Neural Computing and Applications (2021) 33:1171–1191
https://doi.org/10.1007/s00521-020-05038-8

ORIGINAL ARTICLE

GMM discriminant analysis with noisy label for each class

Jian-wei Liu[1] · Zheng-ping Ren[1] · Run-kun Lu[1] · Xiong-lin Luo[1]

Received: 8 June 2019 / Accepted: 13 May 2020 / Published online: 1 June 2020

Abstract

Keywords

1 Introduction

1.

2.

1.1.1 Description of the problem with the noise labels

$\mathbf{x} = (x_1, \ldots, x_d) \in \mathcal{X}, \quad \mathcal{X} = \mathcal{R}^d$

$\Omega = \{\omega_1, \omega_2, \ldots, \omega_K\}$. $\mathbf{x} \in \mathcal{X}$ $p(\mathbf{x}|\omega)$ $p(\omega),\ \omega \in \Omega.$

$\mathcal{S}_\omega$ $\omega \in \Omega$

$$\mathcal{S}_\omega = \{\mathbf{x} \in \mathcal{X}\},\ \omega \in \Omega\ ;\ \mathcal{S} = \bigcup_{\omega \in \Omega} \mathcal{S}_\omega,\ |\mathcal{S}| = \sum_{\omega \in \Omega} |\mathcal{S}_\omega|,$$

$|\mathcal{S}|$ $|\mathcal{S}_\omega|$ $\mathcal{S}_\omega$. $p(\mathbf{x}|\omega)$. L_ω

$$L_\omega = \frac{1}{|\mathcal{S}_\omega|} \sum_{\mathbf{x} \in \mathcal{S}_\omega} p(\mathbf{x}|\omega) p(\omega), \quad (\omega \in \Omega),$$

$1/|\mathcal{S}_\omega|$

$\omega \in \Omega$ $\mathbf{x} \in \mathcal{S}$ $\mathbf{x} \in \mathcal{S}$, $\tilde{\omega} \in \Omega$, ω. $p(\tilde{\omega})$ $\tilde{\omega}$ $p(\omega)$ ω. $p(\omega|\tilde{\omega})$ ω $\tilde{\omega}$. $\tilde{p}(\mathbf{x}|\tilde{\omega})$ $\mathbf{x} \in \mathcal{X}$ $\tilde{\omega}$

$$\tilde{p}(\mathbf{x}|\tilde{\omega}) = \sum_{\omega \in \Omega} p(\mathbf{x}|\omega) p(\omega|\tilde{\omega}), \mathbf{x} \in \mathcal{X},\ \tilde{\omega} \in \Omega,$$

$p(\omega|\tilde{\omega})$, $\omega \in \Omega$ $\mathbf{x} \in \mathcal{X}$. $p(\omega|\tilde{\omega}),\ \omega \in \Omega,\ p(\tilde{\omega}),$

$\tilde{\omega} \in \Omega$ r . r ,
fr .

1.1.2 Gaussian mixture model

r f [r r
r , r . [, 22
, , ,
f r . r r
r , r
r r , f r .
,

$L = \prod_{\mathbf{x} \in \mathcal{S}} p(\mathbf{x}|\tilde{\omega}, \theta_{\tilde{\omega}}) p$

$$\sum_{\omega\in\Omega} q(\omega) = e^{-\left(1+\lambda_{G_1}\right)} \sum_{\omega\in\Omega} p(\mathbf{x}|\omega,\theta_\omega)\gamma_{\tilde{\omega},\omega}\pi_\omega = 1$$

$$\Rightarrow \lambda_{G_1} = \sum_{\omega\in\Omega} p(\mathbf{x}|\omega,\theta_\omega)\gamma_{\tilde{\omega},\omega}\pi_\omega - 1.$$

λ_{G_1} ( ), $q(\omega)$

$$q(\omega) = \frac{p(\mathbf{x}|\omega,\theta_\omega)\gamma_{\tilde{\omega},\omega}\pi_\omega}{\sum_{\omega\in\Omega} p(\mathbf{x}|\omega,\theta_\omega)\gamma_{\tilde{\omega},\omega}\pi_\omega} = p(\omega|x,\tilde{\omega}) \qquad ( )$$

$p(\mathbf{x}|\omega,\theta_\omega) = \sum_{m\in\mathcal{M}} w_{\omega,m} g\left(\mathbf{x}\,|\,\omega,\boldsymbol{\mu}_{\omega,m},\boldsymbol{\Sigma}_{\omega,m}\right)$

