

Enclosing machine learning: concepts and algorithms

Xun-Kai Wei · Ying-Hong Li · Yu-Fei Li ·
Dong-Fang Zhang

Received: 8 October 2006 / Accepted: 30 March 2007 / Published online: 3 May 2007
© Springer-Verlag London Limited 2007

Abstract A novel machine learning paradigm, i.e., enclosing machine learning based on regular geometric shapes was proposed. First, it adopted regular minimum volume enclosing and bounding geometric shapes (sphere, ellipsoid, box) or their unions and so on to obtain one class description model. Second, Data description, two class classification, learning algorithms based on the one class description model were presented. The most obvious feature was that enclosing machine learning emphasized one class description and learning. To illustrate the concepts and algorithms, a minimum volume enclosing ellipsoid (MVEE) case for enclosing machine learning was then investigated in detail. Implementation algorithms for enclosing machine learning based on MVEE were presented. Subsequently, we validate the performances of MVEE learners using real world datasets. For novelty detection, a benchmark ball bearing dataset is adopted. For pattern classification, a benchmark iris dataset is investigated. The performance results show that our proposed method is comparable even better than Support Vector Machines (SVMs) in the datasets studied.

Keywords Cognitive process · Enclosing machine learning · Minimum volume enclosing shapes · Minimum volume enclosing ellipsoid · Cognitive class description · Cognitive class recognizing

1 Introduction

Cognitive process is the instinctive learning ability of human being. Human always transfers the feature information to the brain through perception, and then the brain will process the feature information and remember the feature information for the given objects. According to the cognitive science theory, the human brain can be imitated but can not be completely reproduced. Currently, artificial intelligence is an important direction of function imitation of the human brain.

Neural-computing and neural networks (NN) families which base on the neuron working mechanism have made great achievements in various aspects. Recently, statistical learning and support vector machines (SVM) are drawing extensive attention and show attractive and excellent performances in various areas compared with NN, which imply that artificial intelligence can also be made via advanced statistical computing theory.

It should be noted that as for both NN and SVM, the function imitation of human cognitive process for pattern classification can be explained as follows [1]. Given the training pairs (sample features, class indicator), we can train a NN or a SVM learning machine. The training process of these learning machines actually imitates the learning ability of human being. For clarity, we call this process “cognizing”. Then, the trained NN or SVM can be used for testing an unknown sample and determine the class indicator it belongs to. The testing process of an

X.-K. Wei (✉) · Y.-H. Li · Y.-F. Li · D.-F. Zhang
School of Engineering,
Air Force Engineering University,
Xi'an 710038, Shaanxi Province, China
e-mail: xunkai.wei@ieee.org

Y.-H. Li
e-mail: yinghong_li@126.com

Y.-F. Li
e-mail: horizon_lyf@hotmail.com

D.-F. Zhang
e-mail: zdf0918@yahoo.com.cn

unknown sample actually imitates the recognizing process of human being. We call this process “recognizing”.

From the mathematic point of view, both these two learning machines are based on the hyperplane adjustment, and obtain the optimum or sub-optimum hyperplane combinations after the training process. As for NN, each neuron acts as a hyperplane in the feature space. The feature space is divided into many partitions according to the selected training principle. Each feature space partition is linked to a corresponding class, which accomplishes the “cognizing” process. Given an unknown sample, it only detects the partition where the sample locates in and then assigns the indicator of this sample, which accomplishes the “recognizing” process. Like NN, SVM is based on the optimum hyperplane. Unlike NN, standard SVM determines the hyperplane via solving a convex optimization problem. They have the same “cognizing” and “recognizing” processes except the different solving strategies.

If a totally unknown and novel sample comes, both SVM and NN will not recognize it correctly and would prefer to assign a most close indicator in the learned classes [2, 3]. This is generally wrong in fact, and here comes the topic which this paper concerns with. The root cause of this phenomenon is the learning principle, which is based on feature space partition. This kind of learning principle may amplify each class’s region especially when the samples are small due to incompleteness. This makes it impossible to automatically detect the novel samples. Here comes the problem: how to make it clever enough to automatically identify the novel samples.

