

Abstract—Rival penalized competitive learning (RPCL) has been shown to be a useful tool for clustering on a set of sample data in which the number of clusters is unknown. However, the RPCL algorithm was proposed heuristically and is still in lack of a mathematical theory to describe its convergence behavior. In order to solve the convergence problem, we investigate it via a cost-function approach. By theoretical analysis, we prove that a general form of RPCL, called distance-sensitive RPCL (DSRPCL), is associated with the minimization of a cost function on the weight vectors of a competitive learning network. As a DSRPCL process decreases the cost to a local minimum, a number of weight vectors eventually fall into a hypersphere surrounding the sample data, while the other weight vectors diverge to infinity. Moreover, it is shown by the theoretical analysis and simulation experiments that if the cost reduces into the global minimum, a correct number of weight vectors is automatically selected and located around the centers of the actual clusters, respectively. Finally, we apply the DSRPCL algorithms to unsupervised color image segmentation and classification of the wine data.

Index Terms—Clustering analysis, competitive learning (CL), convergence, cost function, gradient descent.

A competitive learning network (CLN) is a type of neural network that is used for clustering and classification. It consists of a set of nodes (neurons) that compete to become the winner of a given input. The winning node is then used to update its weights, which are used to represent the cluster center. The process of updating the weights is called competitive learning. CLNs are often used in applications where the number of clusters is unknown, such as in image segmentation and classification.

The RPCL algorithm is a type of CLN that is designed to automatically determine the number of clusters. It is based on the idea of penalizing the weights of nodes that are not selected as winners. This encourages the nodes to specialize in representing different clusters. The RPCL algorithm has been shown to be effective in clustering data sets with unknown numbers of clusters.

However, the RPCL algorithm was proposed heuristically and is still in lack of a mathematical theory to describe its convergence behavior. In order to solve the convergence problem, we investigate it via a cost-function approach. By theoretical analysis, we prove that a general form of RPCL, called distance-sensitive RPCL (DSRPCL), is associated with the minimization of a cost function on the weight vectors of a competitive learning network. As a DSRPCL process decreases the cost to a local minimum, a number of weight vectors eventually fall into a hypersphere surrounding the sample data, while the other weight vectors diverge to infinity. Moreover, it is shown by the theoretical analysis and simulation experiments that if the cost reduces into the global minimum, a correct number of weight vectors is automatically selected and located around the centers of the actual clusters, respectively. Finally, we apply the DSRPCL algorithms to unsupervised color image segmentation and classification of the wine data.

The DSRPCL algorithm is a type of CLN that is designed to automatically determine the number of clusters. It is based on the idea of penalizing the weights of nodes that are not selected as winners. This encourages the nodes to specialize in representing different clusters. The DSRPCL algorithm has been shown to be effective in clustering data sets with unknown numbers of clusters.

and the extent of the research. The researchers will also be able to identify the areas that need further research. The researchers will also be able to identify the areas that need further research. The researchers will also be able to identify the areas that need further research.

The researchers will also be able to identify the areas that need further research. The researchers will also be able to identify the areas that need further research. The researchers will also be able to identify the areas that need further research.

The researchers will also be able to identify the areas that need further research. The researchers will also be able to identify the areas that need further research. The researchers will also be able to identify the areas that need further research.

© 2009 International Accounting Association

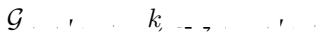
\mathcal{G}

$$W_1^{(0)}, \dots, W_n^{(0)} \quad \mathcal{G}$$

Separation Nature



Correct Division



Correct Location



\mathcal{G}

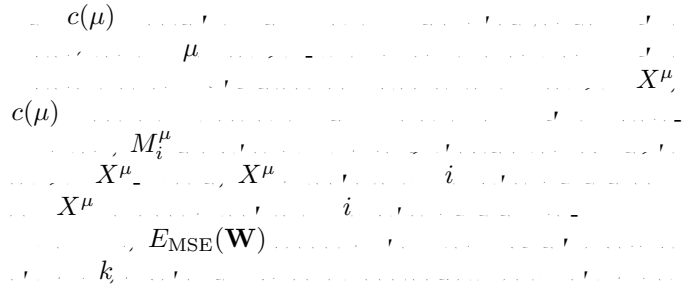
A. $D \quad \mathcal{G} \quad \mathcal{G} \quad \mathcal{G} \quad F \quad \mathcal{G}$

$$S = \{X^\mu\}_{\mu=1}^N \quad X^\mu = [x_1^\mu, x_2^\mu, \dots, x_d^\mu]^\top \quad k$$

$$\begin{aligned} E_{\text{MSE}}(\mathbf{W}) &= \frac{1}{2} \sum_{i,j,\mu} M_i^\mu (x_j^\mu - w_{ij})^2 \\ &= \frac{1}{2} \sum_{i,\mu} M_i^\mu \|X^\mu - W_i\|^2 \\ &= \frac{1}{2} \sum_{\mu} \|X^\mu - W_{c(\mu)}\|^2 \end{aligned}$$

$$\mathbf{W} = [W_1, W_2, \dots, W_k] \quad n = k \quad W_i = [w_{i1}, w_{i2}, \dots, w_{id}]^\top$$