$$\begin{aligned} p(\mathbf{x}|\omega,\theta_\omega) &= \sum_{m\in\mathcal{M}} w_{\omega,m} g\left(\mathbf{x}\,|\,\omega,\boldsymbol{\mu}_{\omega,m},\boldsymbol{\Sigma}_{\omega,m}\right) \\ &= \sum_{m\in\mathcal{M}} h_{\omega,m} \frac{w_{\omega,m} g\left(\mathbf{x}\,|\,\omega,\boldsymbol{\mu}_{\omega,m},\boldsymbol{\Sigma}_{\omega,m}\right)}{h_{\omega,m}} \\ &\geq \sum_{m\in\mathcal{M}} h_{\omega,m} \frac{w_{\omega,m} g\left(\mathbf{x}\,|\,\omega,\boldsymbol{\mu}_{\omega,m},\boldsymbol{\Sigma}_{\omega,m}\right)}{h_{\omega,m}} \\ &= \sum_{m\in\mathcal{M}} h_{\omega,m} \left(w_{\omega,m} g\left(\mathbf{x}\,|\,\omega,\boldsymbol{\mu}_{\omega,m},\boldsymbol{\Sigma}_{\omega,m}\right) - h_{\omega,m} \right) \end{aligned}$$

$\sum_{m\in\mathcal{M}} h_{\omega,m} = 1.$ λ_{G_2} $h_{\omega,m}$,

$$G_2 = \sum_{m\in\mathcal{M}} h_{\omega,m} \left(w_{\omega,m} g\left(\mathbf{x}\,|\,\omega,\boldsymbol{\mu}_{\omega,m},\boldsymbol{\Sigma}_{\omega,m}\right) - h_{\omega,m} \right) + \lambda_{G_2}\left(1 - \sum_{m\in\mathcal{M}} h_{\omega,m}\right)$$

$h_{\omega,m}$

$$w_{\omega,m} g\left(\mathbf{x}\,|\,\omega,\boldsymbol{\mu}_{\omega,m},\boldsymbol{\Sigma}_{\omega,m}\right) - h_{\omega,m} - 1 - \lambda_{\lambda_{G_2}} = 0$$

$$\Rightarrow h_{\omega,m} = w_{\omega,m} g\left(\mathbf{x}\,|\,\omega,\boldsymbol{\mu}_{\omega,m},\boldsymbol{\Sigma}_{\omega,m}\right) - 1 - \lambda_{\lambda_{G_2}}$$

$$\Rightarrow h_{\omega,m} = w_{\omega,m} g\left(\mathbf{x}\,|\,\omega,\boldsymbol{\mu}_{\omega,m},\boldsymbol{\Sigma}_{\omega,m}\right) \cdot e^{-\left(1+\lambda_{\lambda_{G_2}}\right)}$$

$h_{\omega,m}$ λ_{G_2}

$$\sum_{m\in\mathcal{M}} h_{\omega,m} = \sum_{m\in\mathcal{M}} w_{\omega,m} g\left(\mathbf{x}\,|\,\omega,\boldsymbol{\mu}_{\omega,m},\boldsymbol{\Sigma}_{\omega,m}\right) \cdot e^{-\left(1+\lambda_{G_2}\right)}$$

$$\Rightarrow \lambda_{G_2} = \sum_{m\in\mathcal{M}} w_{\omega,m} g\left(\mathbf{x}\,|\,\omega,\boldsymbol{\mu}_{\omega,m},\boldsymbol{\Sigma}_{\omega,m}\right) - 1.$$

λ_{G_2} $h_{\omega,m}$

$$h_{\omega,m} = \frac{w_{\omega,m} g\left(\mathbf{x}\,|\,\omega,\boldsymbol{\mu}_{\omega,m},\boldsymbol{\Sigma}_{\omega,m}\right)}{\sum_{m\in\mathcal{M}} w_{\omega,m} g\left(\mathbf{x}\,|\,\omega,\boldsymbol{\mu}_{\omega,m},\boldsymbol{\Sigma}_{\omega,m}\right)} = r(m|\mathbf{x},\omega) \qquad ( )$$

( ) ( ) ( )

$$Q = \sum_{\tilde{\omega}\in\Omega} I(\tilde{\omega}=\omega) \sum_{\mathbf{x}\in\mathcal{S}} \sum_{\omega\in\Omega} \left(p(\omega|\mathbf{x},\tilde{\omega}) \times \sum_{m\in\mathcal{M}} r(m|\mathbf{x},\omega) \frac{w_{\omega,m} g\left(\mathbf{x}\,|\,\omega,\boldsymbol{\mu}_{\omega,m},\boldsymbol{\Sigma}_{\omega,m}\right)}{r(m|\mathbf{x},\omega)} \right) + \sum_{\mathbf{x}\in\mathcal{S}} \sum_{\omega\in\Omega} p(\omega|\mathbf{x},\tilde{\omega}) \gamma_{\tilde{\omega},\omega} + \sum_{\mathbf{x}\in\mathcal{S}} \sum_{\omega\in\Omega} p(\omega|\mathbf{x},\tilde{\omega}) \pi_\omega \qquad ( )$$

$\Theta = \{\theta_\omega\}_{\omega\in\Omega}$, $\Gamma = \left\{\gamma_{\tilde{\omega},\omega}\right\}_{\tilde{\omega}\in\Omega}$, $\Pi = \{\pi_\omega\}_{\omega\in\Omega}$, $\theta_\omega = \left\{w_{\omega,m},\boldsymbol{\mu}_{\omega,m},\boldsymbol{\Sigma}_{\omega,m}\right\}_{m\in\mathcal{M}}$.