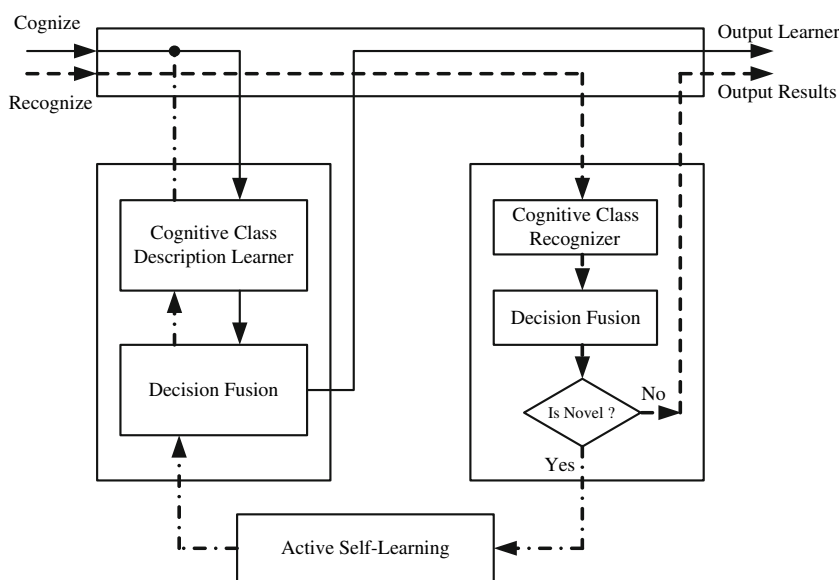
The rest of this paper is organized as follows. Section 2 gives the basic concepts of enclosing learning machine. Section 3 then describes the proposed one-class description

algorithm based on MVEE cognitive learner, and shows how this can be used to build corresponding cognitive learner in kernel-defined feature space. Section 4 presents a novel two-class classification algorithm based on single MVEE cognitive learner. Experimental results are presented in Sect. 5, and Sect. 6 contains some concluding remarks.

2 Enclosing machine learning

As is known , human being generally cognize things of one kind and recognize completely unknown things of a novel kind easily. So the question is, why not make the learning machine “cognize” or “recognize” like human being. In other words, the learning machine should “cognize” the training samples of the same class. Each class is cognized or described by a cognitive learner. It uses some kind of model to describe each class instead of using feature space partition so as to imitate the “cognizing” process. The bounding and closing boundary of each cognitive learner scatters in the feature space. For an unknown sample, the cognitive class recognizer recognizes, then detects whether the unknown sample is located inside a cognitive learner’s boundary to imitate the “recognizing” process. If the sample is completely new (i.e., none of the trained cognitive learner contains the sample), it can again be described by a new cognitive learner and the new obtained learner can be added to the feature space without changing others. We call this feedback process active self-learning. This concludes the basic concepts of enclosing machine learning. The basic flow of enclosing machine learning is depicted in Fig. 1.

Fig. 1 Enclosing machine learning paradigm. The *solid line* denotes the cognizing process. The *dashed line* denotes the recognizing process. The *dash-dotted line* denotes the active self-learning process



Now, we can investigate the definition of the cognitive learner. The cognitive learner should own at least following features:

- a. regular and convenient to calculate,
- b. bounding with the minimum volume,
- c. convex bodies to guarantee optimality,
- d. fault tolerant to guarantee generalization performance.

The basic geometric shapes perhaps are the best choices. Because they are all convex bodies and the operations like intersection, union or complement of the basic geometric shapes can be implemented easily. So we propose to use basic geometric shapes such as sphere, box or ellipsoid. The cognitive learner is then to use these geometric shapes to enclose all the given samples with minimum volume in the feature space. This is the most important reason why we call this learning paradigm enclosing machine learning. Here we give the definition of the cognitive learner and recognizer.