$$M_i^\mu = \begin{cases} 1, & i = c(\mu) \\ \frac{\|X^\mu - W_{c(\mu)}\|}{\|X^\mu - W_i\|}, & \|X^\mu - W_{c(\mu)}\| = \min_j \|X^\mu - W_j\| \\ 0, & \text{otherwise} \end{cases}$$



$$E(\mathbf{W}) = E_1(\mathbf{W}) + E_2(\mathbf{W})$$

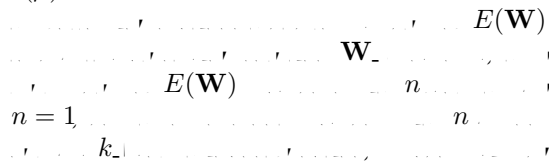
$$E_1(\mathbf{W}) = E_{\text{MSE}}(\mathbf{W}) = \frac{1}{2} \sum_{\mu} \|X^\mu - W_{c(\mu)}\|^2$$

$$E_2(\mathbf{W}) = \frac{2}{P} \sum_{\mu, i \neq c(\mu)} \|X^\mu - W_i\|^{-P}$$

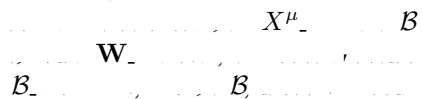
$$\mathbf{W} = \text{vec}[W_1, W_2, \dots, W_n] \quad P$$

$$c(\mu) = 1$$

$$c(\mu) = 0$$



B. $D \quad v \quad v \quad \mathcal{G} \quad \mathcal{G} \quad F \quad \mathcal{G}$



\mathbf{W}
 \mathbf{W}

\mathcal{B}

\mathbf{W}

X^μ

M_i^μ

\mathbf{W} , $E(\mathbf{W})$

\mathbf{W}

w_{ij}

$$\begin{aligned} \frac{\partial E(\mathbf{W})}{\partial w_{ij}} &= \frac{\partial E_1(\mathbf{W})}{\partial w_{ij}} + \frac{\partial E_2(\mathbf{W})}{\partial w_{ij}} \\ &= - \sum_{\mu} \delta_{i,c(\mu)} (x_j^\mu - w_{ij}) + \sum_{\mu,i} (1 - \delta_{i,c(\mu)}) \\ &\quad \times \|X^\mu - W_i\|^{-P-2} (x_j^\mu - w_{ij}) \end{aligned}$$

$\delta_{i,j}$

\mathbf{W}

X^μ

\mathbf{W}

\mathbb{R}^{nd}

M_i^μ

\mathbf{W}'

$W'_i, W'_j, \quad i < j$

$X^{\mu'}$

$$\|W'_i - X^{\mu'}\| = \|W'_j - X^{\mu'}\| = \min_l \|W'_l - X^{\mu'}\| > 0.$$

\mathbf{W}'

$$\|W_i - X^{\mu'}\| = \|W_j - X^{\mu'}\|$$

\mathcal{A}_{l_i}

\mathcal{A}_{l_j}

$$\|W_i - X^{\mu'}\| = \min_l \|W_l - X^{\mu'}\| \leq \|W_j - X^{\mu'}\|$$

$$\mathcal{A}_{l_i} \quad \|W_j - X^{\mu'}\| = \min_l \|W_l - X^{\mu'}\| < \|W_i - X^{\mu'}\|$$

$$\mathcal{A}_{l_j} \quad i < j$$

\mathcal{A}_{l_i}

\mathcal{A}_{l_i}

$$\mathbf{W}' \quad M_i^{\mu'} = 1$$

$$M_j^{\mu'} = 0 \quad \mathcal{A}_{l_j}$$

\mathbf{W}'

$$M_i^{\mu'} = 0 \quad M_j^{\mu'} = 1 \quad \|W'_i - X^{\mu'}\| = \|W'_j - X^{\mu'}\|$$

$$X^{\mu'}$$

$$C_i = \mathcal{S} \cap R_i, \quad i = 1, \dots, n.$$

$$\bar{C}_i = \mathcal{S} - C_i$$

$$E_1(\mathbf{W})$$

$$E_2(\mathbf{W})$$

$$E_2(\mathbf{W})$$

$$X^\mu$$

$$\|X^\mu - W_{r(\mu)}\|^{-P}, \quad r(\mu)$$

$$E_2(\mathbf{W})$$

$$\Delta W_i = \begin{cases} \eta(X^\mu - W_i), & i = c(\mu) \\ -\eta\|X^\mu - W_i\|^{-P-2}(X^\mu - W_i), & i = r(\mu) \\ 0, & \text{otherwise} \end{cases}$$

