

**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** 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.

$k$

$k$

$k$

$k$

$k$

The first part of the document discusses the importance of maintaining accurate records of all transactions. It emphasizes that every entry should be supported by a valid receipt or invoice. This not only helps in tracking expenses but also ensures compliance with tax regulations. The document further outlines the process of reconciling bank statements with the company's ledger to identify any discrepancies. It stresses the need for regular audits to prevent errors and fraud. The second part of the document provides a detailed breakdown of the company's financial performance over the past year. It includes a comparison of actual results against budgeted figures and identifies key areas of improvement. The document concludes with a summary of the overall financial health and a forecast for the upcoming period. It highlights the company's strong position in the market and its commitment to sustainable growth.

Page 1 of 1

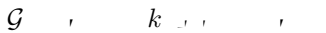
$\mathcal{G}$

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

**Separation Nature**



**Correct Division**



**Correct Location**



$\mathcal{G}$

A. D k C F

$$[x_1^\mu, x_2^\mu, \dots, x_d^\mu]^T$$

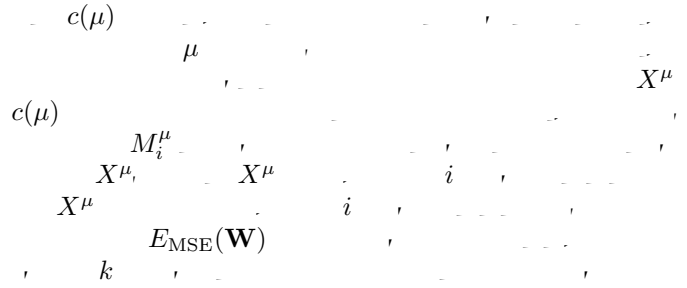
$$\mathcal{S} = \{X^\mu\}_{\mu=1}^N \quad X^\mu =$$

$k$

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

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

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



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

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

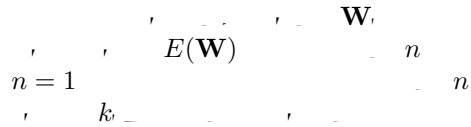
$$E_2(\mathbf{W}) = \frac{2}{P} \sum_{\mu, i \neq c(\mu)} \|X^\mu - W_i\|^{-P}$$

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

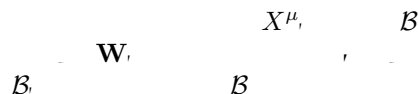
$$c(\mu) = 1$$

$$c(\mu) = 0$$

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



B. D k C F



$$\begin{aligned}
& \mathbf{W}, \\
& \mathcal{B} \\
& \mathbf{W} \\
& X^\mu, \\
& M_i^\mu \\
& \mathbf{W}, \quad E(\mathbf{W}) \\
& \mathbf{W}, \\
& w_{ij}
\end{aligned}$$

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

$\delta_{i,j}$

$\mathbf{W}$

$X^\mu,$

$\mathbf{W}$

$R^{nd}$

$M_i^\mu$

$\mathbf{W}'$

$$\begin{aligned}
W_i' & \quad W_j' \quad i < j \\
X^{\mu'}
\end{aligned}$$

$$\|W_i' - X^{\mu'}\| = \|W_j' - X^{\mu'}\| = \min_l \|W_l' - X^{\mu'}\| > 0.$$

$\mathbf{W}'$

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

$\mathcal{A}_{l_i} \quad \mathcal{A}_{l_j}$

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

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

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

$\mathcal{A}_{l_i}$

$\mathcal{A}_{l_i}$

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

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

$\mathbf{W}'$

$$M_i^{\mu'} = 0 \quad M_j^{\mu'} = 1 \quad \|W_i' - X^{\mu'}\| = \|W_j' - X^{\mu'}\|$$

