Modifying the synthesis of the soft margin optimal hyperplane for support vector machines

Mario Antonio Ruz Canul[1], Jose A. Ruz-Hernandez[2], Alma Y. Alanis[1], Ramon Garcia-Hernandez[3], and Jose-Luis Rullan-Lara[2]

[1] Centro Universitario de Ciencias Exactas e Ingenierias, Universidad de Guadalajara, Blvd. Marcelino Garcia Barragan, Guadalajara 44430, Jalisco, Mexico.
mario.ruz@academicos.udg.mx, alma.alanis@academicos.udg.mx
[2] Facultad de Ingenieria, Universidad Autonoma del Carmen, C.56 No.4 Esq. Avenida Concordia Col. Benito Juarez, Ciudad del Carmen 24180, Campeche, Mexico.
jruz@pampano.unacar.mx, jrullan@pampano.unacar.mx
[3] Tecnologico Nacional de Mexico, Instituto Tecnologico de La Laguna, Blvd. Revolución y Av. Instituto Tecnologico de La Laguna s/n, Torreon 27000, Coahuila, Mexico.
rgarciah@lalaguna.tecnm.mx

Abstract. This paper presents a modification to the Optimal Hyperplane synthesis to improve the linear separating properties of Support Vector Machines. The proposed modification is based on a modified optimization criterion, which is extended to non-linearly separable training data via the Soft Margin Optimal Hyperplane. Performance metrics demonstrate that the proposed modification maximizes the margin during hyperplane construction by increasing the average distance of both training patterns and support vectors, thereby improving classification accuracy.

Keywords: Support Vector Machine, Soft Margin Hyperplane, Optimization, Pattern Classification

1 Introduction

Pattern classification is the process of assigning objects to predefined categories. This process relies on classifiers: software tools designed for accurate categorization. These classifiers utilize carefully selected input features that effectively represent and distinguish between classes within the input space [1]. Support Vector Machines (SVMs), a prominent supervised learning technique used in a wide range of applications, are recognized for their powerful computational capabilities in classification [2]. Despite their development in the 1990s [3], SVMs continue to be relevant and applicable to diverse problems. SVMs operate on the principle of non-linearly mapping input vectors into a high-dimensional feature space, within which a linear decision surface is constructed. Specific properties of this surface promote strong generalization performance. The foundational concept of SVMs is the identification of an optimal separating hyperplane in the

https://doi.org/10.61728/AE20254933

restricted case of linearly separable training data. Subsequently, this approach is generalized to accommodate non-separable data through the introduction of the soft margin hyperplane.They have shown particularly strong performance in areas such as image classification for breast cancer detection, frequently surpassing other methods [4], and have also been successfully combined with artificial neural networks for electrocardiogram heartbeat classification [5]. A comparative analysis by [6] evaluated several machine learning algorithms, including K-Nearest Neighbor, Genetic Algorithms, Decision Tree, Long Short Term Memory networks, and SVMs. Across various problems, SVMs and Long Short Term Memory networks demonstrated superior performance in both qualitative and quantitative assessments.

Recent years have seen the development of novel approaches and modifications aimed at enhancing the performance of SVMs, driven by the continued growth of artificial intelligence and advances in soft computing. One such approach is the Firefly Algorithm-optimized SVM, which has demonstrated improved classification accuracy compared to other methods [7]. A hybrid variable mode decomposition-support vector machine model, coupled with a novel input selection framework, has demonstrated its utility and efficacy for hydrological prediction under climate change scenarios [8]. SVMs integrated with metaheuristic algorithms have gained prominence, as exemplified by the Hybrid Grey Wolf Optimizer-Improved Sine Cosine Algorithm -enhanced SVM for electroencephalography signal classification. Experimental results demonstrate high classification rates indicating the system's potential for epilepsy classification [9]. Improved Fuzzy Least Squares Twin Support Vector Machine are introduced in [10] and improves upon traditional SVMs for pattern classification by using both structural and empirical risk minimization. This method provides benefits in terms of speed and generalization. Quantum Support Vector Machine Feature selection is an innovative method developed in [11] that merges quantum SVMs with multi-objective genetic algorithms. The goal of this technique is to decrease the dimensionality of data while boosting the effectiveness and precision of machine learning models.

