

Research Article

The Generalization Error Bound for the Multiclass Analytical Center Classifier

Zeng Fanzi and Ma Xiaolong

Key Laboratory for Embedded and Network Computing of Hunan Province, Hunan University, Changsha 410082, China

Correspondence should be addressed to Zeng Fanzi; zengfanzi@126.com

Received 29 October 2013; Accepted 5 December 2013

Academic Editors: J. Shu and F. Yu

Copyright © 2013 Z. Fanzi and M. Xiaolong. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents the multiclass classifier based on analytical center of feasible space (MACM). This multiclass classifier is formulated as quadratic constrained linear optimization and does not need repeatedly constructing classifiers to separate a single class from all the others. Its generalization error upper bound is proved theoretically. The experiments on benchmark datasets validate the generalization performance of MACM.

1. Introduction

Multiclass classification is an important and on-going research subject in machine learning. Its application is immense, such as machine vision [1, 2], text and speech categorization [3, 4], natural language processing [5], and disease diagnosis [6, 7]. Two kinds of approaches have been proposed to solve multiclass classification problem [8]. The first multiclass classification approach is extending binary classifier to handle the multiclass case directly. This included neural networks, decision trees, support vector machines, naive Bayes, and K -nearest neighbors. The second approach decomposes the multiclass classification problem into several binary classification tasks. Several methods are used for this decomposition: one-versus-all [9], all-versus-all [10], and error-correcting output coding [11].

The one-versus-all approach reduces the problem of classifying among K classes into K binary problems, where each problem discriminates a given class from the other $K - 1$ classes.

For the all-versus-all method, a binary classifier is built to discriminate between each pair of classes, while discarding the rest of the classes. This requires building $K(K-1)/2$ binary classifiers for K classes problem. When testing a new example, voting is performed among the classifiers and the class with the maximum number of votes wins.

For error-correcting output coding, it works by training N binary classifiers to distinguish between the K different

classes. Each class is given a codeword of length N according to a binary matrix M . Each row of M corresponds to a certain class.

The above multiclass classification algorithms need construct binary classifier repeatedly to separate a single class from all the others for K classes problem, which leads to daunting computation and low efficiency of classification. Reference [12] proposes multiclass support vector machine (MSVM), which corresponds to simple quadratic optimization and need not repeat constructing binary classifier. However, support vector machine corresponds to the center of the largest inscribed hypersphere of feasible space. When the feasible space, that is, the space of hypotheses consistent with the training data, is elongated or asymmetric, support vector machine is not effective [13]. To address the above problems, multiclass classifier based on the analytical center of feasible space (MACM) is proposed. At the same time, in order to validate its generalization performance theoretically, its generalization error upper bound is formulated and proved. And the experiments on benchmark dataset validate the generalization performance of MACM.

2. Multiclass Analytical Center Classifier

To facilitate the discussion of multiclass analytical center classifier, the following definitions are introduced.

Definition 1 (chunk). A vector, $\mathbf{v} = (v_1, \dots, v_{kd}) \in \mathfrak{R}^{kd}$, is broken into k chunks $(\mathbf{v}_1, \dots, \mathbf{v}_k)$ where the i th chunk $\mathbf{v}_i = (v_{(i-1)*d+1}, \dots, v_{i*d})$.

Definition 2 (expansion). Let $\text{Vec}(\mathbf{x}, i) \in \mathfrak{R}^{kd}$ be a vector where $\mathbf{x} \in \mathfrak{R}^d$ is embedded in kd dimensions space by writing the coordinates of \mathbf{x} in the i th chunk of a vector in \mathfrak{R}^{kd} . $\mathbf{0}^\ell$ denotes the zero vector of length ℓ . Then, $\text{Vec}(\mathbf{x}, i)$ can be written formally as the concatenation of three vectors, $\text{Vec}(\mathbf{x}, i) = (\mathbf{0}^{(i-1)*d}, \mathbf{x}, \mathbf{0}^{(k-i)*d}) \in \mathfrak{R}^{kd}$. And define $\text{Vec}(\mathbf{x}, i, j) = \text{Vec}(\mathbf{x}, i) - \text{Vec}(\mathbf{x}, j)$ as a vector where \mathbf{x} is embedded in the i th chunk and $-\mathbf{x}$ is embedded in the j th chunk of a vector in \mathfrak{R}^{kd} .