$X^{\mu'}$

$\mathcal{S}$   $n$

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

$X^\mu$

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

$C_i$

$E_1(\mathbf{W})$

$E_2(\mathbf{W})$

$E_2(\mathbf{W})$

$X^\mu$

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

$E_2(\mathbf{W})$

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

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

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

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

$\alpha_c$

$\eta$

$\alpha_r$

$\eta$

$\eta$

$\eta$

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

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

$E^*$

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

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

$$W_i^{(t)}$$

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

$\mathcal{G}$

$l:$

$$W_i^{(t)}$$

$t$

$T$

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

$$t > T \quad W_i^{(t)}$$

$\mathcal{G}$

$\eta$

$:$

$$W_i^{(t)}$$

$t$

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

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

$E(\mathbf{W})$

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

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

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

$\hat{\mathbf{W}}$

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

$$W_j$$

$$W_i^{(t)}$$

$\eta$

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

$$W_j$$

$$W_i^{(t+1)}$$

$$W_j$$

$$W_i^{(t+1)}$$

$$W_i^{(t+1)}$$

$$W_i^{(t+1)}$$

$$W_i$$

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

$h$

$l:$

$\eta$

$\mathcal{G}$

$$W_i^{(t)}$$

$\mathcal{G}$

$\mathcal{G}$

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

$t$

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

$E(\mathbf{W})$

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

$E^*$

$C > 0$

$E(\mathbf{W})$

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

$t$

$$W_i^{(t)}$$

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

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

$\eta$

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

B. C D

$$\begin{matrix} & & k & & \\ & & | & & \\ & & k & & \\ & & | & & \\ & & n-k & & \end{matrix}$$

$$\begin{matrix} & & k & & \\ & & | & & \\ & & k & & \end{matrix}$$

$$m_1, \dots, m_k$$

$$m_i$$

$$1) \quad \mathbf{W}^* \quad E(\mathbf{W}) \quad n = k: \quad E(\mathbf{W}) =$$

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

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

$$\mathbf{W}^0,$$

$$\mathbf{W}^0$$

$$\mathbf{W}^0 = [m_1, \dots, m_k]$$

$$\mathbf{W} \quad k \quad C_1, \dots, C_k$$

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

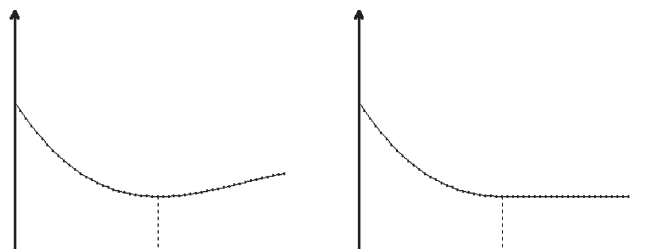
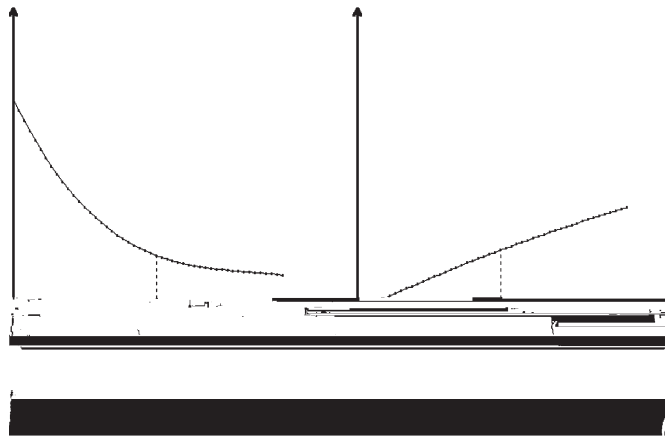
$$E_1(\mathbf{W})$$

$$E_1(\mathbf{W}^0)$$

$$\mathbf{W}$$

$$C_i$$

$$\mathbf{W}^0$$



$n$   $k$   
 $E_2(\mathbf{W})$   $E_1(\mathbf{W})$   $E(\mathbf{W})$   
 $E(\mathbf{W})$

