Fuzzy Support Vector Machines for Classification of Noisy Data Using Fuzzy KKT Conditions

Tsukumchila Jamir
Dept. of Science and Humanities
National Institute of technology Nagaland, India
tchilajamir1993@gmail.com

Dr. P.P. Mishra
Dept. of Science and Humanities
National Institute of Technology Nagaland, India
maths.prem79@gmail.com

Abstract— Support Vector Machines (SVMs) are effective for classification but can be sensitive to noisy data. To enhance the model's robustness to noise and uncertainty, Fuzzy Support Vector Machines (FSVMs) have been developed. The existing FSVMs predominantly focus on fuzziness in class labels while considering feature vectors as crisp, which limits their usefulness. The paper addresses this limitation by introducing FSVMs that handle fuzzy feature vectors. It also derives optimality conditions for these FSVMs using Fuzzy Karush-Kuhn-Tucker (KKT) conditions and provides a comprehensive overview of the KKT conditions for two-class SVMs and FSVMs.

Keywords—KKT optimality conditions, Fuzzy Optimization, FSVM, SVM.

I. INTRODUCTION

Support Vector Machine (SVM) is a robust Statistical Learning Theory (SLT) based supervised learning algorithm used for classification of data and regression introduced by Vapnik, Stitson et al. [5], Vapnik et al. [8]. The key objective in a two-class SVM is to determine the best hyperplane that optimizes the margin between the two classes, Tsang et al. [7]. The traditional SVM often struggle to maintain high classification accuracy in presence of outliers or noisy data. When a training data is identified as an outlier, it may be misclassified leading to significant distortion in the decision boundary. To address this outlier problem, FSVMs were introduced. In FSVM, a membership value is assigned to each data point that reflects the degree of relevance of the sample to a particular class, Lin et al. [2].

In traditional FSVMs, the primary focus is on managing fuzzy membership values related to class labels, with feature vectors treated as crisp or deterministic, Tang et al. [6], Liu[3]. This approach restricts the model's ability to handle scenarios where feature vectors themselves are inherently fuzzy. Although many studies have explored integrating fuzzy memberships into SVMs, the application of fuzzy KKT conditions to FSVMs with fuzzy feature vectors remains underexplored.

This paper addresses this research gap by presenting an overview of KKT conditions for two-class SVMs and developing fuzzy KKT conditions, Jamir and Mishra[1],Stefanini and Arana [4]. By extending the KKT framework to account for the fuzziness in feature vectors, our work aims to enhance the modeling and analysis of FSVMs in contexts where both features and classification boundaries are subject to uncertainty.

II. SVMS AS OPTIMIZATION PROBLEM

In a two-class SVM where the training data set is represented by $\{(x_j, y_j): x_j \in \mathbb{R}^n, y_j \in \{-1,1\}, j = 1,2,\dots,m\}$, the objective is to identify an optimal hyperplane that yields the maximal possible value of the margin. The margin represents the distance between the hyperplane and the support vectors called, which are the points at the closest proximity from both the class. We assume that a hyperplane with normal vector η exists that can accurately separate the two classes. The distance of the closest vector x_j^c from the hyperplane is $\frac{\langle x_j^c, \eta \rangle}{\|\eta\|}$ which is the projection of x_j^c on η. The constraints are based on the sign of $\langle x_j, \eta \rangle$. If $\langle x_j, \eta \rangle > 0$ then x_j is on the same side of the hyperplane as η and if $\langle x_j, \eta \rangle < 0$, then x_j is on the opposite side. Thus,

$$sign(\langle x_j, \eta \rangle + b) = sign(y_j)$$

$$\Rightarrow (\langle x_j, \eta \rangle + b).\, y_j \geq 0, \text{ here b is the bias}$$

By scaling η such that $\langle x_j^c, \eta \rangle = 1$ and considering the misclassification measure ξ_j the optimization model becomes-