Definition 3. Given the sample $(\mathbf{x}, y) \in \mathfrak{R}^d \times \{1, \dots, k\}$, its expansion is defined as $G(\mathbf{x}, y) = \{\text{Vec}(\mathbf{x}, i, j) \mid i = y, j = \{1, \dots, k\}/i\}$; the expansion of the whole sample set $S = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots\}$ is defined as $G(S) = \bigcup_{(\mathbf{x}, y) \in S} G(\mathbf{x}, y)$.

Definition 4 (piecewise linear separability). The point sets $A^i \in \mathfrak{R}^{d \times m_i}$, $i = 1, \dots, k$ (k represents the class labels and m_i the number of samples belonging to i th class), are piecewise linear separable if there exists $\mathbf{w}^i \in \mathfrak{R}^d$, $\gamma^i \in \mathfrak{R}$, $i = 1, \dots, k$, where d represents the dimension of point, such that

$$A_\ell^i \mathbf{w}^i - \gamma^i > A_\ell^i \mathbf{w}^j - \gamma^j, \quad (1)$$

$$i, j = 1, \dots, k, \quad i \neq j, \quad \ell = 1, \dots, m_i.$$

Definition 5 (piecewise linear classifier). Assume $\mathbf{w} = (\mathbf{w}_1, \dots, \mathbf{w}_k)$, where $(\mathbf{w}_1, \dots, \mathbf{w}_k) \in \mathfrak{R}^{k*(d+1)} = \mathfrak{R}^{d+1} \times \dots \times \mathfrak{R}^{d+1}$. Given a new point $\mathbf{x} \in \mathfrak{R}^{d+1}$, a piecewise linear classifier is a function $f: \mathfrak{R}^{d+1} \rightarrow \{1, \dots, k\}$ as follows:

$$f(\mathbf{x}) = \arg \max_{i=1, \dots, k} \mathbf{w}_i \mathbf{x}', \quad (2)$$

where $\arg \max$ returns to a class label corresponding to the maximum value.

To simplify the notation for the formulation of multiclass analytical center classifier, we consider an augmented weight space as follows.

Let

$$\tilde{A}_\ell^i = \begin{bmatrix} A_\ell^i \\ 1 \end{bmatrix} \in \mathfrak{R}^{d+1}, \quad \tilde{\mathbf{w}}^i = \begin{bmatrix} \mathbf{w}^i \\ \gamma \end{bmatrix} \in \mathfrak{R}^{d+1}, \quad (3)$$

$$i = 1, \dots, k, \quad \ell = 1, \dots, m_i;$$

then, inequality (1) can be rewritten as

$$\tilde{A}_\ell^i \tilde{\mathbf{w}}^i - \tilde{A}_\ell^j \tilde{\mathbf{w}}^j > 0, \quad i, j = 1, \dots, k, \quad i \neq j, \quad \ell = 1, \dots, m_i. \quad (4)$$

Let $\tilde{\mathbf{w}} = (\tilde{\mathbf{w}}_1, \dots, \tilde{\mathbf{w}}_k) \in \mathfrak{R}^{k*(d+1)} = \mathfrak{R}^{d+1} \times \dots \times \mathfrak{R}^{d+1}$. According to Definition 2, embedding \tilde{A}_ℓ^i into $\mathfrak{R}^{k*(d+1)}$ space, inequality (4) has the following form:

$$(\text{Vec}(\tilde{A}_\ell^i, i) - \text{Vec}(\tilde{A}_\ell^j, j)) \tilde{\mathbf{w}} > 0, \quad (5)$$

$$i, j = 1, \dots, k, \quad i \neq j, \quad \ell = 1, \dots, m_i.$$

Consider that

$$\text{Vec}(\tilde{A}_\ell^i, i, j) = \text{Vec}(\tilde{A}_\ell^i, i) - \text{Vec}(\tilde{A}_\ell^j, j), \quad (6)$$

$$i, j = 1, \dots, k, \quad i \neq j, \quad \ell = 1, \dots, m_i.$$

Thus, inequality (6) can be rewritten as follows:

$$\text{Vec}(\tilde{A}_\ell^i, i, j) \tilde{\mathbf{w}} > 0, \quad i, j = 1, \dots, k, \quad i \neq j, \quad \ell = 1, \dots, m_i. \quad (7)$$