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

$E(\mathbf{W})$   $n = k$

$E(\mathbf{W})$

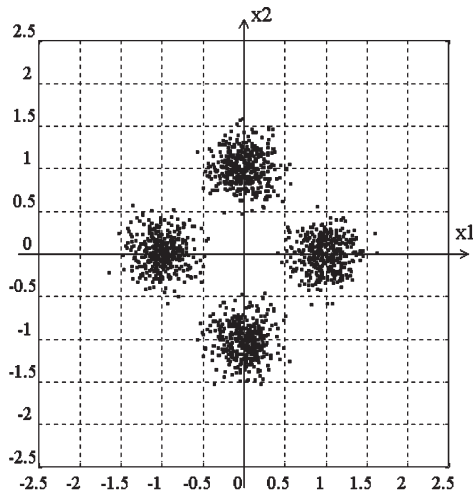
C.  $P$

$E(\mathbf{W})$

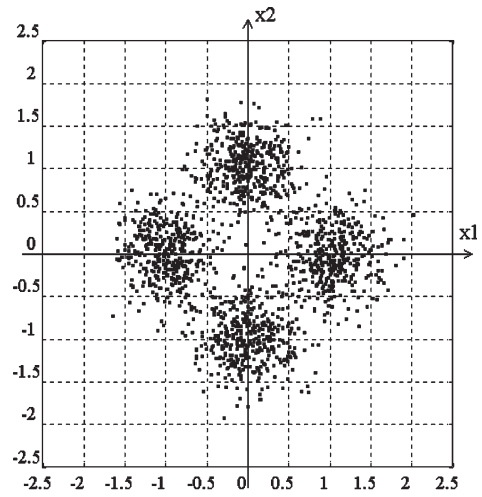
$E(\mathbf{W})$

$n = k$   
 $k$   
 $P$   
 $P$   
 $P$   
 $P$   
 $[P_0, P_1]$   $P$   
 $P \geq 0.01$   
 $E_1(\mathbf{W})$   $E(\mathbf{W})$   
 $n = k$   
 $P_0$   
 $P > 1.9$   
 $P$   $P_1$   
 $P$   
 A.  $E$   $C$   $A$   
 I)  $D$  :  
 $d = 2$   
 $\sigma^2 I$   $\sigma$   
 $m_i$   $\sigma_i$   $\alpha_i$   $N_i$   
 $i$

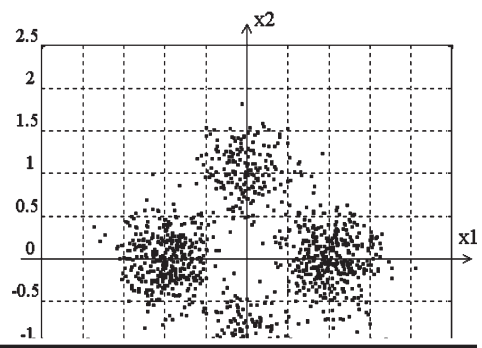
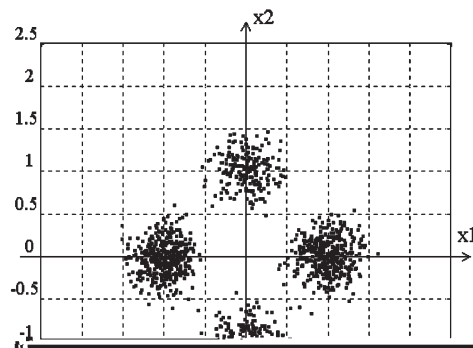




(a)



(b)