Definition 1 A cognitive learner is defined as the bounding and closing boundary of a minimum volume set which encloses all the given samples. The cognitive learner can be either a sphere or an ellipsoid or their combinations. Currently only ellipsoid is investigated for illustration.

Definition 2 A cognitive recognizer is defined as the point detection and assignment algorithm.

Figure 2 gives a geometric illustration of the differences between enclosing machine learning and feature space partition learning paradigm. For the cognizing (or learning) process, each class is described by a cognitive class description learner. For the recognizing (or classification) process, we only need to check which bounding learner the testing sample locates inside. For the learning process, each two classes are separated via a hyperplane (or other boundary forms, such as hypersphere). For the classification process, we need to check whether it is located on the left side or the right side of the hyperplane and then assign

the corresponding class indicator. We can see that the feature space partition learning paradigm in fact amplifies the real distribution regions of each class. But the enclosing machine learning paradigm obtains more reasonable distribution region of each class.

3 MVEE cognitive learner for one class description

3.1 Preliminaries

Our concern is with covering n given points $x_i \in \mathfrak{R}^k$, $X := [x_1 \ x_2 \ \dots \ x_n]$ with an ellipsoid of minimum volume [4, 5]. To avoid trivialities, we also make the following assumption for the remainder of this paper, which guarantees that any ellipsoid containing $X := [x_1 \ x_2 \ \dots \ x_n]$ has positive volume:

Assumption 1 The affine hull of $X := [x_1 \ x_2 \ \dots \ x_n]$ spans \mathfrak{R}^k . Equivalently, $\text{rank} \begin{bmatrix} A^T \\ e^T \end{bmatrix} = k + 1$, where e is a vector of ones, A denotes the $n \times k$ matrix, whose rows are the given points.

Definition 3 For center $c \in \mathfrak{R}^k$ and shape matrix $E \in S_{++}^{k \times k}$, we define the ellipsoid

$$\varepsilon(E, c) := \left\{ x \in \mathfrak{R}^k \mid (x - c)^T E (x - c) \leq 1 \right\}.$$

where $E \in S_{++}^{k \times k}$ determines the shape and directions of the ellipsoid. The length of the axes is given by $[\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_k}]$, where $[\lambda_1, \lambda_2, \dots, \lambda_k]$ are the corresponding eigenvalues of the matrix E .

Definition 4 For $x_i \in \mathfrak{R}^k$, $X := [x_1 \ x_2 \ \dots \ x_n]$, a MVEE cognitive learner is defined as the boundary of all the possible enclosing ellipsoids with the minimum volume.

Under the Assumption 1, a natural formulation of a minimum volume ellipsoid enclosing can be obtained via solving the following convex minimization problem,

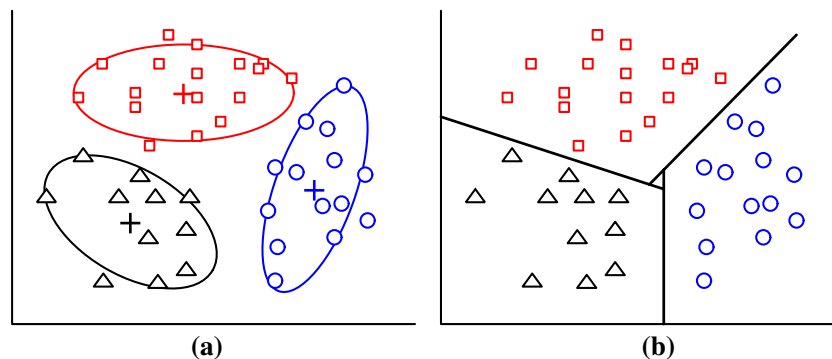


Fig. 2 A geometric illustration of learning a three class samples via enclosing machine learning vs. feature space partition learning paradigm. **a** For the depicted example, the cognitive learner is the bounding minimum volume ellipsoid, while the cognitive recognizer