Inequality (7) represents the feasible space of $\tilde{\mathbf{w}}$ in the higher dimension space $\mathfrak{R}^{k*(d+1)}$. Similar to the binary classification based on the analytical center of version space [8], we define the slack variable $S_{i,j,\ell} = \text{Vec}(\tilde{A}_\ell^i, i, j) \tilde{\mathbf{w}}$, $i, j = 1, \dots, k$, $i \neq j$, $\ell = 1, \dots, m_i$ and then have the following minimization problem, whose solver corresponds to the analytical center of higher dimension space:

$$\min \Phi(\tilde{\mathbf{w}}) = - \sum_{\ell=1}^{m_i} \sum_{i \neq j, i, j=1}^k \ln S_{i,j,\ell} \quad (8)$$

$$\text{s.t. } h(\tilde{\mathbf{w}}) = \frac{1}{2} \tilde{\mathbf{w}} \tilde{\mathbf{w}}' - 1 = 0.$$

In order to further simplify the formulation of multiclass analytical center classifier, we introduce some notations as follows: $M = k * (k-1) * \sum_{i=1}^k m_i$, $B = \{\text{Vec}(\tilde{A}_\ell^i, i, j) \mid i, j = 1, \dots, k, i \neq j, \ell = 1, \dots, m_i\} \in \mathfrak{R}^{M \times k*(d+1)}$; let B_i represent the i th row vector of B . Then, the optimization problem (8) can be rewritten as follows:

$$\min \Phi(\tilde{\mathbf{w}}) = - \sum_{i=1}^M \ln S_i \quad \text{where } S_i = B_i \tilde{\mathbf{w}} \quad (9)$$

$$\text{s.t. } h(\tilde{\mathbf{w}}) = \frac{1}{2} \tilde{\mathbf{w}} \tilde{\mathbf{w}}' - 1 = 0.$$

After solving the optimization problem (9) to get the optimal weight $\tilde{\mathbf{w}}^*$, we have a piecewise linear classifier $f: \mathfrak{R}^{d+1} \rightarrow \{1, \dots, k\}$ computed in the following way:

$$f(\mathbf{x}) = \arg \max_{i=1, \dots, k} \tilde{\mathbf{w}}_i^* \mathbf{x}', \quad (10)$$

where $\arg \max$ returns to a class label corresponding to the maximum value.

If the dataset is not piecewise linear separable, the kernel function is used to map the data into high dimension linear space.

3. Generalization Error Bound of Multiclass Analytical Center Classifier

In order to analyze the generalization error bound theoretically, we introduce the definition of classification margin and data radius and then deduce the margin-based generalization error bound of MACM.

Definition 6 (classification margin). Given the linear classifier $h(\mathbf{x}_j) = \arg \max_{i=1, \dots, k} \bar{\mathbf{w}}_i \bar{\mathbf{x}}_j'$, the classification margin of the sample $(\mathbf{x}_j, y_j) \in \mathfrak{R}^d \times \{1, \dots, k\}$ is defined as follows:

$$r_j = \min_{l=y_j, i=\{1, \dots, k\}/l} (\bar{\mathbf{w}}_l \bar{\mathbf{x}}_j' - \bar{\mathbf{w}}_i \bar{\mathbf{x}}_j'). \quad (11)$$

For the whole training set $S = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_m, y_m)\}$, the minimal margin is as follows:

$$\text{margin}_S(h) = \min_{(\mathbf{x}_j, y_j) \in S} r_j. \quad (12)$$

Definition 7 (data radius). Given dataset $S = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_m, y_m)\}$, the data radius is defined as follows:

$$\zeta(S) = \max_{\mathbf{x}_i \in S} \|\mathbf{x}_i\|. \quad (13)$$

Theorem 8. Define data radius of dataset S as $\zeta(S)$ and data radius of dataset $\text{Vec}(S, i, j)$ as $\zeta(\text{Vec}(S, i, j))$; if $\zeta(S) \leq R$, then $\zeta(\text{Vec}(S, i, j)) \leq 2R$.