$$\alpha_c = \eta, \quad \alpha_r = \eta\|X^\mu - W_{r(\mu)}\|^{-P-2}.$$

$$\frac{\alpha_c}{\alpha_r} = \|X^\mu - W_{r(\mu)}\|^{2+P}.$$

$$\alpha_c/\alpha_r = \|X^\mu - W_{r(\mu)}\|^{2+P}$$

$$P$$

$$\alpha_c = \eta$$

$$\alpha_r$$

$$\eta$$

$$\eta$$

$$\eta$$

$$\eta$$

$$\mathbf{W}^{(t)} = [W_1^{(t)}, \dots, W_n^{(t)}]$$

$$\mathbf{W} = [W_1^{(0)}, \dots, W_n^{(0)}] \in R^{nd} - \mathcal{B}$$

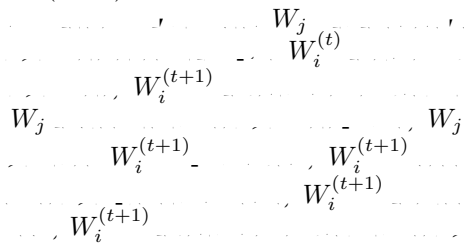
$$I: \quad W_i^{(t)}$$

$$\mathcal{G}: \quad W_i^{(t)}$$

$$\Delta W_i = \eta \sum_{\mu} \|X^{\mu} - W_i\|^{-P-2} (W_i - X^{\mu}).$$

$$\frac{\Delta W_i}{W_i^{(t+1)}}$$

$$\frac{E(\mathbf{W}^{(t)})}{E(\mathbf{W}^{(t)})}$$



W_i

\mathcal{G}

η

$W_i^{(t)}$

\mathcal{G}

\mathcal{G}

\mathcal{G}

$E(\mathbf{W}^{(t)})$

t

E^*

t

$\{W_i^{(t)}\}$

$\{W_i^{(t)}\}$

$W_i^{(t)}$

$$E^* \quad \{W_i^{(t)}\} \quad E(\mathbf{W}^{(t)})$$

$\{W_i^{(t)}\}$

\mathcal{G}

$\{W_i^{(t)}\}$

T

$t > T, W_i^{(t)}$

\mathcal{G}

η

$E(\mathbf{W})$

$\hat{\mathbf{W}}^*$

$E(\hat{\mathbf{W}})$

$\hat{\mathbf{W}}$

$\hat{\mathbf{W}}^*$

η

$E(\hat{\mathbf{W}})$

$E(\hat{\mathbf{W}})$

$$\|W_i - W_j\| \geq \delta \quad i \neq j \quad \delta$$

$E(\mathbf{W})$

$\{\mathbf{W} : E(\mathbf{W}) \leq C\}$

$$C > 0 \quad E(\mathbf{W})$$

η

$$E(\mathbf{W}^{(t)})$$

B. $C_{1, \dots, D}$ v $\mathcal{G}, \mathcal{G}, \mathcal{G}$

Let us consider a system with n nodes, k of which are connected to a central node. The remaining $n-k$ nodes are connected to each other in a ring topology. The system is described by the following equations:

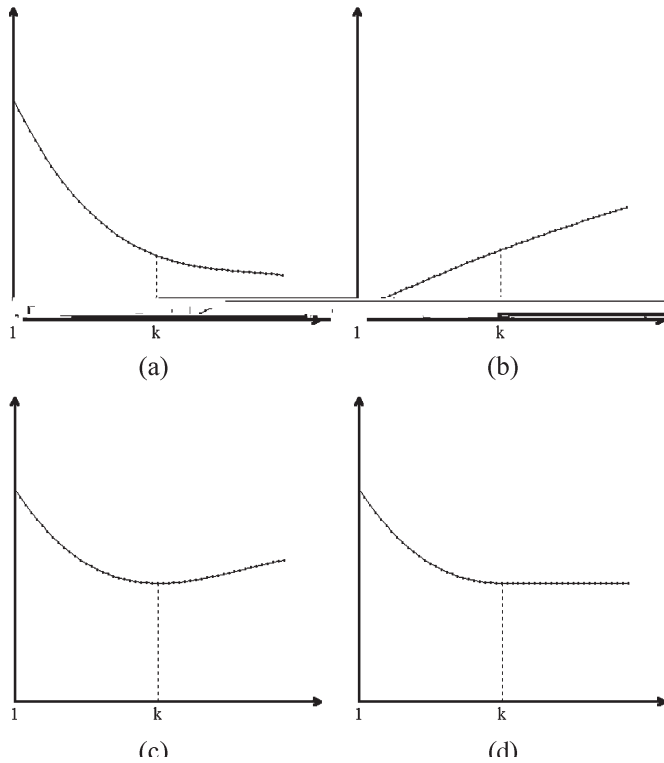
$$\begin{aligned} \dot{\mathbf{W}} &= \mathbf{W}^* - \mathbf{W} \\ \dot{\mathbf{W}} &= \mathbf{W}^* - \mathbf{W} \end{aligned}$$

$$\begin{aligned} \dot{\mathbf{W}} &= \mathbf{W}^* - \mathbf{W} \\ \dot{\mathbf{W}} &= \mathbf{W}^* - \mathbf{W} \end{aligned}$$

$$1) \quad \mathcal{G} \quad m_i \quad \mathcal{G} \quad \mathbf{W}^* \quad \mathcal{G} \quad E(\mathbf{W}) \quad n = k: \quad E(\mathbf{W}) = E_1(\mathbf{W}) + E_2(\mathbf{W})$$

$$\begin{aligned} \mathbf{W}^0 &= [m_1, \dots, m_k] \\ \mathbf{W}^0 &= [m_1, \dots, m_k] \end{aligned}$$

$$\begin{aligned} \mathbf{W} &= k \quad C_1, \dots, C_k \\ E_1(\mathbf{W}) &= C_i \\ E_1(\mathbf{W}) &= E_1(\mathbf{W}^0) \quad \mathbf{W} \end{aligned}$$