is actually the point location detection algorithm of the testing sample. **b** All the three classes are separated by three hyperplanes. Obviously, each distribution region of the given class samples is amplified

$$\begin{aligned} \min_M & -\ln \det M \\ \text{s.t.} & (Mx_i - z)^T(Mx_i - z) \leq 1, \quad \forall i = 1, 2, \dots, n, \\ & M \succ 0 \end{aligned} \tag{1}$$

where, $M = \sqrt{E}$, $z = c\sqrt{E}$, square root of X is defined as: $\sqrt{X} = V^T \sqrt{D_{[d_{ii}]}} V$, $D_{[d_{ii}]}$ is an element-wise square root of eigenvalues.

Definition 5 Decompose $\tilde{M} \in \mathbb{R}^{(k+1) \times (k+1)} = (\sum_{i=1}^n \alpha_i \tilde{x}_i \tilde{x}_i^T)^{-1} = \begin{pmatrix} s & v^T \\ v & F \end{pmatrix}$, $\tilde{x}_i = \begin{pmatrix} 1 \\ x_i \end{pmatrix}$, denote $v = -Fz$, $F \in \mathbb{R}^{k \times k}$, $v \in \mathbb{R}^k$, s is a const, denote $\delta = 1 - s + \tilde{z}^T F \tilde{z}$, the linear transformation $f: \mathbb{R}^{k+1} \varepsilon(\tilde{M}, 0) \rightarrow \mathbb{R}^k \varepsilon(M, z)$ is defined as $\begin{cases} z = \tilde{z} \\ M = \delta^{-1} F \end{cases}$. Ellipsoid $\varepsilon(E, c)$ can be computed from $\begin{cases} E = (\delta^{-1} F)^T (\delta^{-1} F) \\ c = -F^{-1} v (E)^{-\frac{1}{2}} \end{cases}$.

Lemma 1 The minimization the volume of the ellipsoid $\varepsilon(M, z)$ in \mathbb{R}^k is equivalent to minimization the volume of the ellipsoid $\varepsilon(\tilde{M}, 0)$ in augmented \mathbb{R}^{k+1} centered at the origin using linear transformation f [6].

According to Lemma 1 and Definition 5, Eq. (1) can be rewritten as:

$$\begin{aligned} \min_{\tilde{M}} & -\ln \det \tilde{M} \\ \text{s.t.} & \tilde{x}_i^T \tilde{M} \tilde{x}_i \leq 1, \quad \forall i = 1, 2, \dots, n, \\ & \tilde{M} \succ 0 \end{aligned} \tag{2}$$

The dual form is:

$$\begin{aligned} \max_{\alpha_i} & \ln \det \sum_{i=1}^n \alpha_i \tilde{x}_i \tilde{x}_i^T \\ \text{s.t.} & \sum_{i=1}^n \alpha_i = k + 1 \\ & 0 \leq \alpha_i \leq 1, \quad \forall i = 1, 2, \dots, n \end{aligned} \tag{3}$$

where α is the dual variable.

We name this one class description method OCMVEE for clarity, where OCMVEE stands for one class minimum volume enclosing ellipsoid.

According to KKT conditions, we can get $\tilde{M}^{-1} = \sum_{i=1}^n \alpha_i \tilde{x}_i \tilde{x}_i^T$. We define $A: A_{ii} = \alpha_i = \sqrt{\alpha_i} \geq 0$, $X = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)$. Because $\sum_{i=1}^n \alpha_i \tilde{x}_i \tilde{x}_i^T = X^T A^2 X$, $\tilde{M}^{-1} = (AX)^T (AX) = X^T A^2 X$, $(AX)(AX)^T = AXX^T A$. Using singular value decomposition method, we can naturally define $X^T A^2 X = P A P^T$, $AXX^T A = AKA = Q A Q^T$, and thus we can obtain the following important equation $P = X^T A Q A^{-1/2}$. Using eigenspectrum analysis, we can infer following important lemma.