$S_3$

$S_4$

$S_5$

$S_1$

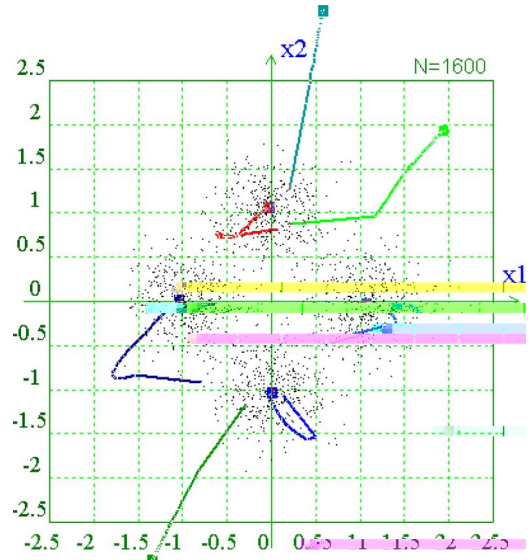
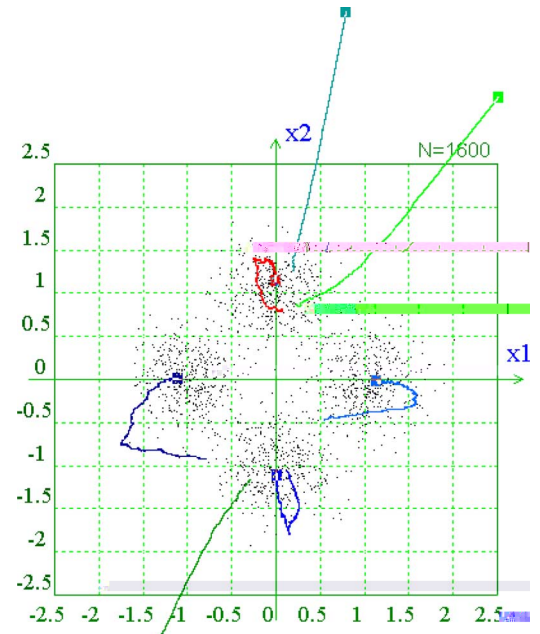
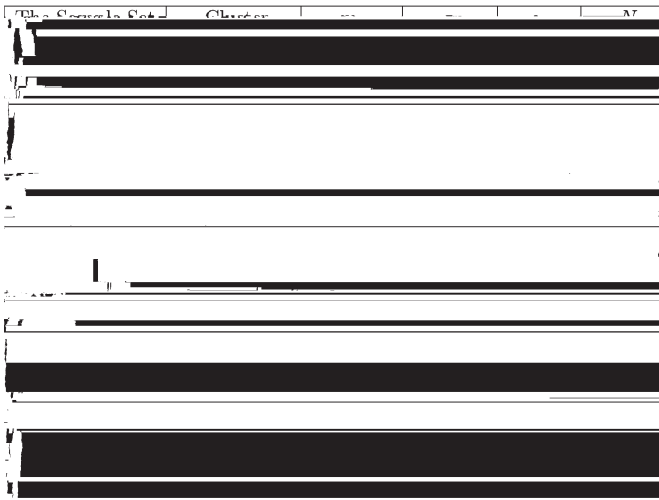
$S_2$

2)

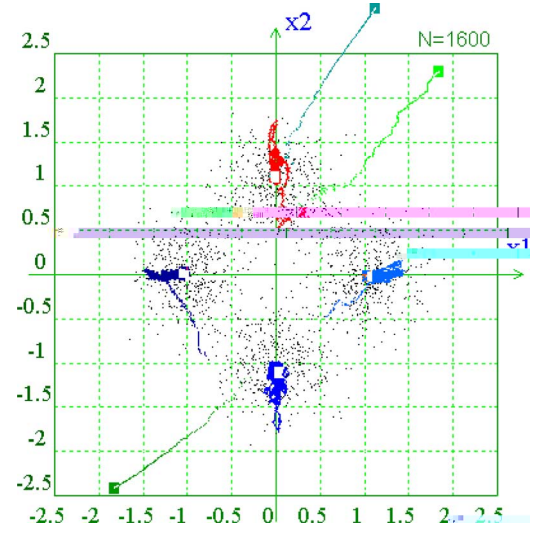
$h \quad D \quad C \quad A \quad h \quad :$