In the present paper, authors propose a modification to the criterion usually utilized to determine optimal separating hyperplane or soft margin optimal hyperplane in synthesis of SVMs, [12]. This modification improves properties of SVMs, for linearly and non-linearly separable patterns. In order to do so, respective optimization process prioritizes to minimize effects corresponding to large classification errors. Furthermore, simulation results for this new optimization criterion are obtained, using Quadratic Programming (QP). This algorithm is tested using the well-known Pima Indians Diabetes database [13].

Paper structure presents in Section 2 the mathematical preliminaries where the construction of Optimal Hyperplane (OH) and Soft Margin Optimal Hyperplane (SMOH) for SVMs are presented. In Section 3, proposed modification for synthesis of optimal hyperplane named PSMOH is introduced and new objective functions to acquire correspond parameter of model are developed. Section 4 presents simulation results employing the Pima Indians Diabetes Database

[13] to compare performance of conventional methods and proposed new one, where distances of patterns with obtained hyperplane for each case and average distance of the support vectors are analyzed. Furthermore, metrics such as precision, sensitivity and specificity for classification are obtained. Finally, Section 5 describes conclusions of proposed PSMOH and future works for particular cases where $p > 2$.

2 Mathematical Preliminaries

In this section main concepts for construction of OH for simple case of linearly separable patterns are described [3]. Following with a more difficult case of non-linearly separable patterns where the SMOH is introduced.

2.1 Optimal Hyperplane of SVM

SVM is commonly used for two class classification problems where a training set is used to train the model. Training set O of M training examples can be defined as follows:

$$O = \{(X_1, y_1), (X_2, y_2), \ldots, (X_M, y_M)\} \tag{1}$$

where $X_i \in R^n$ is the input pattern for i^{th} example and $y_i \in \{-1, 1\}$ is a class label with $i = 1, 2, ..., M$. Correct classification of data points in a linearly separable case is done by

$$W^T X_i + b \geq 1, for \quad y_i = 1 \tag{2}$$

$$W^T X_i + b < 1, for \quad y_i = -1 \tag{3}$$

Equations (2) and (3) can be combined into one set of inequalities:

$$y_i(W^T X_i + b) \geq 1, \quad i = 1, 2, \ldots, M \tag{4}$$

Signed distance G_i, between X_i and the OH, can be calculated by:

$$G(X_i) = G_i = \frac{(W^T X_i + b)}{||W||} \tag{5}$$

where $||W||$ is the Euclidean norm of W. In order to maximize distance between classes and W, its norm should be minimized satisfying (2) at the same time according to [3]. W norm can be minimized by means of:

$$F(W) = \frac{W^T W}{2} \tag{6}$$

Then, by [3] to determine W and b for the OH it is required to minimize (6) satisfying constraint (4). Lagrange multiplier approach can be applied [12, 14] as:

$$L(W, b, A) = \frac{W^T W}{2} - \sum_{i=1}^{M} a_i [y_i (W^T X_i + b) - 1] \tag{7}$$

where $A = (\alpha_1, \alpha_2, \ldots, \alpha_M)$ is the Lagrange multipliers vector subject to $\alpha_i \geq 0$ for $i = 1, 2, \ldots, M$. To minimize (7) with respect to W, b and A, the respective saddle point should be determined by minimizing it with respect to W and b, and maximizing it with respect to the Lagrange multipliers. According to [3, 15], after obtaining the respective minimization, then (7) can be transformed to $LL(A)$ as:

$$LL(A) = L(W, b, A) = \sum_{i=1}^{M} a_i - \frac{A^T D A}{2} \tag{8}$$

with D as a positive semi-definite matrix [3], with $i = 1, 2, \ldots, M$ defined by:

$$D = \sum_{i=1}^{M} \sum_{j=1}^{M} y_i y_j X^T X \tag{9}$$

$LL(A)$ depends only on (1) and Lagrange multipliers; therefore, minimization of (6), including constraints (4), is reduced to maximize LL(A) by means of Lagrange multipliers (α_i) satisfying:

$$\sum_{i=1}^{M} \alpha_i y_i = 0, \tag{10}$$

$$\alpha_i \geq 0 \tag{11}$$

with $i = 1, 2, \ldots, M$. It is worth noting that (10) is derived from the minimization of (7) with respect to b. If the solution of $LL(A)$ exists, then it defines W and b for OH, and W can be formulated as a function of optimal Lagrange multipliers as follows:

$$W = \sum_{i=1}^{M} \alpha_i y_i X_i \tag{12}$$

with, again $i = 1, 2, \ldots, M$. Although the OH has seen application in several works, its effectiveness is limited to linearly separable data [1].