Lemma 2 For $\tilde{M}^{-1} = (AX)^T (AX) = X^T A^2 X$, $(AX)(AX)^T = AXX^T A$, following identities are concluded [7, 8]:

$$\begin{aligned} \ln \det (X^T A^2 X) &= \sum_{i:\lambda_i \neq 0} \ln(\lambda_i) + (k + 1 - \#\{\lambda_i \neq 0\}) \ln(0) \\ \ln \det (AXX^T A) &= \sum_{i:\lambda_i \neq 0} \ln(\lambda_i) + (n - \#\{\lambda_i \neq 0\}) \ln(0) \end{aligned} \tag{4}$$

where, λ_i are nonzero eigenvalues. (Proof Omitted)

3.2 Regularized self-adaptive MVEE cognitive learner

According to Lemma 2, it is obvious that there are probable existences of zero eigenvalues. So it is wise to introduce a regularized item to avoid this situation. And it is natural to add a regularized item μI in the $\ln \det(\bullet)$ objective function. According to Lemma 2, we can easily conclude following identities:

$$\begin{aligned} \ln \det (X^T A^2 X + \mu I) &= \sum_{i:\lambda_i \neq 0} \ln(\lambda_i + \mu) + (k + 1 - \#\{\lambda_i \neq 0\}) \ln(\mu) \\ \ln \det (AXX^T A + \mu I) &= \sum_{i:\lambda_i \neq 0} \ln(\lambda_i + \mu) + (n - \#\{\lambda_i \neq 0\}) \ln(\mu) \end{aligned} \tag{5}$$

To realize this regularized operation, we can add an additional item $\mu \text{trace}(\tilde{M})$ due to $\text{trace}(\tilde{M}) = \sum_{i=1}^{k+1} \frac{1}{\lambda_i}$, where λ_i is eigenvalue of \tilde{M}^{-1} . Then the primal regularized self-adaptive MVEE can be written as:

$$\begin{aligned} \min_{\tilde{M}, \xi_i} & -\ln \det \tilde{M} + \mu \text{trace}(\tilde{M}) + \frac{1}{n} \sum_{i=1}^n \xi_i + v \rho, \\ \text{s.t.} & \tilde{x}_i^T \tilde{M} \tilde{x}_i \leq \rho + \xi_i, \quad \forall i = 1, 2, \dots, n, \\ & \tilde{M} \succ 0, \xi_i \geq 0, \rho \geq 0, \quad \forall i = 1, 2, \dots, n \end{aligned} \tag{6}$$

where $v \geq 0$ is now a user-specified parameter that equals the fraction of objects outside the optimized ellipsoid. ρ is a variable controls the volume according to v , ξ_i is slack variable tolerates the misclassified samples.

By introducing dual variables $\alpha_i, \beta_i, \gamma \geq 0$, the Lagrange dual form is:

$$\begin{aligned} \max_{\alpha_i} & \ln \det \left(\sum_{i=1}^n \alpha_i \tilde{x}_i \tilde{x}_i^T + \mu I \right) \\ \text{s.t.} & \sum_{i=1}^n \alpha_i = v, \\ & 0 \leq \alpha_i \leq \frac{1}{n}, \quad \forall i = 1, 2, \dots, n \end{aligned} \tag{7}$$

ρ^*, ξ_i^* can be determined using following KKT conditions:

$$\begin{cases} \alpha_i^* (\rho^* + \zeta_i^* - \tilde{x}_i^T U^* U^{*T} \tilde{x}_i) = 0 \\ \beta_i^* \zeta_i^* = 0, (\frac{1}{n} - \alpha_i^*) \zeta_i^* = 0 \\ \gamma^* \rho^* = 0 \end{cases} \quad (8)$$

For a given sample \tilde{x}_i , we only need to check whether it is located inside the MVEE. If it satisfies $\tilde{x}_i^T \tilde{M} \tilde{x}_i \leq 1$, then the sample is inside the MVEE, otherwise the sample is outside the MVEE.