Proof. Consider the following:

$$\begin{aligned} & \zeta(\text{Vec}(S, i, j)) \\ &= \max_{\mathbf{x}_i \in S} \|\text{Vec}(\mathbf{x}_i, i, j)\| \\ &= \max_{\mathbf{x}_i \in S} \|\text{Vec}(\mathbf{x}_i, i) - \text{Vec}(\mathbf{x}_i, j)\| \\ &\leq \max_{\mathbf{x}_i \in S} (\|\text{Vec}(\mathbf{x}_i, i)\| + \|\text{Vec}(\mathbf{x}_i, j)\|) \\ &= \max_{\mathbf{x}_i \in S} \|\text{Vec}(\mathbf{x}_i, i)\| + \max_{\mathbf{x}_i \in S} \|\text{Vec}(\mathbf{x}_i, j)\|. \end{aligned} \quad (14)$$

Because $\|\text{Vec}(\mathbf{x}, i)\| = \|(\mathbf{0}, \mathbf{x}, \mathbf{0})\| = \|\mathbf{x}\|$, $\zeta(\text{Vec}(S, i, j)) \leq 2 * \max_{\mathbf{x}_i \in S} \|\mathbf{x}_i\| = 2R$. This ends the proof of Theorem 8. \square

Theorem 9 (see [14]). Consider thresholding of a real-valued function \mathcal{H} with unit weight vectors $\|\mathbf{w}\| = 1$ on the inner product space X and fix margin $r \in \mathfrak{R}^+$. For any probability distribution \mathcal{D} on $X \times \{-1, 1\}$ with support in a ball of radius R around the origin, with probability $1 - \delta$ over m random samples S , any hypothesis $h \in \mathcal{H}$ with $\text{margin}_S(h) \geq r$ on S has error more than

$$\begin{aligned} \text{err}_{\mathcal{D}}(h) &= \varepsilon(r, m, \delta, R) \\ &= \frac{2}{m} \left(\frac{64R^2}{r^2} \log \frac{emr}{8R^2} \log \frac{32m}{r^2} + \log \frac{4}{\delta} \right), \end{aligned} \quad (15)$$

provided $m > 2/\varepsilon, 64R^2/r^2 < m$.

From Definition 3 and inequality (7), it is shown that to correctly classify the sample $A^i, i = 1, \dots, k$, $G(\bar{A})\bar{\mathbf{w}} = G(\bigcup_{i=1}^k \bar{A}^i)\bar{\mathbf{w}} > \mathbf{0}$ is satisfied. Here, one introduces the samples' pairs $P_+(A) = \{G(\bar{A}), +\mathbf{1}\} \in \mathfrak{R}^{k*(d+1)} \times \{+1\}$ and $P_-(A) = \{-G(\bar{A}), -\mathbf{1}\} \in \mathfrak{R}^{k*(d+1)} \times \{-1\}$, where $\mathbf{0}, \mathbf{1}$ denote the corresponding dimension vector with elements 0 or 1, respectively. So, one can construct the new samples' set $P(A) = P_+(A) \cup P_-(A) \in \mathfrak{R}^{k*(d+1)} \times \{-1, +1\}$.

Theorem 10. The binary classification of sample $P(A)$ by analytical center classifier is equivalent to the multiclass classification of sample $A = \{A^i\}_{i=1}^k$ by multiclass analytical center classifier.

Proof. Assume that $\{X_i^+, Y^+\} \in P_+(A), \{X_i^-, Y^-\} \in P_-(A), i = 1, \dots, |P_+(A)|$; then, binary classification is to solve the following feasible problem:

$$\begin{aligned} Y^+ (X_i^+ \mathbf{w} + b) &> 0, \quad i = 1, \dots, |P_+(A)|, \\ Y^- (X_i^- \mathbf{w} + b) &> 0, \quad i = 1, \dots, |P_-(A)|. \end{aligned} \quad (16)$$

Suppose the bias b equals 0, because $P_-(A)$ and $P_+(A)$ are symmetrical on the origin. The feasible constraints can be rewritten as follows:

$$X_i^+ \mathbf{w} > 0, \quad i = 1, \dots, |P_+(A)|. \quad (17)$$

The feasible constraints (17) define the feasible space of weight vector $\mathbf{w} \in \mathfrak{R}^{k*(d+1)}$; the binary classification by analytical center classifier can be formulated as follows:

$$\begin{aligned} \min \Phi(\mathbf{w}) &= - \sum_{i=1}^{|P_+(A)|} X_i^+ \mathbf{w} > 0 \\ \text{s.t.} \quad & \frac{1}{2} \mathbf{w}' \mathbf{w} - 1 = 0. \end{aligned} \quad (18)$$

Because $X_i^+ \in G(A)$ and $|P_+(A)| = |G(A)|$, the problem (18) is equivalent to problem (8). This ends the proof of Theorem 10. \square

