexandria University, Alexandria, Egypt
[3]Department of Mechatronics, Faculty of Engineering,
Horus University, Egypt

ABSTRACT

A support vector machine (SVM) learns the decision surface from two different classes of the input points. In many applications, there are misclassifications in some of the input points and each is not fully assigned to one of these two classes. In this paper a bi-objective quadratic programming model with fuzzy parameters is utilized and different feature quality measures are optimized simultaneously. An α-cut is defined to transform the fuzzy model to a family of classical bi-objective quadratic programming problems. The weighting method is used to optimize each of these problems. An important contribution will be added for the proposed fuzzy bi-objective quadratic programming model by getting different efficient support vectors due to changing the weighting values. The experimental results show the effectiveness of the α-cut with the weighting parameters on reducing the misclassification between two classes of the input points. An interactive procedure will be added to identify the best compromise solution from the generated efficient solutions.

KEYWORDS

Support vector machine (SVMs); Classification; Multi-objective problems; Weighting method; fuzzy mathematics; Quadratic programming; Interactive approach.

1. INTRODUCTION

Support Vector Machines (SVMs) are a classification technique developed by Vapnik at the end of '60s [1]. The theory of support vector machines (SVMs) is a new classification technique and has drawn much attention on this topic in recent years [6]. Since then the technique has been deeply improved, being applied in many different contexts. In many applications, SVM has been shown to provide higher performance than traditional learning machines [6]. SVMs are known as maximum margin classifiers, since they find the optimal hyperplane between two classes as shown in figure1, defined by a number of support vectors [4].
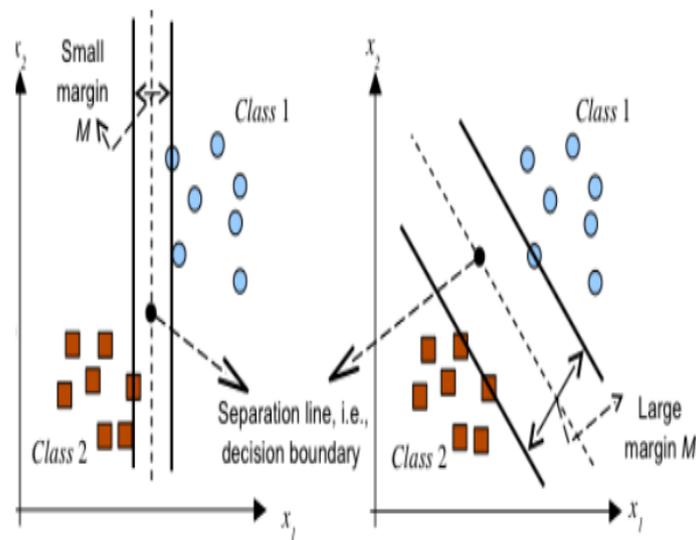
Figure 1: maximization of the margin between two classes

The well-known generalization feature of the technique is mainly due to the introduction of a penalty factor, named C that allows us to prevent the effects of outliers by permitting a certain amount of misclassification errors.

In this paper, the idea is to apply the fuzzy multi-objective programming technique for developing the set of all efficient solutions for the classification problem with minimum errors. An α-cut is taken to transform the fuzzy multi-objective problem model to a classical one (α problem).The weighting method is used to solve the α problem proposed to generate the set of efficient solutions for the proposed model. The remainder of this paper is organized as follows. A brief review for the SVM is described in section 2. The proposed fuzzy bi-objective model for the Support Vector Machine will be derived in section 3. NEXT, section 4 presents three numerical examples corresponding to three different α-cut. Section 5 provides our general conclusions.

2. SUPPORT VECTOR MACHINES

SVM is an efficient classifier to classify two different sets of observations into their relevant class as shown in figure 2 where there are more than straight line separates between the two sets. SVM mechanism is based upon finding the best hyperplane that separates the data of two different classes of the category.