2.2 Soft Margin Optimal Hyperplane

SMOH is introduced in [15, 16] as an alternative to relax the linearly separable classes condition. The respective constraints are specified as:

$$y_i(W^T X_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \ldots, M \tag{13}$$

with $E = (\xi_i, \xi_2, \ldots \xi_M)^T$ and $\xi_i \geq 0$ for $i = 1, 2, \ldots, M$. Search region for (13) is larger than one for (4), allowing to find a better solution for optimization. The e_i values could be interpreted as deviations from SMOH and they are minimized by means of the following criterion:

$$H = \sum_{i=1}^{M} e_i^{\sigma} \quad with \quad \sigma > 0 \tag{14}$$

In [15], it is also recommended to select σ "sufficiently small". In [16], the following statement is presented:"For computational reasons, however, in this work it is consider the case $\sigma = 1$. This case corresponds to the smallest $\sigma > 0$ that is still computationally simple". For $\sigma < 1$, respective optimization problem may not be convex, due to computational complexity. Then $\sigma = 1$ is the smallest value which can guarantee computational feasibility, and so it is selected for (14). W and b for the SMOH can be determined by minimization of (15) with constraint (13)

$$SMF(W) = \frac{W^T W}{2} + C \sum_{i=1}^{M} e^i \tag{15}$$

where $C > 0$ is arbitrarily defined according to the problem to be solved. In fact, it is a weighting factor to balance the importance of two components in (15). To solve $SMF(W)$, Lagrange multipliers can be applied, whose solution corresponds to the problem:

$$SMLL(A) = \sum_{i=1}^{M} \alpha_i - \frac{A^T D A}{2} \tag{16}$$

subject to

$$\sum_{i=1}^{M} \alpha_i y_i = 0 \tag{17}$$

$$0 \leq \alpha_i \leq C \tag{18}$$

where D is a positive semi-definite matrix given by (9). Maximization of (16), with constraints (17) and (18), is a quadratic programmatic (QP) problem that can be easily solved [14]. When comparing the two dual maximization problems, OH (8) and SMOH (16), it can be seen that they are very similar; the only differences are the respective constraints (11) and (18).

3 Proposed Modification PSMOH

Optimization results depend on the respective criterion, which should be adequate to the problem to be solved. In Fig. 1, the effects of $e > 0$ on the function e^{σ} are illustrated for different positive values of σ. There, $''*''$ corresponds to $e = 0.1$, $''+''$ to $e = 0.5$, $''\circ''$ to $e = 0.9$ and $''\times''$ to $e = 1.1$. From the effects obtained of e it is evident that for criterion (14), when $\sigma < 1$, the effect is similar and independent of the values of e, and the function e^{σ} converges to 1. Then, using (14), the optimization process considers small e values (which do not produce classification errors) in a similar way, which means that (14) is not an adequate criterion for classification.

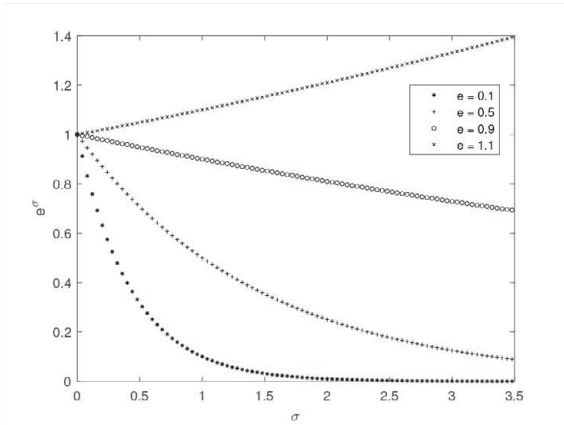

Fig. 1. e effect on e^{σ} function.