Minimize $\frac{\|\eta\|^2}{2} + \sum_{j=1}^{m} \xi_j$

Subject to $(\langle x_j, \eta \rangle + b).\, y_j \geq 1 - \xi_j$ (1)

III. KKT CONDITIONS FOR SVMS

KKT conditions are fundamental in SVM formulation providing the necessary conditions and framework for determining the optimal hyperplane. By incorporating Lagrangian multipliers λ_j's and applying KKT conditions, we can express the dual of the problem described in equation (1) as follows:

$$\mathcal{L}(\eta, b, \lambda_j) = \frac{\|\eta\|^2}{2} + \sum_{j=1}^{m} \lambda_j \left(1 - y_j(\langle x_j, \eta \rangle + b)\right)$$

$$= \sum_{j=1}^{m} \lambda_j - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \lambda_i \lambda_j\, y_i y_j \langle x_i, x_j \rangle \quad (2)$$

Subject to $\lambda_j \geq 0, j = 1, 2, \dots, m$

$$\sum_{j=1}^{m} \lambda_j\, y_j = 0 \quad (3)$$

And we can obtain the optimal margin as-

$$\eta = \sum_{j=1}^{m} \lambda_j y_j x_j \quad (4)$$

IV. FSVMS

The conventional FSVM enhances the standard SVM framework by incorporating a fuzzy membership value for each data point, representing the confidence in the accuracy of its label. Although the data points themselves are not inherently fuzzy, this method mitigates the impact of outliers by assigning them lower membership values, thereby diminishing their impact on the decision boundary and improving the model's robustness.

Thus, in a two-class FSVM the training data set is represented by $\{(x_j, y_j, \mu_i): x_j \in \mathbb{R}^n, y_j \in \{-1,1\}, \mu_i \in [0,1],\ j = 1,2, \dots, m\}$. The objective is to identify an optimal hyperplane while minimizing the influence of outliers on the margin. This is accomplished by incorporating the membership value into the error term ξ_j. Consequently, the error term has a reduced effect on the objective function for outliers. The FSVM model is formulated as: Minimize $\frac{\|\eta\|^2}{2} + \sum_{j=1}^{m} \mu_i \xi_j$

Subject to $(\langle x_j, \eta\rangle + b).\, y_j \geq 1 - \xi_j$

$\xi_j \geq 0, j = 1, 2, \dots, m$

Let us now consider the case when each feature vector is represented by a fuzzy number.

A. Fuzzy Number

A fuzzy number $\tilde{z}$ is a fuzzy set characterized by a membership function that is normal and satisfies the additional conditions of upper semi-continuity and convexity. Furthermore, for every $\alpha \epsilon [0,1]$, the $\alpha - level$ set of $\tilde{z}$ is compact

If $\mathcal{F}$ is the set of all fuzzy numbers then, for any $z \epsilon \mathcal{F}$, $[\tilde{z}]^{\alpha} = [\underline{\tilde{z}_\alpha}, \overline{\tilde{z}_\alpha}\ ]$.

In this case, the training data set is represented by $\{(\widetilde{x_j}, y_j, \mu_i): \widetilde{x_j} \in \mathcal{F}, y_j \in \{-1,1\}, \mu_i \in [0,1],$ $j = 1,2, \dots, m\}$. And the FSVM model becomes-

Minimize $\frac{\|\eta\|^2}{2} + \sum_{j=1}^{m} \mu_i \xi_j$

Subject to $(\langle \widetilde{x_j}, \eta\rangle + b).\, y_j \geq 1 - \xi_j$

$\xi_j \geq 0, j = 1, 2, \dots, m$

V. FUZZY KKT CONDITIONS FOR FSVMS

B. Theorem

A totally differentiable fuzzy objective function $\widetilde{f}$: $\mathbb{R}^n \to \mathcal{F}_Q$ with differentiable constraint functions $g_j's$ has a non-inferionsolution $x_I{}^*$ if $\exists \lambda_j(\alpha)$, $j = 1,2, \dots, m$ such that the following conditions are met for every $\alpha \epsilon [0,1]$ and $j = 1,2, \dots, m$ -