The best hyperplane is the one that maximizes the margin, i.e., the distance from the nearest training points [2].

Support vector machine has been utilized in many applications such as biometrics, chemoinformatics, and agriculture. SVM has penalty parameters, and kernel parameters that have a great influence on the performance of SVM [3]. We review the basis of the theory of SVM in classification problems [7].

Let a set S of labelled training points

$$(y_1, x_1)\ldots (y_l, x_l). \qquad (1)$$

Where, $x_i \in \mathcal{R}^N$ belongs to either of two classes and is given a label $y_i = \{-1,1\}$ for $i = 1, ..., l$.
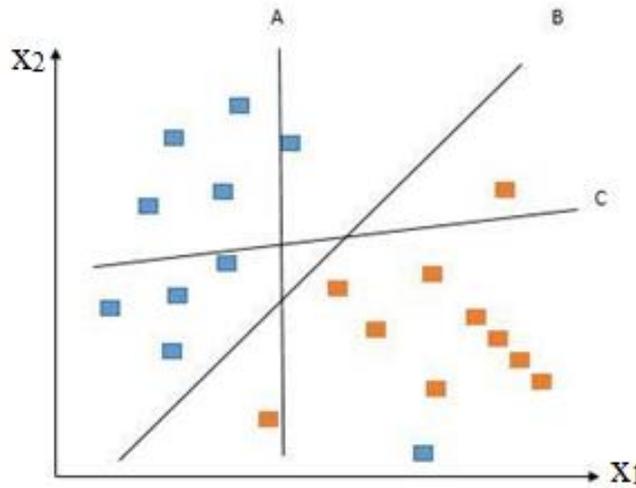

Figure 2: Data classification using support vector machine

In some cases, to get the suitable hyperplane in an input space, mapping the input space into a higher dimension feature space and searching the optimal hyperplane in this feature space.

Let $z = \varphi(x)$ denote the corresponding feature space vector with mapping φ from $\mathcal{R}^N$ to a feature space ƶ. We wish to find the hyperplane

$$w.z + b = 0 \qquad (2)$$

defined by the pair (w, b) according to the function

$$f(x_i) = sign(w.z_i + b) = \begin{cases} 1, & if\, y_i = 1 \\ -1, & if\, y_i = -1 \end{cases} \qquad (3)$$

where $w \in$ ƶ and $b \in \mathcal{R}$. For more precisely the equation will be

$$\begin{cases} (w.z_i + b) \geq 1, & if\, y_i = 1 \\ (w.z_i + b) \leq -1, & if\, y_i = -1, \end{cases} i = 1, ..., l \qquad (4)$$

For the linearly separable set S, we can find a unique optimal hyperplane for which the margin between the projections of the training points of two different classes is maximized.

For the data that are not linearly separable figure 3, the previous analysis can be generalized by introducing some nonnegative variables $\xi_i \geq 0$ then,

$$y_i(w.z_i + b) \geq 1 - \xi_i, \qquad i = 1, ..., l. \qquad (5)$$

The term $\sum_{i=1}^{l} \xi_i$ can be thought of as some measure of the amount of misclassifications.

The optimal hyperplane problem is then regarded as the solution to the problem

$$minimize \frac{1}{2}w.w + C \sum_{i=1}^{l} \xi_i$$

$$subject\ to\ y_i(w.z_i + b) \geq 1 - \xi_i\ , \qquad (6)$$

$$i = 1, \dots, l$$

$$\xi_i \geq 0, \quad i = 1, \dots, l$$

where, C is a constant. The parameter C can be regarded as a regularization parameter [5]. SVM algorithms use a set of mathematical functions that are defined as the kernel.

The function of kernel is to take data as input and transform it into the required form. Different SVM algorithms use different types of kernel functions. For example, linear, nonlinear, polynomial, radial basis function (RBF), and sigmoid.