In order to select a more adequate criterion, its respective optimization process should be in priority minimize the effects corresponding to large e (producing classification errors). In order to do so, the following criterion could be used:

$$K = \sum_{i=1}^{M} Q_i e_i \tag{19}$$

Now the weight values are defined such that they will be larger for large e values, and smaller for small ones, as follows:

$$Q_i = e_i^{p-1} \tag{20}$$

for $i = 1, 2, ..., M$ and $p > 1$. Substituting (20) in (19), the following criterion is obtained:

$$K = \sum_{i=1}^{M} Q_i e_i = \sum_{i=1}^{M} e_i^{p-1} e_i = \sum_{i}^{M} e_i^{p} \tag{21}$$

with $p > 1$. From (21), the respective optimization process prioritizes large e values, which produce classification errors.

In Fig. 1, it is possible to see difference of influences for large values of p (or σ as described in the figure) and also for small increases of e when p is large; hence, the proposed criterion (21) is very different from criterion (14) in [15]. In fact for (21) p should be big enough; instead of σ small enough as stated in (14). Then, $\sigma = 1$ (as suggested in [1]) is the maximum value for criterion (14), and $p = 1$ is the minimum one for criterion (21). Additionally, in [16] it is stated that due to computationally, efficiency $\sigma = 1$ is the only allowable value for (14). Due to similar considerations, for (21), p can take any integer value bigger or equal to 1 ($p \geq 1$), its specific value depends on the problem to be solved.

To define the SMOH parameters, on the basis of criterion (21), it is required to minimize the following function, named $PSMF(W)$.

$$PSMF(W) = \frac{W^T W}{2} + C \sum_{i=1}^{M} e_i^p \tag{22}$$

with $p \geq 1$ and $C > 0$ where C is arbitrarily defined according to the problem to be solved. In fact, it is a weighting factor case as it explained for (15). In order to guarantee convexity, p can be any integer value larger or equal to 1. Considering that $PSMF(W)$ is convex and (13) depends linearly on W, b and e_i, then to solve this minimization problem no negative Lagrange multipliers (α) can be applied [3], with the respective function formulated as:

$$PSML(W, b, A) = \frac{W^T W}{2} + C \sum_{i=1}^{M} e_i^p - \sum_{i=1}^{M} \alpha_i [y_i (w^T X_i + b) - 1 + e_i] - \sum_{i=1}^{M} m_i e_i \tag{23}$$

where α_i are the Lagrange multipliers for (13), and m_i are the ones for $e_i > 0$. To solve (22), with constrains (13), the Lagrange multipliers can be applied, whose solution corresponds to the respective saddle point of (23), obtained by minimizing with respect to w, b and e_i, and simultaneously, maximizing with respect to the Lagrange multipliers.

The conditions for the minimum of $PSML$ (23) are:

$$\frac{\partial PSML}{\partial W} = W - \sum_{i=1}^{M} \alpha_i y_i X_i = 0 \tag{24}$$

$$\frac{\partial PSML}{\partial b} = - \sum_{i=1}^{M} \alpha_i y_i = 0 \tag{25}$$

$$\frac{\partial PSML}{\partial e_i} = pC e_i^{p-1} - \alpha_i - m_i = 0 \tag{26}$$

From (24), W can be obtained as:

$$W = \sum_{i=1}^{M} \alpha_i y_i X_i = 0 \tag{27}$$

and from (25) the following constraint is defined:

$$\sum_{i=1}^{M} \alpha_i y_i = 0 \tag{28}$$

According to Kühn-Tucker conditions [3], the optimum point for $e_i \neq 0$ corresponds to $m_i = 0$; then from (26), e_i is formulated as:

$$e_i = \frac{\alpha_i}{pC}^{\frac{1}{p-1}} \tag{29}$$

Substituting (27)-(29) in (23), and then transforming it, the following function is obtained:

$$\begin{aligned} PSMLL(A) &= \sum_{i=1}^{M} \alpha_i - \frac{A^T D A}{2} + \left[\frac{C}{(pC)^{\frac{p}{p-1}}} - \frac{1}{pC}^{\frac{1}{p-1}} \right] \sum_{i}^{M} \alpha_i^{\frac{p}{p-1}} \\ &= \sum_{i=1}^{M} \alpha_i - \frac{A^T D A}{2} + C(1-p) \sum_{i}^{M} \left[\frac{\alpha_i}{pC} \right]^{\frac{p}{p-1}} \end{aligned} \tag{30}$$

$PSMLL(A)$ depends only on the training examples (1) and the Lagrange multipliers. Hence, the minimization of (22) with constraints (13) is transformed to maximization of $PSMLL(A)$, with constraints (11) and (28). The values of the PSMOH parameters can be determined by means of (13), (27) and (29), and the optimal values of the Lagrange multipliers.