$n = k$

k

P

P

P

P

P

$[P_0, P_1]$ P

$P \geq 0.01$

$E_1(W)$ $E(W)$ $n = k$

P_0

$P > 1.9$

$E_2(W)$ $E(W)$ $E(W)$

$E(W)$

$E(W^{(t)})$ $E(W)$

$E(W)$ $n = k$

$E(W)$

A. ϑ E ϑ C A y

1) D $d = 2$

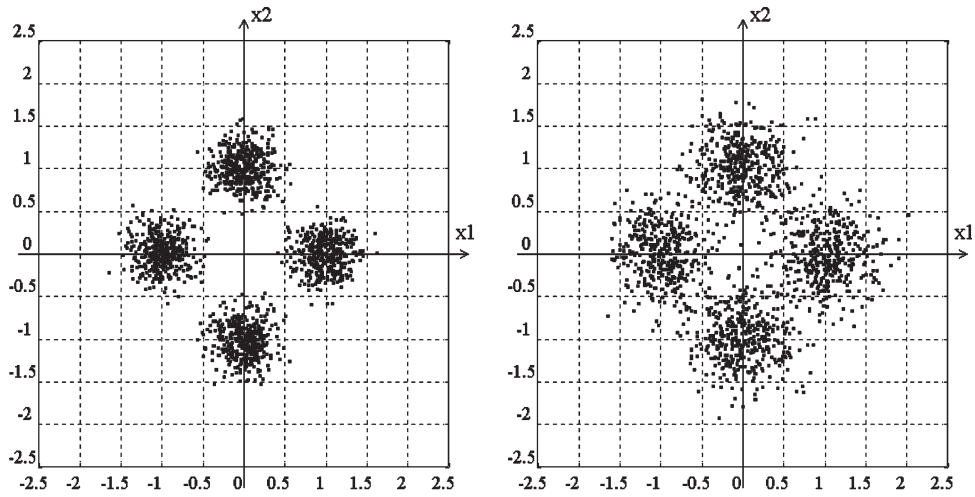
$\sigma^2 I$ σ

m_i σ_i α_i N_i

i

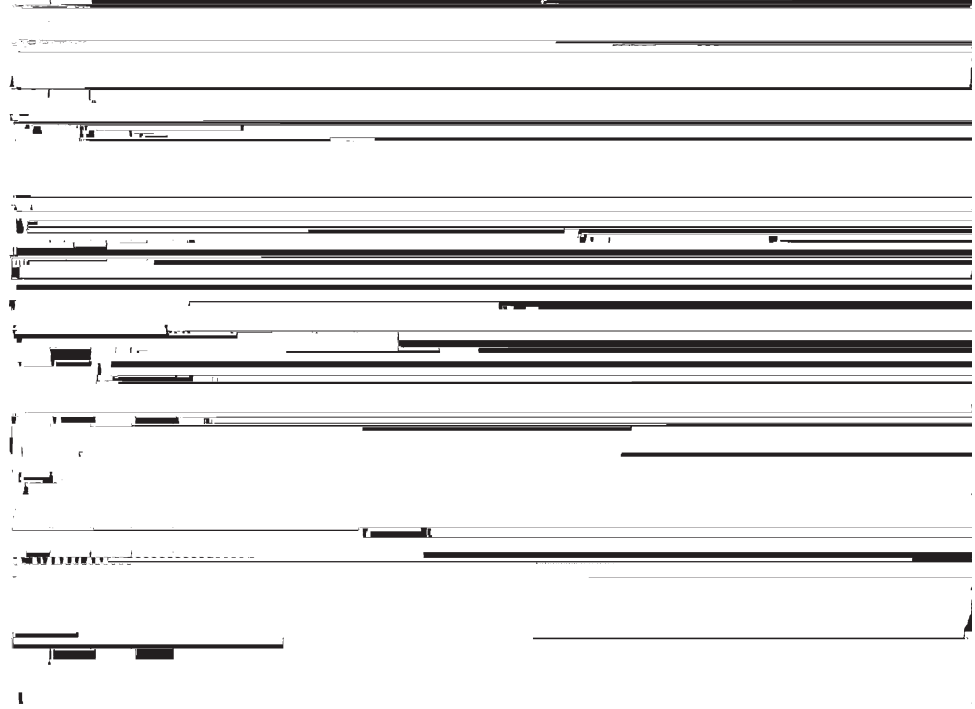
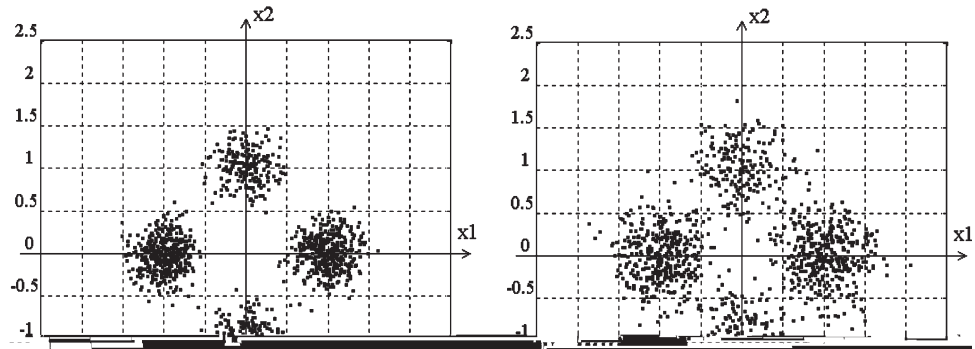
C. $\vartheta \vartheta P$