$E(W) \quad E(W^{(t)})$   
 $n$

$E(W)$



$n$   $k = 4$   $\eta$



$n$   $k = 4$   $\eta$

$S_1$   $S_2$

$E(W)$

$S_2$

$E_2(W)$

$E_2(W)$

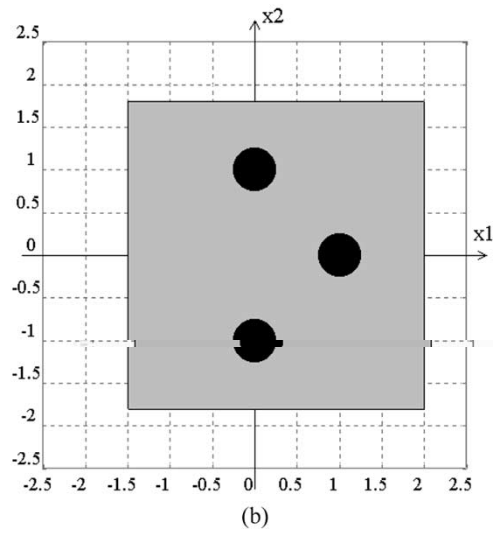
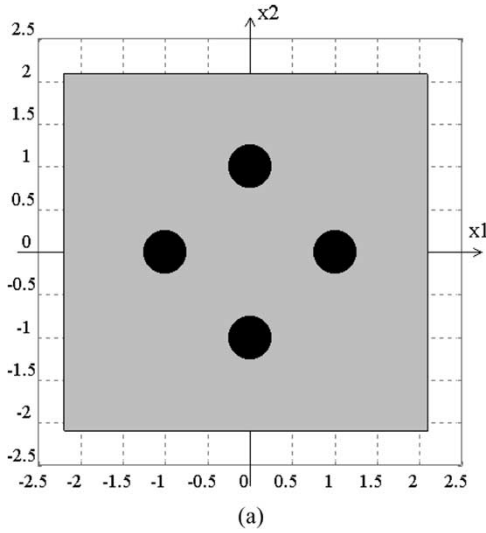
$\eta/m$   $m = \lceil t/5 \rceil$

$m = \lceil t/5N \rceil$

$[x]$   $x$   $t$

$N$

$n = 7$   $k = 4$



$$[-2.1, 2.1] \times [-2.1, 2.1]$$

$$[-1.5, 2.0] \times [-1.5, 1.5]$$

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

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

$$X^\mu \in S = \{X^1, X^2, \dots, X^N\}$$

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

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

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

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

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

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

$$n \leq 2k$$

$$(A \ C) A \quad k \quad A \quad C$$

$$S_3 \ S_4 \ S_5$$

$$n$$

$$k$$

n

$$\eta/(T+1)$$

$$1/T$$

M

M

n

\eta

$$10^{-6}$$

T

k<sub>0</sub>

k<sub>1</sub>

c<sub>0</sub>

c<sub>1</sub>

$$[a, b]$$

$$\eta_0 = 0.003 \quad M = 100 \quad k_1 = 0.005 \quad k_0 = 1.200 \quad c_1 = 0.015$$

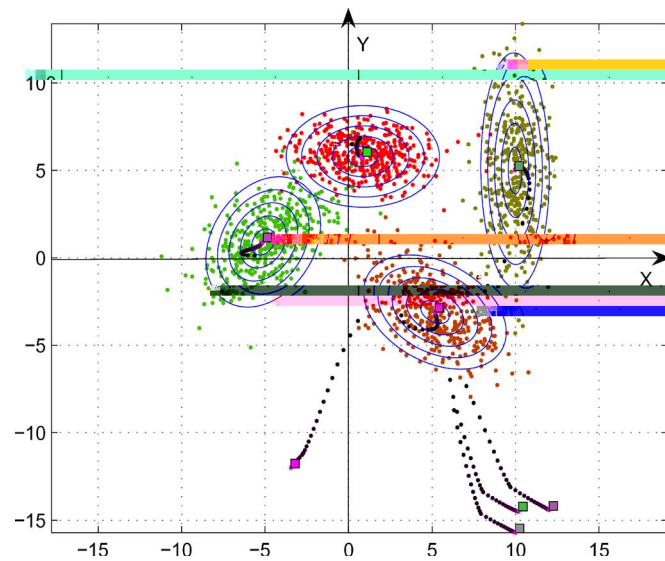
$$c_0 = 1.000 \quad [a, b] = [-1.2, 1.2]$$

$$T = 10000$$

$$w_{ij} \quad T \quad [a, b] \quad T = 0$$

$$\lambda = \exp(-k_1 T - k_0) \quad \eta = \eta_0 / (c_1 T + c_0) \quad t = 0$$

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



$n$   $k = 4$   $\eta$

-3

B. E  $U$   $C$

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

$n$   $n$   $k$

$E(\mathbf{W})$

$n$   $n$   $k$

$n$   $n$   $n$   $k$

$n$   $n$

$S_1$   $S_2$

$w_{ij}$   
[a, b]