In particular, the optimal value for b is calculated using the following equation:

$$b = \frac{(1-(\frac{\alpha_i}{pC}))^{\frac{p}{p-1}}}{y_i} - W^T X_i \tag{31}$$

In connection with (1), minimizing the functional in (22) includes $N + M$ variables, whereas its dual (30) requires only M variables. Considering (32) and $p = 1$ in (30), hence.

$$\frac{\alpha_i}{C} \leq 1 \tag{32}$$

For $i = 1, ..., M$.

3.1 First case of the proposed modification, $p = 1$

Consider the scenario wherein the parameter p approaches the value of 1 from the right-hand side. In order to prevent the final term of equation (30) from becoming excessively large, it is necessary to:

$$PSMLL_1(A) = \sum_{i=1}^{M} \alpha_i - \frac{A^T D A}{2} \tag{33}$$

Moreover, constraint (11) must be converted into constraint (18). Consequently, when $p = 1$, the dual problem (30) associated with PSMOH simplifies to the dual problem (16) related to SMOH, incorporating constraints (17) and (18), as outlined in references [15, 16].

3.2 Second case of the proposed modification, $p = 2$

Considering $p = 2$ in (30), then:

$$PSMLL_2(A) = \sum_{i=1}^{M} \alpha_i - \frac{A^T(D+\frac{0.5}{C}I)A}{2} \tag{34}$$

where I is the identity matrix. The problem to maximize (34) considering restriction (11) and (28) is a typical problem of QP.

3.3 Third case of the proposed modification, $p = 3$

Considering $p = 3$ in (30), then:

$$PSMLL_3(A) = \sum_{i=1}^{M} \alpha_i - \frac{A^T D A}{2} - 2C \sum_{i=1}^{M} \left[\frac{\alpha_i}{3C}\right]^{\frac{3}{2}} \tag{35}$$

3.4 Forth case of the proposed modification, $p = 4$

In this case $p = 4$ in (30) resulting:

$$PSMLL_4(A) = \sum_{i}^{M} \frac{\alpha_i - A^T D A}{2} - 3C \sum_{i=1}^{M} \left[\frac{\alpha_i}{4C}\right]^{\frac{4}{3}} \tag{36}$$

Although Sections 3.3 and 3.4 introduce new functions to be maximized under constraints (11) and (28), they cannot be resolved using QP. In instances where $p > 2$, as established in (35) for $p > 3$ and (36) $p > 4$ an alternative method must be employed to solve them. For classification problems, the distance between the training examples and OH, SMOH and PSMOH can be calculated using (5). It is easy to see that these distances are different for a diverse vector, except for support ones. Due to this fact, it is very important to compare the Generalized Proximity Measure (GPM) for OH, SMOH and PSMOH.

$$GPM = \frac{1}{M} \sum_{i=1}^{M} |(G_i)| \tag{37}$$

where $|(G_i)|$is the absolute value of G_i.

4 Simulation Results

In this section the proposed modification for identifying the optimal hyperplane (PSMOH) in SVM considering $p = 2$ with the QP as optimization approach is applied to the Pima Indians Diabetes Database from the UCI Repository of Machine Learning Databases [13] and compared to the obtained results with OH, and SMOH. A dataset of 768 samples with 8 attributes is employed in this

study. These attributes are detailed in Table 1. A class value of 1 indicates a positive test result for diabetes. The class distribution comprises 500 samples classified as negative (class -1) and 268 samples classified as positive (class 1), for a total of 768 samples.

Table 1. Attributes of Pima Indians Diabetes Database used for OH, SMOH, and PSMOH methods.

Number	Attribute
1	Number of times pregnant
2	Plasma glucose concentration (2-hour oral glucose tolerance test)
3	Diastolic blood pressure (mm Hg)
4	Triceps skin fold thickness (mm)
5	2-hour serum insulin (μU/ml)
6	Body mass index (weight in kg/m^2)
7	Diabetes pedigree function
8	Age (years)
9	Class variable (-1,1)

4.1 Optimization via quadratic programming

This section presents a comparative performance analysis of the proposed modification, PSMOH, against the OH and SMOH algorithms. Sensitivity, specificity, and accuracy metrics were calculated for each approach. Furthermore, GPM of training patterns and resulting support vectors, along with constructed hyperplane, are presented.