i) $\nabla(\underline{\tilde{f}_\alpha} + \overline{\tilde{f}_\alpha})(\,x_I{}^*\,) - \sum_{j=1}^{m} \lambda_j(\alpha) \nabla g_j(x_I{}^*\,) = 0$

ii) $\lambda_j\,(\alpha) g_j(x_I{}^*\,) = 0$

iii) $\lambda_j(\alpha) \geq 0$

Using the above Theorem,

$$\tilde{\mathcal{L}}(\eta, b, \lambda_{1j}, \lambda_{2j}, \xi_j) = \frac{\|\eta\|^2}{2} + \sum_{j=1}^{m} \mu_i \xi_j - \sum_{j=1}^{m} \lambda_{1j} \left(1 - \xi_j - y_j(\langle \widetilde{x_j}, \eta\rangle + b)\right) - \sum_{j=1}^{m} \lambda_{2j}\, \xi_j$$

ISBN : 978-81-978522-2-0

$$= \frac{\|\eta\|^2}{2} + \sum_{j=1}^{m} \mu_i \xi_j - \sum_{j=1}^{m} \lambda_{1j} \left(1 - \xi_j - y_j \left(\langle \left[\underline{\tilde{x}_{J_\alpha}}, \overline{\tilde{x}_{J_\alpha}}\right], \eta \rangle + b\right)\right) - \sum_{j=1}^{m} \lambda_{2j} \xi_j$$

$$= \frac{\|\eta\|^2}{2} + \sum_{j=1}^{m} \mu_i \xi_j - \sum_{j=1}^{m} \lambda_{1j} \left(1 - \xi_j - y_j \left(\left[\langle \underline{\tilde{x}_{J_\alpha}}, \eta \rangle, \langle \overline{\tilde{x}_{J_\alpha}}, \eta \rangle\right] + b\right)\right) - \sum_{j=1}^{m} \lambda_{2j} \xi_j \quad (5)$$

Thus,

$$\frac{\partial \tilde{\mathcal{L}}(\eta, b, \lambda_{1j}, \lambda_{2j}, \xi_j)}{\partial b} = 0 \Rightarrow \sum_{j=1}^{m} \lambda_{1j} y_j = 0$$

$$\frac{\partial \tilde{\mathcal{L}}(\eta, b, \lambda_{1j}, \lambda_{2j}, \xi_j)}{\partial \eta_i} = 0 \qquad \Rightarrow \eta_i -$$

$$\sum_{j=1}^{m} y_j \lambda_j \left[\langle \underline{\widetilde{x_{Ji}}_\alpha}, \eta_i \rangle, \langle \overline{\widetilde{x_{Ji}}_\alpha}, \eta_i \rangle\right] 0, here\ x_{ji}\ represent\ i^{th}\ component\ of\ x_j$$

$$\Rightarrow \eta = \sum_{j=1}^{m} y_j \lambda_j \left[\underline{\tilde{x}_{J_\alpha}}, \overline{\tilde{x}_{J_\alpha}}\right]$$

And, $\frac{\partial \tilde{\mathcal{L}}(\eta, b, \lambda_{1j}, \lambda_{2j}, \xi_j)}{\partial \xi_j} = 0 \Rightarrow \sum_{j=1}^{m} \mu_i + \lambda_{1j} - \lambda_{2j} = 0$