Theorem 11. Consider the classifiers' set \mathcal{H} from Definition 5 with $\sum_{i=1, \dots, k} \|\mathbf{w}_i\| = 1$ on the inner product space X , where $h : \mathfrak{R}^{d+1} \rightarrow \{1, \dots, k\}$ and fix margin $r \in \mathfrak{R}^+$. For any probability distribution \mathcal{D} on $X \times \{1, \dots, k\}$ with support in a ball of radius R around the origin, with probability $1 - \delta$ over m random samples S , any hypothesis $h \in \mathcal{H}$ with $\text{margin}_S(h) \geq r$ on S has error more than

$$\begin{aligned} \text{err}_{\mathcal{D}}(h) &= \varepsilon(r, m, \delta, R, k) \\ &= \frac{4(k-1)}{m} \left(\frac{256R^2}{r^2} \log \frac{emr}{32R^2} \log \frac{32m}{r^2} + \log \frac{4}{\delta} \right) \end{aligned} \quad (19)$$

provided $m > 2/\varepsilon, 64R^2/r^2 < m$.

Proof. Because the sample in $P(S)$ is not independent, the generalization error bound cannot be attained from Theorem 9. Theorem 9 is independent of the sample distribution, so we can construct a new sample distribution \mathcal{D}' . According to the new distribution and dataset S to generate the independent sample set $P'(S)$ with m samples, that is, for every $(\mathbf{x}, y) \in S$, define $P'(\mathbf{x}, y)$ as the point sampled uniformly and randomly from $P(S)$ according to the distribution \mathcal{D}' ; then, we have $P'(S) = \bigcup_{(\mathbf{x}, y) \in S} P'(\mathbf{x}, y)$. From Theorem 8, the data radius $\zeta(P'(S))$ of $P'(S)$ satisfies

$\zeta(P^f(S)) \leq 2R$. The generalization error of hypothesis h over $P^f(S)$ from Theorem 9 can be calculated as follows:

$$\begin{aligned} \text{err}_{\mathcal{D}'}(h) &= \varepsilon(r, m, \delta, R) \\ &= \frac{2}{m} \left(\frac{256R^2}{r^2} \log \frac{\text{emr}}{32R^2} \log \frac{32m}{r^2} + \log \frac{4}{\delta} \right). \end{aligned} \quad (20)$$

Event E_i which denotes a sample in $P(S)$ is wrongly classified and event C which denotes the misclassification occurs in $P^f(S)$. From the above analysis, the misclassification of any sample in $P(S)$ causes the misclassification of the point in $P^f(S)$, so that the probability of events E_i and C satisfies the following inequality:

$$P(E_i) \leq P(C). \quad (21)$$

Because the cardinality of $P(S)$ equals $2(k-1)$, the probability of sample misclassification in $P(S)$ is written as follows:

$$P\left(\bigcup_{i=1}^{2(k-1)} E_i\right). \quad (22)$$

From union bound theorem, we have the following inequality:

$$P\left(\bigcup_{i=1}^{2(k-1)} E_i\right) \leq \sum_{i=1}^{2(k-1)} P(E_i) \leq \sum_{i=1}^{2(k-1)} P(C). \quad (23)$$

So the generalization error of hypothesis h over S is

$$\begin{aligned} \text{err}_{\mathcal{D}}(h) &= \varepsilon(r, m, \delta, R, k) \\ &= \sum_{i=1}^{2(k-1)} \frac{2}{m} \left(\frac{256R^2}{r^2} \log \frac{\text{emr}}{32R^2} \log \frac{32m}{r^2} + \log \frac{4}{\delta} \right) \\ &= \frac{4(k-1)}{m} \left(\frac{256R^2}{r^2} \log \frac{\text{emr}}{32R^2} \log \frac{32m}{r^2} + \log \frac{4}{\delta} \right). \end{aligned} \quad (24)$$

This ends the proof of Theorem 11. \square

4. Computational Experiments

In this section, we present the computational results comparing multiclass analytical center classifier (MACM) and multiclass support vector machine (MSVM) [12]. A description of each of the datasets follows this paragraph. The kernel function for the piecewise nonlinear MACM and MSVM methods is $k(\mathbf{x}, \mathbf{x}_i) = (\mathbf{x}\mathbf{x}_i^t/N + 1)^n$, where n is the desired polynomial.