$E(W)$ $E(W)$



(a)

(b)



S_3 S_4 S_5 S_1 S_2

2) \emptyset \emptyset D C A \emptyset : $E(W)$
 $E(W^{(t)})$
 $E(W)$ n

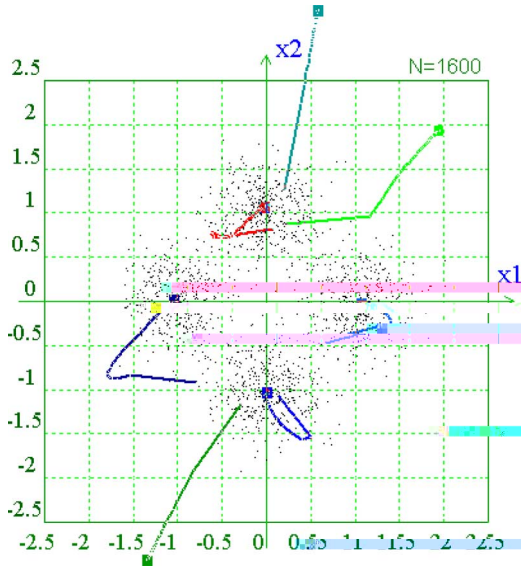
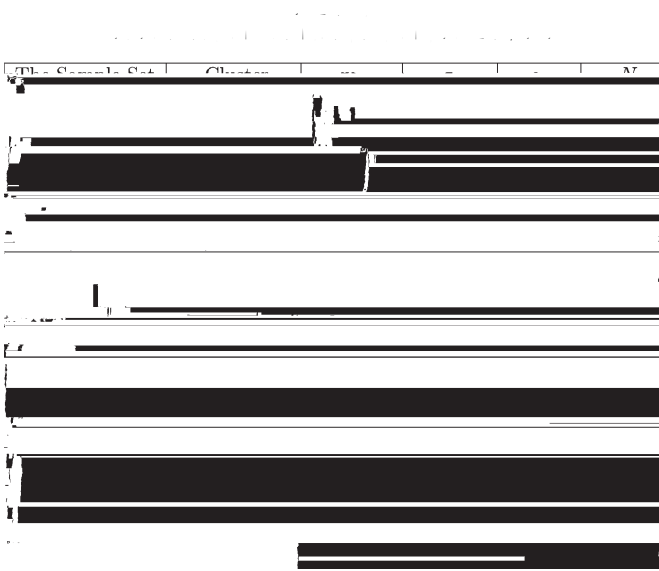


Figure 1: Scatter plot of $N=1600$ data points in the x_1 - x_2 plane. The plot shows several clusters of points. A red line connects the origin to a point in the upper right cluster. A green line connects the origin to a point in the lower left cluster. A blue line connects the origin to a point in the lower right cluster. A pink horizontal bar is located at approximately $x_2 = 0$. A cyan horizontal bar is located at approximately $x_2 = -1.5$. A yellow horizontal bar is located at approximately $x_2 = -0.5$.

S_1 S_2 $E(W)$ $E_2(W)$ η/m $m = \lfloor t/5 \rfloor$

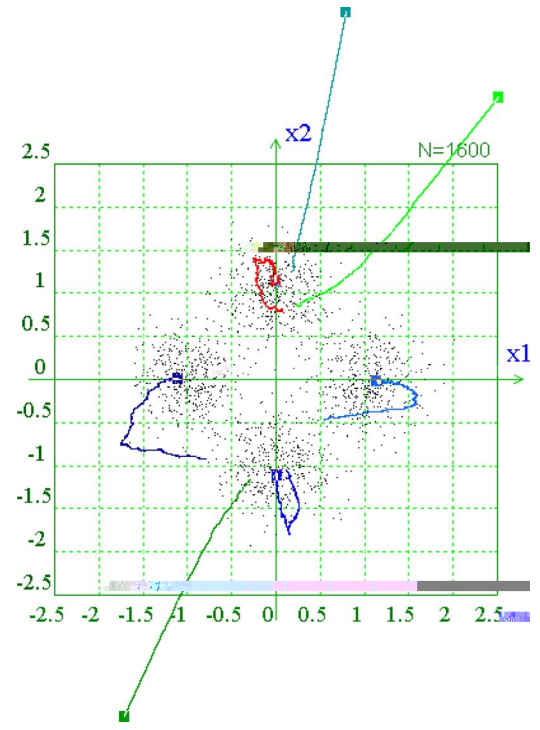


Figure 2: Scatter plot of $N=1600$ data points in the x_1 - x_2 plane. The plot shows several clusters of points. A red line connects the origin to a point in the upper right cluster. A green line connects the origin to a point in the lower left cluster. A blue line connects the origin to a point in the lower right cluster. A red horizontal bar is located at approximately $x_2 = 1.5$. A green horizontal bar is located at approximately $x_2 = 1.5$. A blue horizontal bar is located at approximately $x_2 = 0$. A pink horizontal bar is located at approximately $x_2 = -1.5$. A yellow horizontal bar is located at approximately $x_2 = -0.5$.