Basically, the training part consists in finding the best separating plane (with maximal margin) based on specific vector called support vector. If the decision is not feasible in the initial description space, you can increase space dimension thanks to kernel functions and may be find a hyperplane that will be your decision separator.
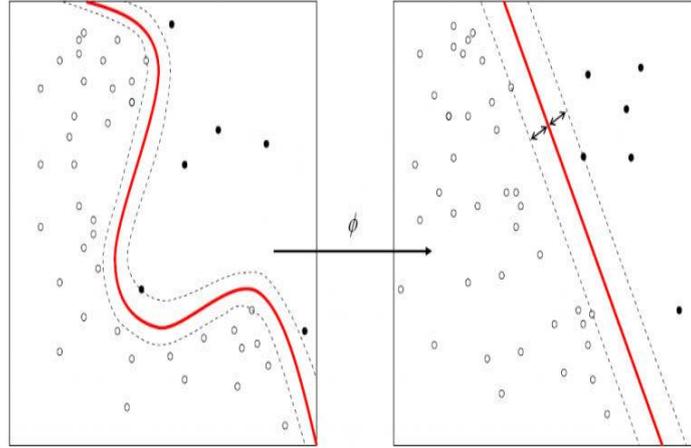

Figure 3: linearly separable and nonlinearly separable

3. FORMULATION OF THE FUZZY BI-OBJECTIVE QUADRATIC PROGRAMMING MODEL OF SVM

In this section, we make a detail description about the idea and formulation of the fuzzy biobjective programming model for the SVM. SVM is a powerful tool for solving classification problems, but due to the nonlinearity separable in some of the input data, there is an error in measuring the amount of misclassification. In the same time, in many real-world applications, each of the input points does not exactly belong to one of the two classes [11].

From this point of view, we reformulate the classical model of the SVM to the following bi-objective programming model with fuzzy parameters.

3.1. The fuzzy bi-objective support vector machine (FSVM):

Now, we add another objective function with fuzzy parameters $\tilde{\upsilon}_i, i = 1,2, \ldots, l$ for the previous model in section 2 to be in the form

$$\text{Min} \parallel w \parallel^2,$$

$$\text{Min} \sum_{i=1}^{l} \tilde{\upsilon}_i \, \xi_i$$

$$\text{Subject to} \qquad\qquad (7)$$

$$y_i(w.x_i + b) \geq 1 + \xi_i \quad , i = 1,2, \ldots, l$$

$$\xi_i \geq 0 \qquad , i = 1,2, \ldots, l$$

By taken an α-cut for the membership functions corresponding to the fuzzy parameters $\tilde{\upsilon}_i$, $i = 1,2, \ldots, l$, we get the following α-problem:

$$\text{Min} \parallel w \parallel^2,$$

$$\text{Min} \sum_{i=1}^{l} \alpha_i \xi_i$$

$$\text{Subject to} \qquad\qquad (8)$$

$$y_i(w.x_i + b) \geq 1 + \xi_i \quad , i = 1,2, \ldots, l$$
$$\xi_i \geq 0 \qquad , i = 1,2, \ldots, l$$
$$\sigma \leq \alpha_i \leq 1 \qquad , i = 1,2, \ldots, l$$

With sufficient small $\sigma > 0$.

Where the parameter ξ_i is a measure of the error in the SVM and the term $\alpha_i \xi_i$ is a measure of the error with different degrees α_i. The (α-problem) is solved by the weighting method to get the set of all efficient solutions.

This problem is a bi-objective quadratic programming problem. The first objective is to maximize the gap between the two hyperplanes which used to classify the input points. The second objective is to minimize the error (with different degrees α_i , $i = 1,2, \ldots, l$) in measuring the amount of misclassification in case of nonlinearity separable input points [11].

Problem 8 can be solved by the weighting method to get the set of all efficient solutions for the classification problem.