Since PSMOH with $p = 2$ remains a QP problem, similar to OH and SMOH formulations, all three approaches are evaluated under comparable conditions. Simulation is performed using the following parameters: $N = 8$ (number of attributes), $M = 768$ (number of samples), and $C = 7680$ (soft margin weighting factor). The value of C is selected empirically, based on the number of training patterns, and used in the construction of both SMOH and PSMOH models.

4.2 Optimal Hyperplane (OH) Analysis

Results obtained for the classification of the training patterns illustrate optimization performance of (8) using the QP. In Fig. 2 a) and b) measured distance of patterns and support vectors with OH illustrates that patterns are quite far from the optimal hyperplane. In this case, OH is not adequate for non-linearly separable data presented in patterns, resulting in low accuracy of classification and all patterns are considered as support vector to construct the hyperplane, which in this case have an average distance between the support vector and hyperplane of around 180 presented in Fig. 2 c).

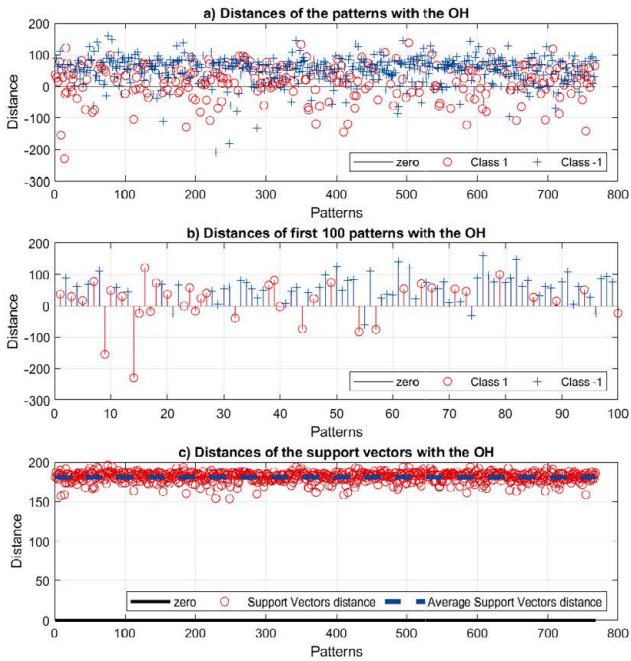

Fig. 2. a) Distances between patterns and optimal hyperplane using OH method b) Distances of the first 100 patterns with the OH, c) Distances between support vector machines and optimal hyperplane using OH method.

4.3 Soft Margin Optimal Hyperplane (SMOH) Analysis

SMOH is implemented for classification of training patterns. Contrarily to OH classification accuracy with SMOH method presents better results in performance metrics. In Figure 3 a) and b) the signed distances of all patterns and first 100 patterns with SMOH are described, where values are now smaller compared to the ones presented in the OH analysis. Additionally, support vectors for margin of the SMOH are better selected as the C parameter is adjusted for training data with distance of support vector with the SMOH and average distance presented in Figure 3 c).

4.4 Proposed Modification of the Optimal Hyperplane (PSMOH) Analysis

The PSMOH method presents better results compared to other two conventional methods. In this case, as described in the beginning of this section, $p = 2$ is

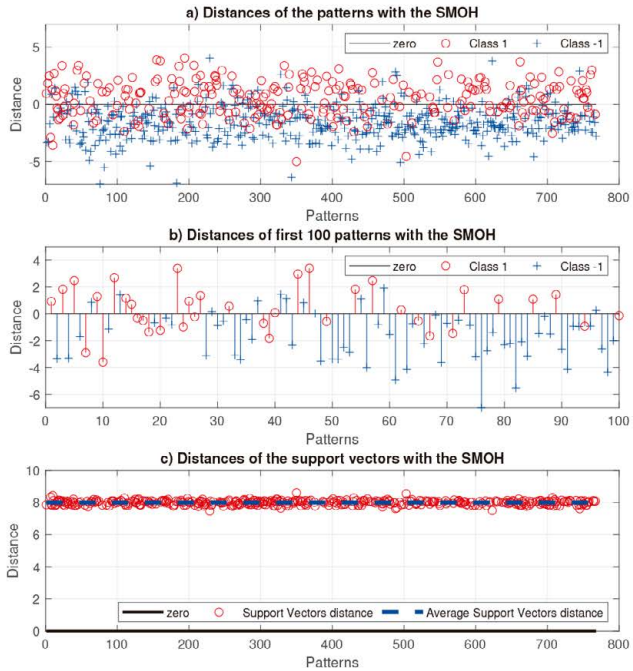

Fig. 3. a) Distances between patterns and optimal hyperplane using SMOH method. b) Distances of the first 100 patterns with the SMOH, c) Distances between support vector machines and optimal hyperplane using SMOH method