3.3 Kernel regularized self-adaptive MVEE cognitive learner

Matrix $X^T A^2 X$ and $AXX^T A$ have the same nonzero eigenvalues. According to (5), we have following identity:

$$\begin{aligned} \ln \det (AXX^T A + \mu I) \\ = \ln \det (X^T A^2 X + \mu I) + (n - (k + 1)) \ln (\mu) \end{aligned} \quad (9)$$

Later, we will explain how to use this important equation.

As for the inner product XX^T , we can find a kernel K which satisfies Mercer condition and then replace the inner product, i.e., $XX^T = K$. We get $AXX^T A = AKA$ and (9) can be rewritten as:

$$\ln \det (AKA + \mu I) = \ln \det (X^T AAX + \mu I) + (n - k - 1) \ln (\mu) \quad (10)$$

Hence we can optimize $\ln \det (AKA + \mu I)$ instead of $\ln \det (X^T AAX + \mu I)$. The corresponding kernel-regularized self-adaptive MVEE can be written as:

$$\begin{aligned} \max_{\alpha_i} \quad & \ln \det (AKA + \mu I) \\ s.t. \quad & \sum_{i=1}^n \alpha_i = v, \\ & 0 \leq \alpha_i \leq \frac{1}{n}, \quad \forall i = 1, 2, \dots, n \end{aligned} \quad (11)$$

So as to connect (11) with the dual variable α , we define $G \in R^{n \times n}$ to be a Cholesky factor of K , i.e. $K = GG^T$. Then we get $AKA = AGG^T A$, $G^T A^2 G = \sum_{i=1}^n \alpha_i g_i g_i^T + \mu I$, where g_i is the i th column of K . According to (9), $AGG^T A$ and $G^T A^2 G$ have the same eigenvalues, such that $\ln \det (AGG^T A + \mu I) = \ln \det (G^T A^2 G + \mu I) = \sum_{i=1}^n \alpha_i g_i g_i^T + \mu I$. Thus, we obtain the final dual kernel-regularized self-adaptive MVEE:

$$\begin{aligned} \max_{\alpha_i} \quad & \ln \det \left(\sum_{i=1}^n \alpha_i g_i g_i^T + \mu I \right) \\ s.t. \quad & \sum_{i=1}^n \alpha_i = v, \\ & 0 \leq \alpha_i \leq \frac{1}{n}, \quad \forall i = 1, 2, \dots, n \end{aligned} \quad (12)$$

Equation (12) is convex, and can be solved via the state of the art convex programming solvers such as Yalmip [9]. And then we can get the kernel-regularized MVEE cognitive learner. Now we should consider the point detection of the test samples. For ellipsoid $\varepsilon(\tilde{M}, 0)$, the Mahalanobis distance is defined as $d(\tilde{x}, \tilde{M}) = \tilde{x}^T \tilde{M} \tilde{x}$, $\tilde{x} \in R^{k+1}$. Due to the existence of matrix \tilde{M} , the Mahalanobis distance $d(\tilde{x}, \tilde{M})$ cannot be directly expressed in kernel form. Thus, we have to find another way to deduce the kernel form of the Mahalanobis distance.

Note that $X^T A^2 X = P \Lambda P^T$, we have $\sum_{i=1}^n \alpha_i \tilde{x}_i \tilde{x}_i^T + \mu I = P(\Lambda + \mu I)P^T + P^\perp (\mu I) P^{\perp T}$, $P^\perp = I - P$, $\tilde{M} = (\sum_{i=1}^n \alpha_i \tilde{x}_i \tilde{x}_i^T + \mu I)^{-1}$, $P = X^T A Q A^{-\frac{1}{2}}$. Then the Mahalanobis distance can be rewritten as:

$$d(\tilde{x}, \tilde{M}) = \frac{1}{\mu} k(\tilde{x}, \tilde{x}) - \frac{1}{\mu} \mathbf{k}^T A Q (\Lambda + \mu I)^{-1} Q^T A \mathbf{k} \quad (13)$$

where, $\mathbf{k} = (k(\tilde{x}_1, x), k(\tilde{x}_2, x), \dots, k(\tilde{x}_n, x))^T$, Q, A can be determined via $AKA = Q \Lambda Q^T$.