The right choice of weightage for each of these objectives is critical to the quality of the classifier learned, especially in case of the class imbalanced data sets. Therefore, costly parameter tuning has to be undertaken to find a set of suitable relative weights [10].

3.2. The weighting method

In this method each objective $f_i(X), i = 1,2, \ldots, k$, is multiplied by a scalar weigh $w_i \geq 0 \, and \sum_{i=1}^{k} w_i = 1$. Then, the k weighted objectives are summed to form a weighted-sums objective function [8][12].

$$Assume W \text{ as } \begin{cases} w \in R^k : w_i \geq 0, \\ i = 1, 2, \dots, k \\ and \ \sum_{i=1}^{k} w_i = 1 \end{cases} \qquad (9)$$

be the set of nonnegative weights. Then the weighting problem is defined as:

$$P(W): Min \sum_{i=1}^{k} w_i f_i$$
$$\text{Subject to } M = \begin{cases} X \in R^n : g_r(X) \leq 0, \\ r = 1, 2, \dots, m \end{cases}. \qquad (10)$$

Then, in this paper the weighting method takes the form

$$\text{Inf } z = w_1 \parallel w \parallel^2 + w_2 \sum_{i=1}^{l} \alpha_i \xi_i$$
$$\text{Subject to}$$
$$y_i(w.x_i + b) \geq 1 + \xi_i , i = 1, 2, \dots, l$$
$$\xi_i \geq 0 \qquad , i = 1, 2, \dots, l \qquad (11)$$

$$w_1 > 0, w_2 \geq 0$$
$$w_1 + w_2 = 1$$
$$\sigma \leq \alpha_i \leq 1 \qquad , i = 1, 2, \dots, l$$
$$\text{With sufficient small } \sigma > 0$$

Here we use "Inf "instead of "Min" since the set of constraints is unbounded, where $w_1 0$.

Also, we avoid the redundant solutions by adding the constraint $w_1 + w_2 = 1$.

3.3. An interactive procedure to identify the best compromise solution

For the version of our bi-objective (SVM) model which applies to determine the best compromise solution, we need the following hypothesis (after the interaction with the decision maker)[13]:

The best compromise solution for the set of the generated efficient solution is that efficient one corresponding to

$$\min_{\alpha} N^+ \leq \min_{\alpha} N^-$$

Where, N^- is the number of support vectors of the negative class,
N^+ is the number of support vectors of the positive class.

We must notice that this hypothesis can be reversed according to the preference of the decision maker (see Yaochu Jin, 2006) [9].

4. NUMERICAL EXAMPLES

By using python program, we can solve the previous problem and show the effect of different values of the weighting parameters. The data set that is used in these examples consist of 51 points and each point has two features, table 1 shows part of this data.

Table 1: Description of part of datasets used in our study.

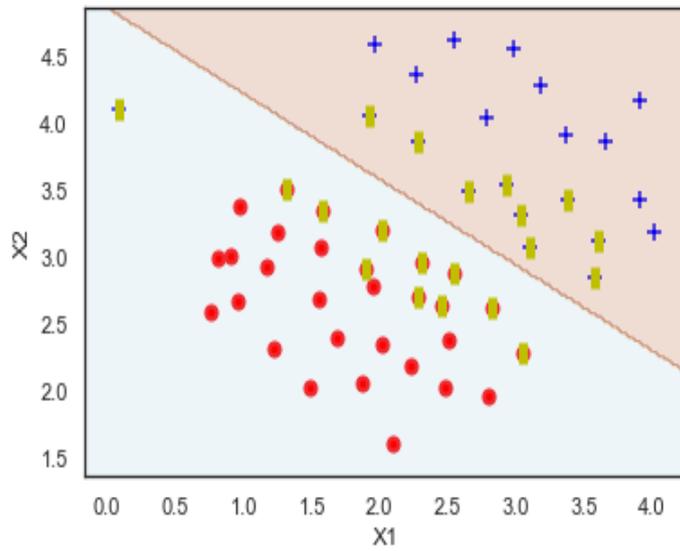| X1	X2	Y
1.9643	4.5957	1
2.2753	3.8589	1
2.9781	4.5651	1
2.932	3.5519	1
3.5772	2.856	1
0.9044	3.0198	0
0.76615	2.5899	0
0.086405	4.1045	0