Wine Recognition Data. The wine dataset uses the chemical analysis of wine to determine the cultivar. There are 178 points with 13 features. This is a three class dataset distributed as follows: 59 points in class 1, 71 points in class 2, and 48 points in class 3.

Glass Identification Database. The Glass dataset is used to identify the origin of a sample of glass through chemical

TABLE 1: The generalization error of MACM and MSVM.

Dataset	Classifier	Degree of polynomial	
		1	3
Wine	M-ACM	97.74	98.65
	M-SVM	97.19	97.75
Glass	M-ACM	56.46	69.38
	M-SVM	55.14	66.15

analysis. This dataset is comprised of six classes of 214 points with 9 features. The distribution of points by class is as follows: 70 float processed building windows, 17 float processed vehicle windows, 76 nonfloat processed building windows, 13 containers, 9 tableware, and 29 headlamps.

Table 1 contains the results for MACM and MSVM on wine and glass datasets. As anticipated, MACM produces better testing generalization than MSVM.

5. Summary

In this paper, the multiclass analytical center classifier based on the analytical center of feasible space, which corresponds to a simple quadratic constrained linear optimization, is proposed. At the same time, in order to validate its generalization performance theoretically, its generalization error upper bound is formulated and proved. By the experiments on wine recognition and glass identification dataset, it is shown that the multiclass analytical center classifier outperforms multiclass support vector machine in generalization error.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant nos. 61370096 and 61173012 and the Key Project of Natural Science Foundation of Hunan Province under Grant no. 12JJA005.

References

- [1] J. S. Prakash, K. A. Vignesh, C. Ashok, and R. Adithyan, "Multi-class support vector machines classifier for machine vision application," in *Proceedings of the International Conference on Machine Vision and Image Processing*, pp. 197–199, 2013.
- [2] A. J. Joshi, F. Porikli, and N. P. Papanikolopoulos, "Scalable active learning for multiclass image classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2259–2273, 2012.
- [3] S. Ai-Xiang, L. Ming-Hui, H. Shun-Liang, and Z. Jun, "A new hypersphere multi-class support vector machine applied in text classification," in *Proceedings of the IEEE 3rd International Conference on Communication Software and Networks (ICCSN '11)*, pp. 478–481, Xi'an, China, May 2011.
- [4] L. La, Q. Guo, D. Yang, and Q. Cao, "Multiclass boosting with adaptive group-based kNN and its application in text categorization," *Mathematical Problems in Engineering*, vol. 2012, Article ID 793490, 24 pages, 2012.
- [5] Y. Ni, C. J. Saunders, S. Szedmak, and M. Niranjana, "Structure learning for natural language processing," in *Proceedings of the*

- IEEE International Workshop on Machine Learning for Signal Processing (MLSP '09)*, pp. 1–6, September 2009.
- [6] H. Ankişhan and D. Yılmaz, “Comparison of SVM and ANFIS for snore related sounds classification by using the largest Lyapunov exponent and entropy,” *Computational and Mathematical Methods in Medicine*, vol. 2013, Article ID 238937, 13 pages, 2013.
- [7] P. Adarsh and D. Jeyakumari, “Multiclass SVM-based automated diagnosis of diabetic retinopathy,” in *Proceedings of the International Conference on Communication and Signal Processing (ICCSP '13)*, pp. 206–210, Melmaruvathur, India, April 2013.
- [8] M. Aly, *Survey on Multiclass Classification Methods*, 2005.
- [9] A. O. Hatch and A. Stolcke, “Generalized linear kernels for one-versus-all classification: application to speaker recognition,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '06)*, Toulouse, France, May 2006.
- [10] M. A. Bagheri, G. A. Montazer, and S. Escalera, “Error correcting output codes for multiclass classification: application to two image vision problems,” in *Proceedings of the 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP '12)*, pp. 508–513, Shiraz, Iran, May 2012.
- [11] C. W. Hsu and C. J. Lin, “A comparison of methods for multiclass support vector machines,” *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002.
- [12] E. J. Bredensteiner and K. P. Bennett, “Multicategory classification by support vector machines,” *Computational Optimization and Applications*, vol. 12, no. 1–3, pp. 53–79, 1999.
- [13] T. B. Trafalis and A. M. Malysheff, “An analytic center machine,” *Machine Learning*, vol. 46, no. 1–3, pp. 203–223, 2002.
- [14] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, New York, NY, USA, 2000.