selected for (22) resulting the following function:

$$PSMF(W) = \frac{W^T W}{2} + C \sum_{i=1}^{M} e_i^2 \tag{38}$$

with the dual problem given by the objective function (34). In Figure 4 a) and b) distances of patterns with the PSMOH constructed hyperplane and support vectors distances including their average distances are illustrated. Finally, Figure 4 c) presents distances between support vectors and constructed hyperplane with PSMOH method. In Table 2, PSMOH shows superior classification performance compared to OH and SMOH. Although OH exhibits high sensitivity (69.7761%), it has low specificity (10.8%) and accuracy (31.3802%). In contrast, PSMOH achieves higher accuracy (78.2552%) with balanced sensitivity (59.3280%) and significantly improved specificity (89.20%), even surpassing SMOH in accuracy (77.3437%), sensitivity (59.3280%), and specificity (80%). This improved perfor-

mance is attributed to the construction of a more robust decision hyperplane, as evidenced in Table 3 by reduced GPM and ADSV values (1.6932 and 8.21, respectively) compared to OH (large GPM, ADSV 181.0552) and SMOH (1.6480 and 7.9920). These metrics indicate enhanced margin and class separation for PSMOH, effectively minimizing classification errors in non-linearly separable data.

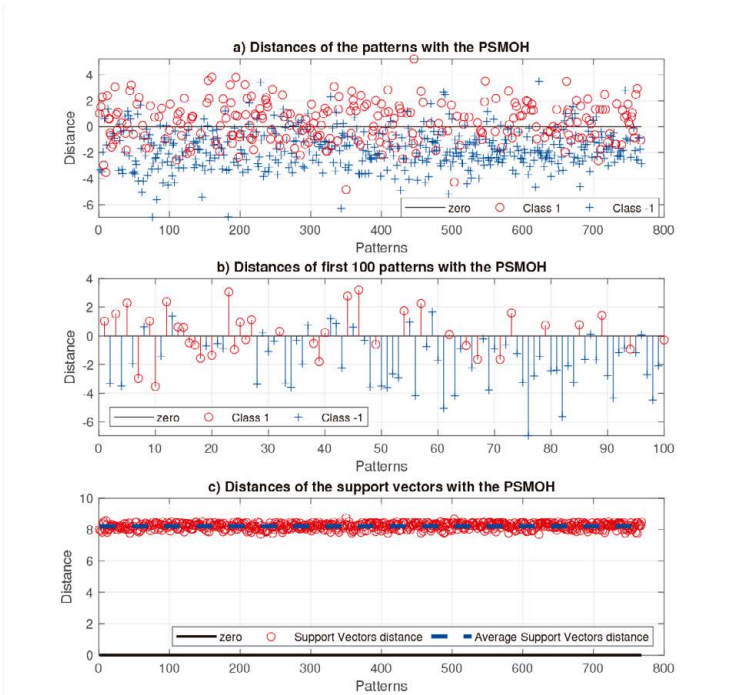

Fig. 4. a) Distances between patterns and optimal hyperplane using PSMOH method. b) Distances of the first 100 patterns with the PSMOH, c) Distances between support vector machines and optimal hyperplane using PSMOH method

5 Conclusions

Proposed modification to optimal hyperplane synthesis for support vector machines offers an advantageous alternative due to its expanded restriction domains compared to the OH and SMOH methods. This improvement stems from proposed criterion (21) and extension of restriction domain (4) to encompass

Table 2. Classification metrics for the OH, SMOH, and PSMOH methods.

Method	Sensitivity (%)	Specificity (%)	Accuracy (%)
OH	69.7761	10.8000	31.3802
SMOH	59.3280	80.0000	77.3437
PSMOH	57.8359	89.2000	78.2552

Table 3. The obtained GPM and ADSV for OH, SMOH, and PSMOH methods.