4 MVEE cognitive learner for two class classification

A novel two class classification algorithm based on single MVEE cognitive learner is presented. The main idea is to use MVEE to enclose all the samples of one class (m samples, $y_i = 1$) and to try to exclude the samples of the other class (n samples, $y_i = -1$). We name this method MVEEC (Minimum Volume Enclosing Ellipsoid for Classification) for clarity. The problem can be formulated as:

$$\begin{aligned} \min_{\tilde{M}, \zeta_i, \xi_j} \quad & -\ln \det (\tilde{M}) + \frac{1}{l} \sum_{i=1}^l \zeta_i + v \rho \\ s.t. \quad & y_i \tilde{x}_i^T \tilde{M} \tilde{x}_i \leq y_i \rho + \zeta_i, \quad \forall i = 1, 2, \dots, l \\ & \tilde{M} \succ 0, \rho > 0, \zeta_i \geq 0, y_i \in \{1, -1\}, \quad \forall i = 1, 2, \dots, l \end{aligned} \quad (14)$$

where, l is the total number of the samples, i.e., $l = m + n$.

According to the optimality conditions and KKT conditions, the Lagrange dual of (14) is:

$$\begin{aligned} \max_{\alpha} \quad & \ln \det \left(\sum_{i=1}^l y_i \alpha_i \tilde{x}_i \tilde{x}_i^T \right) \\ s.t. \quad & \sum_{i=1}^l y_i \alpha_i = v, \\ & 0 \leq \alpha_i \leq \frac{1}{l}, \quad \forall i = 1, 2, \dots, l \end{aligned} \quad (15)$$

where, ρ^*, ζ_i^* can be computed via following KKT conditions:

Table 1 Ball-bearing dataset novelty detection success rate

Name	σ	ν	μ	Normal	Fault 1	Fault 2	Fault 3	Fault 4
OCSVM	–	0.001	–	89.7	99.7	94.9	54.9	52.1
OCSVM	320	0.001	–	90.3	100	95.2	68.3	85.1
OCMVEE	–	0.001	0.03	92.5	100	97.1	90.8	97.2
OCMVEE	320	0.001	0.03	90.1	100	100	97.5	98.1

$$\left\{ \begin{array}{l} \alpha_i^* (y_i \rho^* + \xi_i^* - y_i \tilde{x}_i^T U^* U^{*T} \tilde{x}_i) = 0, \forall (\alpha_i^* = \frac{1}{l}) \\ (\frac{1}{l} - \alpha_i^*) \xi_i^* = 0, \forall (0 \leq \alpha_i^* < \frac{1}{l}) \end{array} \right. \quad (16)$$

5 Experiments

This section investigates the enclosing learning machine on two well-known benchmark datasets: a ball bearing dataset for novelty detection [10], an iris dataset for classification [11]. For obtaining a quantity investigation of the performances of MVEE learner, One Class SVM (OCSVM) and NuSVM (a two class classification algorithm) are adopted in corresponding dataset for performance comparison under the same backgrounds. That is to say, we compare OCMVEE method with OCSVM for the novelty detection problem. And we compare MVEEC method with NuSVM for the two-class classification problem. We use libsvm [12] for implementation of OCSVM and NuSVM. The MVEE-enclosing learning machines are programmed in Matlab via Yalmip. Both linear kernel and RBF kernel for the three methods are investigated. The optimum parameters of OCSVM, NuSVM, MVEE, and MVEEC are determined via fivefold Cross-Validation.