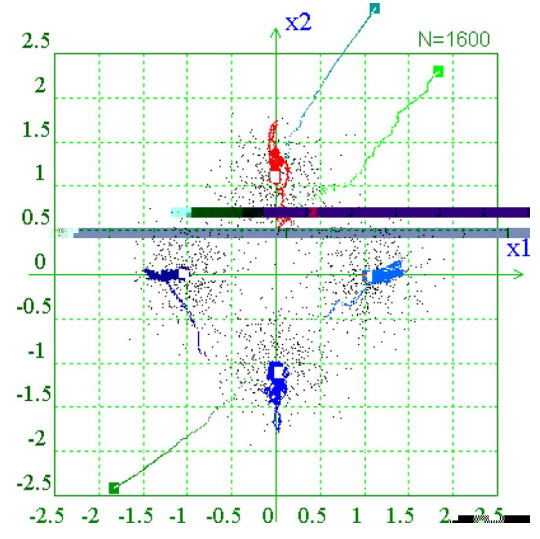
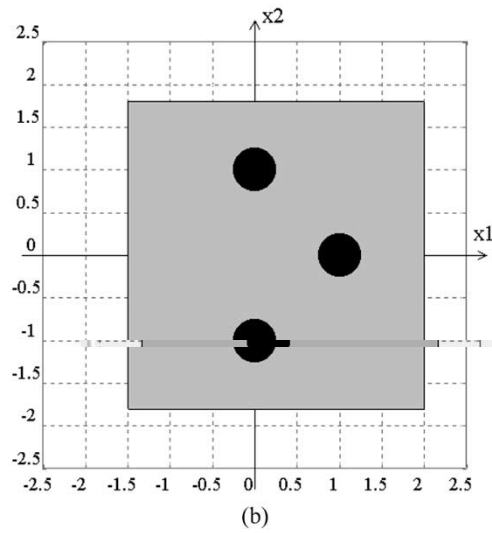
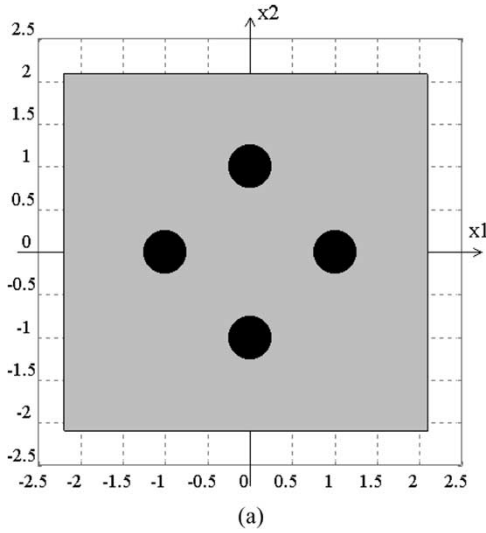


Figure 3: Scatter plot of $N=1600$ data points in the x_1 - x_2 plane. The plot shows several clusters of points. A red line connects the origin to a point in the upper right cluster. A green line connects the origin to a point in the lower left cluster. A blue line connects the origin to a point in the lower right cluster. A red horizontal bar is located at approximately $x_2 = 0.5$. A green horizontal bar is located at approximately $x_2 = 0.5$. A blue horizontal bar is located at approximately $x_2 = 0$. A pink horizontal bar is located at approximately $x_2 = -1.5$. A yellow horizontal bar is located at approximately $x_2 = -0.5$.

$m = \lfloor t/5N \rfloor$ $[x]$ N $n = 7$ $k = 4$



$-2.1, 2.1 \times -2.1, 2.1$
 $-1.5, 2.0 \times -1.5, 1.5$

$$\Delta E_t(\mathbf{W}) = |E(\mathbf{W}^{(t)}) - E(\mathbf{W}^{(t-1)})| < 10^{-6}$$

$$X^\mu = \begin{cases} \xi & \xi > \lambda \\ \xi & \xi \leq \lambda \end{cases} \quad \mathcal{S} = \{X^1, X^2, \dots, X^N\}$$

$$\Delta W_i = \begin{cases} \eta(X^\mu - W_i), & i = c(\mu) \\ -\eta \|X^\mu - W_i\|^{-P-2}(X^\mu - W_i), & \xi \leq \lambda \end{cases}$$

$$\Delta W_i = \begin{cases} -\eta(X^\mu - W_i), & i = c(\mu) \\ \eta \|X^\mu - W_i\|^{-P-2}(X^\mu - W_i), & \xi \leq \lambda \end{cases}$$