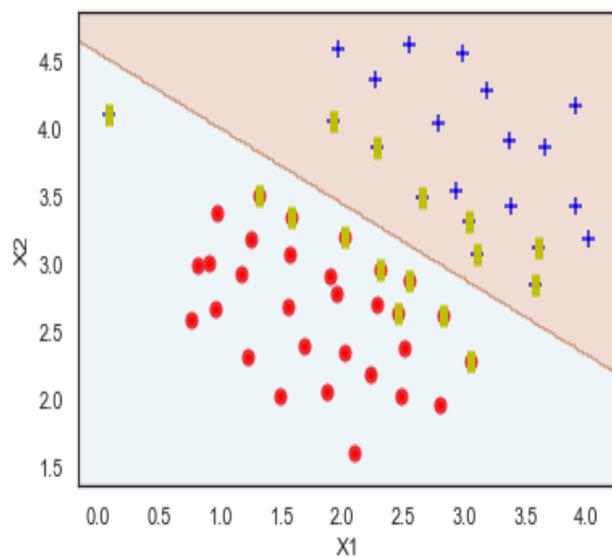Figure 6: $w_2 = \frac{1}{2}, w_1 = \frac{1}{2}$, α=$\frac{1}{4}$, number of support vectors = 20

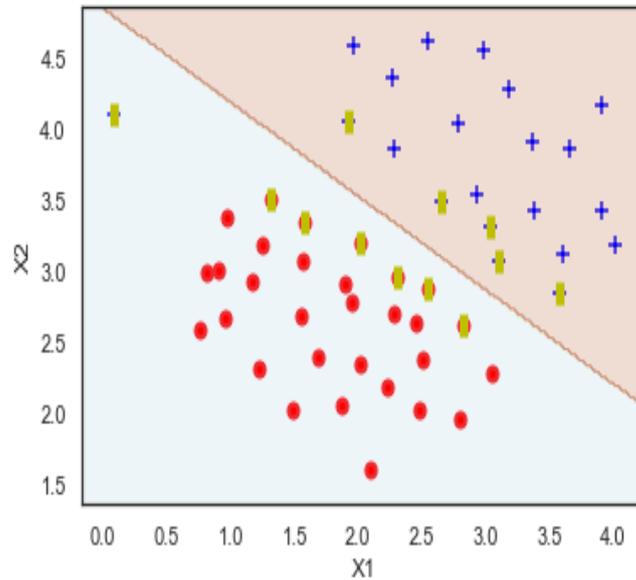Figure 5: $w_2 = \frac{1}{2}, w_1 = \frac{1}{2}$, α=$\frac{1}{2}$, number of support vectors = 16

Figure 4: $w_2 = \frac{1}{2}, w_1 = \frac{1}{2}$, α=1, number of support vectors = 12

So, the previous results, by using different degrees (α) at the same weights (w_1 & w_2), show how these parameters (α, w_1, w_2) effect on the performance of SVM. When the value of α is increased the number of support vectors is reduced.

There are good reasons to prefer SVMs with few support vectors (SVs). In the hard-margin case, the number of SVs (#SV) is an upper bound on the expected number of errors made by the leave-one-out procedure [9].

So, we can control the performance of SVM according to our requirements by adjusting the values of the parameters (α, w_1, w_2).

According to our hypothesis that presented in section 3.3, the best compromise solution is that corresponding to α=1.

5. CONCLUSIONS

This paper introduced the fuzzy multi-objective programming technique for developing the set of all efficient solutions for the classification problem with minimum errors. The weighting method is used to solve our fuzzy model after defuzzification by using the α − cut technique. The experimental evaluation was carried out using 51 datasets, each one has two features. The experimental results show the effect of the parameters (α, w_1, w_2) on the misclassification between two sets.