Applying this to (5) we get,

Maximize $\tilde{\mathcal{L}}(\eta, b, \lambda_{1j}, \lambda_{2j}, \xi_j)$

$$= -\left\{\frac{1}{2} \sum_{j=1}^{m} y_j \lambda_j \left[\underline{\tilde{x}_{J_\alpha}}, \overline{\tilde{x}_{J_\alpha}}\right] + \sum_{j=1}^{m} \mu_i \xi_j - \sum_{j=1}^{m} \lambda_{1j}(1 - \xi_j) - \sum_{j=1}^{m} \lambda_{2j} \xi_j\right\}$$

$$= \sum_{j=1}^{m} \lambda_{1j} - \frac{1}{2} \sum_{j=1}^{m} y_j \lambda_j \left[\underline{\tilde{x}_{J_\alpha}}, \overline{\tilde{x}_{J_\alpha}}\right]$$

$$= \sum_{j=1}^{m} \lambda_{1j} - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} y_i y_j \lambda_i \lambda_j \left[\underline{\tilde{x}_{\iota_\alpha}}\, \underline{\tilde{x}_{J_\alpha}}, \overline{\tilde{x}_{\iota_\alpha}}\ \overline{\tilde{x}_{J_\alpha}}\right]$$

Subject to $\lambda_{1j} \geq 0, j = 1, 2, \ldots, m$

$\sum_{j=1}^{m} \lambda_{1j}\, y_j = 0$

Thus for optimal ${\lambda_{1j}}^{*}$ we can obtain the optimal hyperplane using-

$$\eta^{*} = \sum_{j=1}^{m} y_j \lambda_j \left[\underline{\tilde{x}_{J_\alpha}}, \overline{\tilde{x}_{J_\alpha}}\right]$$

VI. CONCLUSION

This paper advances Fuzzy Support Vector Machines (FSVMs) by incorporating fuzzy Karush-Kuhn-Tucker (KKT) conditions for feature vectors. Our approach enhances FSVMs by integrating fuzziness into both feature vectors and classification boundaries, providing a more robust framework for handling uncertain and imprecise data. This work paves the way for future research and practical applications in domains where both features and labels exhibit inherent fuzziness.

ISBN : 978-81-978522-2-0

REFERENCES

[1] Jamir, T., Mishra, P.P. (2024).KKT optimality conditions for fuzzy optimization problems using H1 and H2 derivative. *OPSEARCH* , https://doi.org/10.1007/s12597-024-00786-w.

[2] Jamir, T., Mishra, P.P. (2024).KKT optimality conditions for fuzzy optimization problems using H1 and H2 derivative. *OPSEARCH* , https://doi.org/10.1007/s12597-024-00786-w.

[3] Liu, J. (2021). Fuzzy support vector machine for imbalanced data with borderline noise. *Fuzzy sets and systems*, *413*, 64-73. https://doi.org/10.1016/j.fss.2020.07.018.

[4] Stefanini, L., & Arana-Jiménez, M. (2019). Karush–Kuhn–Tucker conditions for interval and fuzzy optimization in several variables under total and directional generalized differentiability. *Fuzzy Sets and Systems*, *362*, 1-34. https://doi.org/10.1016/j.fss.2018.04.009.

[5] Stitson, M. O., Weston, J. A. E., Gammerman, A., Vovk, V., & Vapnik, V. (1996). Theory of support vector machines. University of London, 117(827), 188-191.

[6] Tang, W. M. (2011). Fuzzy SVM with a new fuzzy membership function to solve the two-class problems. *Neural processing letters*, *34*, 209-219. https://doi.org/10.1007/s11063-011-9192-y.

[7] Tsang E. C. C., Yeung D. S. and Chan P. P. K., (2003). Fuzzy support vector machines for solving two-class problems,*Proceedings of the 2003 International Conference on Machine Learning and Cybernetics, IEEE*, 2, 1080-1083. https://doi.org/ doi: 10.1109/ICMLC.2003.1259643.

[8] Vapnik, V.N., Chervonenkis, A.Y. (2015). On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities. In: Vovk, V., Papadopoulos, H., Gammerman, A. (eds) Measures of Complexity. *Springer*. https://doi.org/10.1007/978-3-319-21852-6_3