$$t < M, \quad t = t + 1$$

$$\lambda < \varepsilon, \quad T = T + 1$$

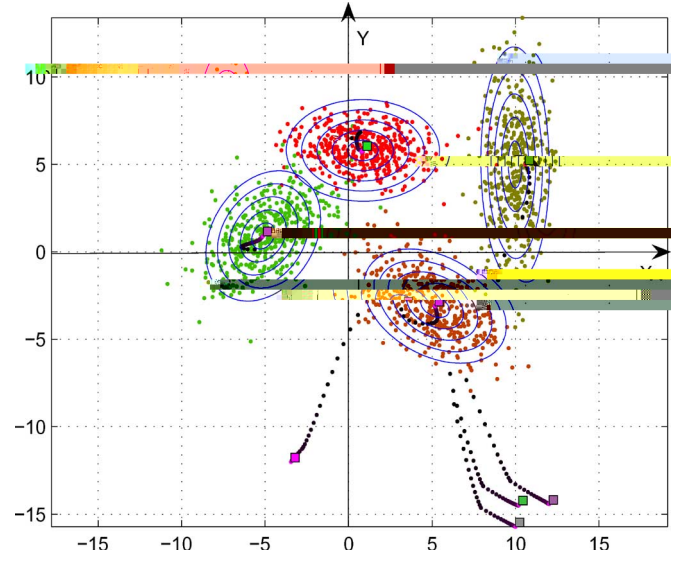
$E(\mathbf{W})$
 $E(\mathbf{W})$
 $n \leq 2k$
 $(A \ C) A \ \mathcal{S}_3 \ \mathcal{S}_4 \ \mathcal{S}_5$
 n
 k

$\eta/(T+1)$
 $1/T$
 M
 M
 n
 η
 10^{-6}
 T
 $k_0 \ k_1 \ c_0 \ c_1$
 $[a, b]$

$$w_{ij} = \frac{\eta_0}{c_1 T + c_0} \exp(-k_1 T - k_0), \quad \eta = \eta_0 / (c_1 T + c_0), \quad T = 0, \quad t = 0$$

$\eta_0 = 0.003$ $M = 100$ $k_1 = 0.005$ $k_0 = 1.200$ $c_1 = 0.015$
 $c_0 = 1.000$ $[a, b] = [-1.2, 1.2]$
 $T = 10000$

The set of sample data	S_1	S_2	S_3	S_4	S_5
VP = 100%	4-36	4-9	4-8	4-6	3-5
VP \geq 97%	4-62	4-21	4-15	4-9	3-9
VP \geq 95%	4-65	4-38	4-25	4-14	3-10



n k 4 η
 -3

θ

S_1 S_2

B E θ θ θ

$E(\mathbf{W}^{(t)})$
 $E(\mathbf{W})$ n k
 $E(\mathbf{W})$ n k
 n
 n
 n n k
 n
 n

w_{ij}
 $[a, b]$

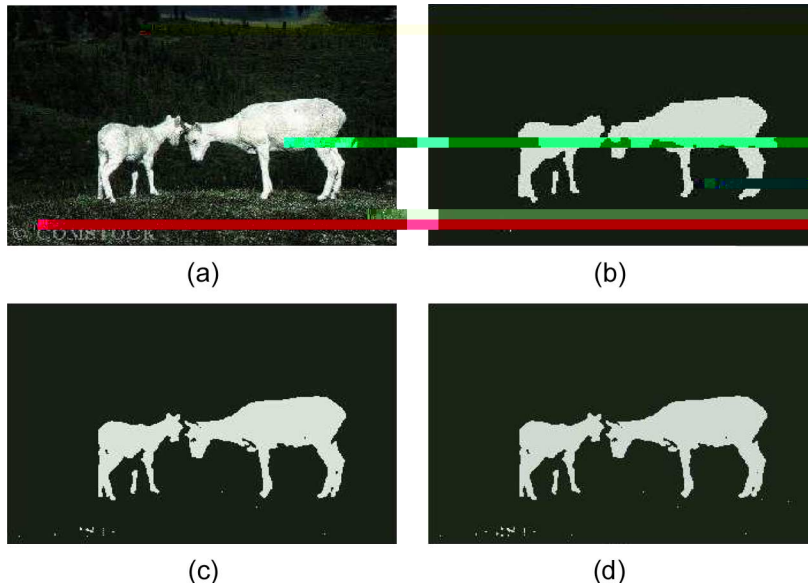


Figure 2: Results of the proposed method. (a) Original image. (b) Binary mask of the sheep. (c) Mask with a red horizontal line indicating a segmentation error. (d) Mask with a green horizontal line indicating a segmentation error.