Our future work is to try to construct a utility function to select the best compromised hyperplane from the generated set of the efficient solutions.

An interactive procedure is added to identify the best compromise hyperplane from the generated efficient set.

REFERENCES

[1] Cortes, Corinna; Vapnik, Vladimir N (1995) "Support vector networks" (PDF). Machine learning. 20 (3):273297. CiteSeerX 10.1.1.15.9362. DOI:10.1007/BF00994018.

[2] Asmaa Hamad1,3(B), Essam H. Houssein1,3, Aboul Ella Hassanien2,3, and Aly A. Fahmy2 :Hybrid Grasshopper Optimization Algorithm and Support Vector Machines for Automatic Seizure Detection in EEG Signals. Faculty of Computers and Information, Minia University, Minya, Egypt. January 2018. DOI: 10.1007/978-3-319-74690-6_9.

[3] Alaa Tharwat1;_, Thomas Gabel1, Aboul Ella Hassanien2;_ Parameter Optimization of Support Vector Machine using Dragon_y Algorithm. Faculty of Computer Science and Engineering, Frankfurt University of Applied Sciences, Frankfurt am Main, Germany ,Faculty of Computers and Information, Cairo University, Egypt. January 2018 DOI: 10.1007/978-3-319-64861-3_29.

[4] Gray, D., Bowes, D., Davey, N., Sun, Y., Christianson, B.: Using the Support Vector Machine as a Classification Method for Software Defect Prediction with Static Code Metrics. In: Palmer Brown, D., Draganova, C., Pimenidis, E., Mouratidis, H. (eds.) EANN 2009. Communications in Computer and Information Science, vol. 43, pp. 223–234. Springer,Heidelberg (2009).

[5] Chun-Fu Lin and Sheng-De Wang: "Fuzzy Support Vector Machines". Article in IEEE Transaction on neural networks, VOL.13, NO.2, March 2002. DOI:10.1109/72.991432.

[6] C. Burges, "A tutorial on support vector machines for pattern recognition," Data Mining and Knowledge Discovery, vol.2, no.2,1998.

[7] C. Cortes and V. N. Vapnik, "Support vector networks," Machine Learning, vol.20,pp.273-297,1995.

[8] Chankong V. and Haimes Y.Y., Multi-objective Decision-Making: Theory and Methodology (North Holland Series in System Science and Engineering, 1983).

[9] YaochuJin (Ed.), Multi-objective Machine Learning Studies in Computational Intelligence, Vol. 16, pp. 199-220, Springer-Verlag, 2006.

[10] Shounak Datta and Swagatam Das, Multiobjective Support Vector Machines: Handling Class Imbalance With Pareto Optimality, IEEE Transactions on Neural Networks and Learning Systems, 2019.DOI:10.1109/TNNLS.2018.2869298.

[11] Y.Wang, Shouyang Wang, K.K.Lai, "A new fuzzy support vector machine to evaluate credit risk," IEEE Transactions on Fuzzy Systems, VOL.13, NO.6, January 2006, DOI:10.1109.

[12] Mohammed Zakaria Moustafa, Mohammed Rizk Mohammed, Hatem Awad Khater and Hager Ali Yahia, Building a Bi-objective Quadratic Programming Model for The Support Vector Machine, 8th International Conference on Artificial Intelligence, Soft Computing (AISC 2020), DOI: 10.5121/csit.2020.100208.

[13] Mohammed Zakaria Moustafa, Mohammed Rizk Mohammed, Hatem Awad Khater and Hager Ali Yahia, A bi-objective Model for SVM with an interactive procedure to identify the best compromise solution, International Journal of Artificial Intelligence and Applications (IJAIA), Vol.11, No.2, March 2020.