5.1 Novelty detection

The data comes from a real ball bearing type 6204 (steel cage), with rotational frequency 24.5625 Hz (Tacho-signal used for the measurement), sampling frequency 16384 Hz, minimum frequency 0.7 Hz, Data acquisition system (B&K analyzer). Each file consists of 10–12 vectors including 2,048 samples. Each instance consisted of 2,048 samples of acceleration. After preprocessing with a discrete Fast Fourier Transform each such instance had 32 attributes. The dataset consisted of five categories: normal data from new ball bearings and four types of abnormalities,

Table 2 Iris dataset total classification error rate

Name	σ	ν	μ	Error
NuSVM	–	0.001	–	0.06
NuSVM	10	0.001	–	0.05
MVEEC	–	0.001	0.01	0.03
MVEEC	10	0.001	0.01	0.02

i.e., fault 1 (outer race completely broken), fault 2 (broken cage with one loose element), fault 3 (damaged cage with four loose elements) and fault 4 (a badly worn ball bearing with no evident damage) (Table 1).

5.2 Pattern classification

The data set contains three classes of 50 instances each, where the three classes refer to Iris Setosa, Iris Versicolour and Iris Virginica. Each class has four attributes: sepal length, sepal width, petal length, petal width. Class Setosa is linearly separable from the other two classes Versicolour and Virginica. Since Versicolour and Virginica class are nonlinearly separable, we only report this two class classification problem (Table 2).

6 Conclusions

We propose a novel machine learning paradigm based on minimum volume enclosing shapes called enclosing machine learning and illustrated the concepts and algorithms using minimum volume enclosing ellipsoid. We have developed MVEE class description algorithm and two class classification algorithm, and validated the algorithms via benchmark datasets. The results prove the proposed MVEE-enclosing learning machines are comparable even better than SVMs in the datasets studied.

Acknowledgments This paper is jointly supported by NSFC and CAAC under grant #60672179, and also supported by the Doctorate Foundation of the Engineering College, Air Force Engineering University of China under grant #BC0501.

References

- Li Y-H, Wei X-K, Liu J-X (2004) Engineering applications of support vector machines, 1st edn. Weapon Industry Press, Beijing
- Li Y-H, Wei X-K (2005) Fusion development of support vector machines and neural networks. *J Air Force Eng Univ* 4:70–73
- Wei X-K, Li Y-H, Feng Y (2006) Comparative study of extreme learning machine and support vector machines. In: Jun Wang et al (eds) *Advances in neural networks—ISNN 2006. Lecture Notes in Computer Science*, vol 3971. Springer, Heidelberg, pp 1089–1095
- Sun P, Freund RM (2004) Computation of minimum-volume covering ellipsoids. *Oper Res* 5:690–706

5. Vandenberghe L, Boyd S, Wu S-P (1998) Determinant maximization with linear matrix inequality constraints. *SIAM J Matrix Anal Appl* 2:499–533
6. Glineur F (1998) Pattern separation via ellipsoids and conic programming. *Mémoire de D.E.A., Faculté Polytechnique de Mons, Mons, Belgium*
7. Shawe-Taylor J, Williams C, Cristianini N, Kandola JS (2002) On the eigenspectrum of the gram matrix and its relationship to the operator eigenspectrum. In: CesaBianchi N et al (eds) *Proceedings of the 13th international conference on algorithmic learning theory (ALT2002)*. Lecture Notes in Artificial Intelligence, vol 2533. Springer, Heidelberg, pp 23–40
8. Dolia AN, Page SF, White NM, Harris CJ (2004) D-optimality for minimum volume ellipsoid with outliers. In: *Proceedings of the seventh international conference on signal/image processing and pattern recognition (UkrOBRAZ'2004)* 73–76
9. Löfberg J (2006) YALMIP: A toolbox for modeling and optimization in MATLAB. <http://control.ee.ethz.ch/~joloef>
10. Ball-Bearing Dataset (2007) <http://www.hi.is/~runson/svm>
11. Newman DJ, Hettich S, Blake CL, Merz CJ (1998) Repository of machine learning databases. <http://www.ics.uci.edu/~mllearn/MLRepository>
12. Chang C-C, Lin C-J (2001) LIBSVM: a library for support vector machines. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>