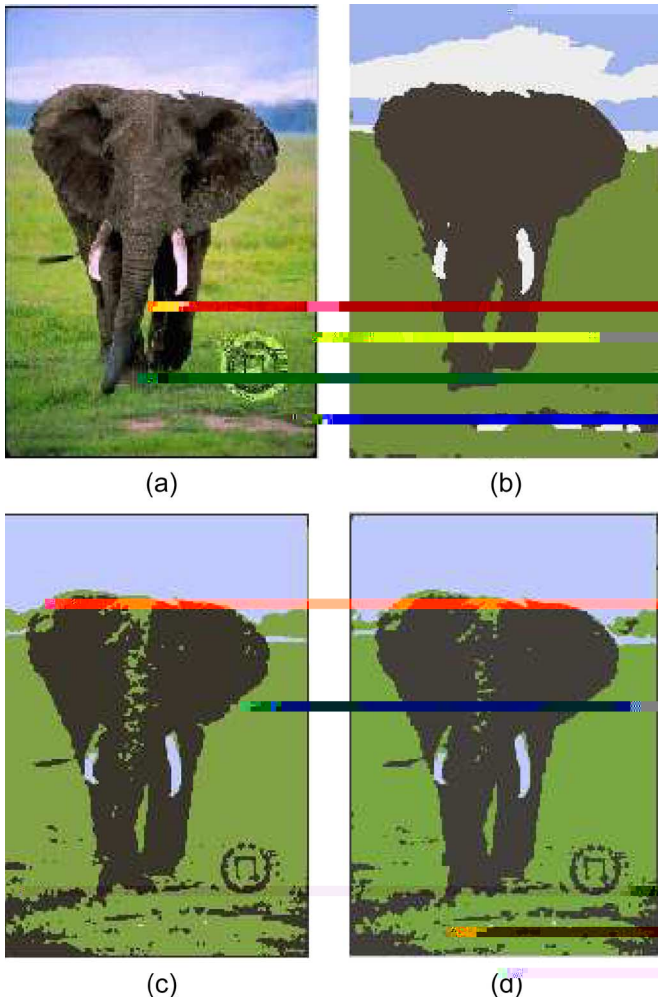
Let us consider the problem of finding the optimal value of k for a given image. In this paper, we propose a method for finding the optimal value of k for a given image. Let k^* be the optimal value of k for a given image. We propose a method for finding k^* for a given image. The method is based on the following idea: the optimal value of k is the value of k that minimizes the number of segmentation errors. In this paper, we propose a method for finding k^* for a given image. The method is based on the following idea: the optimal value of k is the value of k that minimizes the number of segmentation errors.

Let us consider the problem of finding the optimal value of k for a given image. In this paper, we propose a method for finding the optimal value of k for a given image. Let k^* be the optimal value of k for a given image. We propose a method for finding k^* for a given image. The method is based on the following idea: the optimal value of k is the value of k that minimizes the number of segmentation errors. In this paper, we propose a method for finding k^* for a given image. The method is based on the following idea: the optimal value of k is the value of k that minimizes the number of segmentation errors.

Let us consider the problem of finding the optimal value of k for a given image. In this paper, we propose a method for finding the optimal value of k for a given image. Let k^* be the optimal value of k for a given image. We propose a method for finding k^* for a given image. The method is based on the following idea: the optimal value of k is the value of k that minimizes the number of segmentation errors. In this paper, we propose a method for finding k^* for a given image. The method is based on the following idea: the optimal value of k is the value of k that minimizes the number of segmentation errors.

Let us consider the problem of finding the optimal value of k for a given image. In this paper, we propose a method for finding the optimal value of k for a given image. Let k^* be the optimal value of k for a given image. We propose a method for finding k^* for a given image. The method is based on the following idea: the optimal value of k is the value of k that minimizes the number of segmentation errors. In this paper, we propose a method for finding k^* for a given image. The method is based on the following idea: the optimal value of k is the value of k that minimizes the number of segmentation errors.

Let us consider the problem of finding the optimal value of k for a given image. In this paper, we propose a method for finding the optimal value of k for a given image. Let k^* be the optimal value of k for a given image. We propose a method for finding k^* for a given image. The method is based on the following idea: the optimal value of k is the value of k that minimizes the number of segmentation errors. In this paper, we propose a method for finding k^* for a given image. The method is based on the following idea: the optimal value of k is the value of k that minimizes the number of segmentation errors.



... ..

D. D B v D C A

... ..

... ..

... ..

... ..

Figure 2. The original image and its segmented image. (a) Original image. (b) Segmented image. (c) Segmented image with red and blue bars. (d) Segmented image with red and blue bars.

... ..

C. C D

... ..

EEE

EEE

§ § : A A §

§ EEE

EEE §

§

A v § § § y

§ §

§

§

A §

§

§ EEE

§ §

§ II §

§ §

EEE

EEE

§

y §

EEE

EEE

EEE

§

EEE

EEE

§

§ §

y

EEE

§

§ § §

§

§

§

§ §

§ EEE § § §

A v § §

y

EEE

EEE

EEE

EEE

§ D E § A

D

§

A y

A § § § § &

§ §

§ § 19 § § A F y § § §

§

C § § §

D

§



Jinwen Ma received his B.S. degree in chemistry from the University of Science and Technology of China, Hefei, China, in 2004. He received his M.S. degree in chemistry from the University of Science and Technology of China, Hefei, China, in 2007. He received his Ph.D. degree in chemistry from the University of Science and Technology of China, Hefei, China, in 2010. He is currently an associate professor in the School of Chemistry, University of Science and Technology of China, Hefei, China.

His research interests include the synthesis and characterization of novel polymers and their applications in catalysis and materials science.

He has published over 20 papers in the field of polymer chemistry and materials science. He is also a member of the Chinese Chemical Society and the American Chemical Society.



Taijun Wang received his B.S. degree in chemistry from the University of Science and Technology of China, Hefei, China, in 2004. He received his M.S. degree in chemistry from the University of Science and Technology of China, Hefei, China, in 2007. He received his Ph.D. degree in chemistry from the University of Science and Technology of China, Hefei, China, in 2010. He is currently an associate professor in the School of Chemistry, University of Science and Technology of China, Hefei, China.

His research interests include the synthesis and characterization of novel polymers and their applications in catalysis and materials science.

He has published over 20 papers in the field of polymer chemistry and materials science. He is also a member of the Chinese Chemical Society and the American Chemical Society.