Method	GPM	ADSV
OH	54.3583	181.0556
SMOH	1.6480	7.9920
PSMOH	1.6932	8.2144

that of (13), leading to improved global optimal computing. PSMOH exhibits enhanced classification accuracy due to maximized margin during hyperplane construction, achieved by increasing the average distance of training patterns and support vectors. This superior performance and generalization, compared to OH and SMOH, results from its optimized criterion, prioritizing minimization of significant classification errors and margin maximization. Future work will explore the $p > 2$ cases using alternative optimization algorithms. Genetic and bio-inspired algorithms, which have demonstrated efficacy in minimizing and maximizing complex objective functions (as exemplified in (35) and (36)), will be considered. These investigations will encompass further functions, including potential special cases of the PSMOH method and analysis with the addition of kernels functions to enhance the classification performance of non-linearly separable data.

6 Note added

Initial analysis of this novel method has been conducted by Dr. Evgen Shelomov, whose untimely passing occurred in November 2011. This paper extends his initial analysis of the synthesis and explores its potential application to diverse areas, including pattern classification.

References

1. Abe, S., Support Vector Machines for Pattern Classification, Advances in Pattern Recognition, (2005) doi:10.1007/978-1-84996-098-4_1
2. Chandra, M.A., Bedi, S.S. Survey on SVM and their application in image classification. Int. j. inf. tecnol. 13, 1–11 (2021). doi:10.1007/s41870-017-0080-1
3. Cortes, C., Vapnik V. N., Support vector networks. Machine Learning. 20, (1995) 273-297. doi:10.1007/BF00994018
4. Lo C-S., Wang C-M. Support vector machine for breast MR image classification. Comput Math Appl 64:1153–1162 (2012) doi:10.1016/j.camwa.2012.03.033

5. Homaeinezhad, M. R., Tavakkoli, E., Atyabi, S. A., Ghaffari, A., and Ebrahimpour, R. . Synthesis of multiple-type classification algorithms for robust heart rhythm type recognition: Neuro-svm-pnn learning machine with virtual QRS image-based geometrical features. Scientia Iranica, 18(3), 423-431. (2011). doi:10.1016/j.scient. 2011.05.033.
6. Bansal, M., Goyal, A., and Choudhary, A. A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning. Decision Analytics Journal, 3, 100071. (2022). doi:10.1016/j.dajour.2022.100071
7. Zhang, H., Shi, Y., Yang, X., and Zhou, R. A firefly algorithm modified support vector machine for the credit risk assessment of supply chain finance. Research in International Business and Finance, 58, 101482.(2021). doi:10.1016/j.ribaf.2021. 101482
8. Meng, E., Huang, S., Huang, Q., Fang, W., Wang, H., Leng, G., and Liang, H. A hybrid VMD-SVM model for practical streamflow prediction using an innovative input selection framework. Water Resources Management, 35, 1321-1337. (2021). doi:10.1007/s11269-021-02786-7
9. Divya, P., and Devi, B. A. Hybrid metaheuristic algorithm enhanced support vector machine for epileptic seizure detection. Biomedical Signal Processing and Control, 78, 103841. (2022). doi:10.1016/j.bspc.2022.103841
10. Laxmi, S., Gupta, S. K., and Kumar, S. Intuitionistic fuzzy least square twin support vector machines for pattern classification. Annals of Operations Research, 339(3), 1329-1378. (2024). doi:10.1007/s10479-022-04626-2
11. Wang, H. A novel feature selection method based on quantum support vector machine. Physica Scripta, 99(5), 056006.(2024) doi:10.1088/1402-4896/ad36cf
12. Vapnik V. N., Statistical Learning Theory. (John Wiley and Sons, New York, 1998).
13. Smith, J.W., Everhart, J.E., Dickson, W.C., Knowler, W.C., and Johannes, R.S. Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. In Proceedings of the Symposium on Computer Applications and Medical Care (pp. 261–265). IEEE Computer Society Press. (1988)
14. Aoki M. Introduction to Optimization Techniques. Fundamentals and Applications of Nonlinear Programming- (McMillan, USA, 1971).
15. Haykin S., Neural Networks: A comprehensive foundation, 2nd ed. (Prentice Hall, Upper-Saddle River, NJ, 1999).
16. Vapnik V. N., The Nature of Statistical Learning Theory, 2nd ed., (Springer-Verlag, USA, 1999)