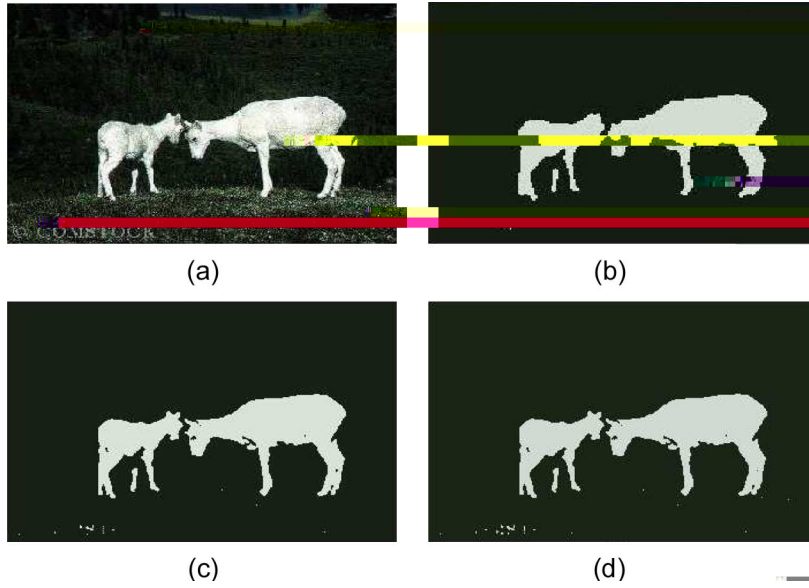


Figure 1: Segmentation results for the sheep image. (a) Original image with a red bar at the bottom and a yellow-green horizontal line across the sheep. (b) Segmentation mask with the same line. (c) Another segmentation mask. (d) Another segmentation mask. A legend at the bottom right shows a color bar with red, green, and blue segments.

Let  $\mathcal{I}$  be the set of all images in the dataset. For each image  $I \in \mathcal{I}$ , let  $\mathcal{S}_I$  be the set of all segmentation masks for  $I$ . Let  $\mathcal{S}$  be the set of all segmentation masks for all images in the dataset. Let  $k^*$  be the number of clusters in the segmentation mask. Let  $k$  be the number of clusters in the segmentation mask. Let  $k = 6$  be the number of clusters in the segmentation mask.

$$k^* = \dots$$

$$k = 6$$



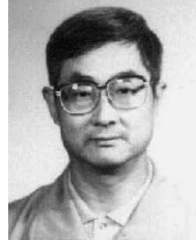
EEE N N  
: A A h  
EEE C  
N C  
A N  
N  
C  
C  
A  
N N  
N N  
C N N  
N N  
N N  
11: C  
EEE N N  
EEE N N  
N C  
h  
N N  
EEE N N  
EEE h h  
EEE h h  
EEE N N

EEE N N  
EEE  
N  
N  
EEE  
9: C N  
C N  
C N  
EEE C N N  
A  
N N  
N N  
N N  
N N  
EEE N N  
EEE N N  
N N  
N N  
D E  
h  
D N N  
h A  
A A h h & h  
19: C N h A F  
D U<sup>C</sup> h



**Jinwen Ma**

Dr. Jinwen Ma received his B.S. degree in Applied Mathematics from Tsinghua University, Beijing, China, in 2002, and his M.S. and Ph.D. degrees in Applied Mathematics from the University of California, San Diego, in 2004 and 2007, respectively. He was a postdoctoral fellow at the University of California, San Diego, from 2007 to 2009. He is currently an Assistant Professor at the Department of Mathematics, University of California, San Diego. His research interests include numerical methods for partial differential equations, computational fluid dynamics, and data science.



**Taijun Wang**

Dr. Taijun Wang received his B.S. degree in Applied Mathematics from Tsinghua University, Beijing, China, in 2002, and his M.S. and Ph.D. degrees in Applied Mathematics from the University of California, San Diego, in 2004 and 2007, respectively. He was a postdoctoral fellow at the University of California, San Diego, from 2007 to 2009. He is currently an Assistant Professor at the Department of Mathematics, University of California, San Diego. His research interests include numerical methods for partial differential equations, computational fluid dynamics, and data science.