E step $p(\omega|\mathbf{x},\tilde{\omega})$ $r(m|\mathbf{x},\omega)$ ( ) ( ),

$$p(\omega|\mathbf{x},\tilde{\omega}) = \frac{p(\mathbf{x}|\omega,\theta_\omega)\gamma_{\tilde{\omega},\omega}\pi_\omega}{\sum_{\omega\in\Omega} p(\mathbf{x}|\omega,\theta_\omega)\gamma_{\tilde{\omega},\omega}\pi_\omega} \qquad ( )$$

$$r(m|\mathbf{x}_n,\omega_n = k) = \frac{w_{\omega,m} g\left(\mathbf{x}\,|\,\omega,\boldsymbol{\mu}_{\omega,m},\boldsymbol{\Sigma}_{\omega,m}\right)}{\sum_{m\in\mathcal{M}} w_{\omega,m} g\left(\mathbf{x}\,|\,\omega,\boldsymbol{\mu}_{\omega,m},\boldsymbol{\Sigma}_{\omega,m}\right)} \qquad ( )$$

M step ( ) $\Theta = \{\theta_\omega\}_{\omega\in\Omega}$, $\Gamma = \left\{\gamma_{\tilde{\omega},\omega}\right\}_{\tilde{\omega}\in\Omega}$, $\Pi = \{\pi_\omega\}_{\omega\in\Omega}$.

1.1.3 Updating

$\boldsymbol{\mu}_{\omega,m}$

$$\frac{\partial Q}{\partial \boldsymbol{\Sigma}_{\omega,m}} = -\sum_{\tilde{\omega}\in\Omega} I(\tilde{\omega}=\omega) \sum_{\mathbf{x}\in\mathcal{S}} \Big\{ p(\omega|\mathbf{x},\tilde{\omega}) \Pr(m|\mathbf{x},\omega) \cdot \Big[\boldsymbol{\Sigma}^{-1}_{\omega,m} - \boldsymbol{\Sigma}^{-1}_{\omega,m} (\mathbf{x}-\boldsymbol{\mu}_{\omega,m})(\mathbf{x}-\boldsymbol{\mu}_{\omega,m})^T \boldsymbol{\Sigma}^{-1}_{\omega,m} \Big] \Big\} = 0 \Rightarrow \boldsymbol{\Sigma}_{\omega,m} = \frac{\sum_{\tilde{\omega}\in\Omega} I(\tilde{\omega}=\omega) \sum_{\mathbf{x}\in\mathcal{S}} p(\omega|\mathbf{x},\tilde{\omega}) \Pr(m|\mathbf{x},\omega) (\mathbf{x}-\boldsymbol{\mu}_{\omega,m})(\mathbf{x}-\boldsymbol{\mu}_{\omega,m})^T}{\sum_{\tilde{\omega}\in\Omega} I(\tilde{\omega}=\omega) \sum_{\mathbf{x}\in\mathcal{S}} p(\omega|\mathbf{x},\tilde{\omega}) \Pr(m|\mathbf{x},\omega)} \quad (11)$$

$w_{\omega,m}$ … $\lambda_{w_{\omega,m}}$ … $\sum_{m\in\mathcal{M}} w_{\omega,m} = 1$,

$$Q_{\lambda_{w_{\omega,m}}} = Q + \lambda_{w_{\omega,m}} \left(1 - \sum_{m\in\mathcal{M}} w_{\omega,m} \right),$$

$Q_{\lambda_{w_{\omega,m}}}$ w.r.t. $w_{\omega,m}$

$$\frac{\partial Q_{\lambda_{w_{\omega,m}}}}{\partial w_{\omega,m}} = 0 = \sum_{\tilde{\omega}\in\Omega} I(\tilde{\omega}=\omega) \sum_{\mathbf{x}\in\mathcal{S}} p(\omega|\mathbf{x},\tilde{\omega}) \Pr(m|\mathbf{x},\omega) \frac{1}{w_{\omega,m}} - \lambda_{w_{\omega,m}}$$

$$\Rightarrow \lambda_{w_{\omega,m}} w_{\omega,m} = \sum_{\tilde{\omega}\in\Omega} I(\tilde{\omega}=\omega) \sum_{\mathbf{x}\in\mathcal{S}} p(\omega|\mathbf{x},\tilde{\omega}) \Pr(m|\mathbf{x},\omega)$$

$$\overset{\text{sum over } w_{\omega,m}}{\Rightarrow} \lambda_{w_{\omega,m}} \sum_{m\in\mathcal{M}} w_{\omega,m} = \lambda_{w_{\omega,m}} = \sum_{\tilde{\omega}\in\Omega} I(\tilde{\omega}=\omega) \sum_{\mathbf{x}\in\mathcal{S}} p(\omega|\mathbf{x},\tilde{\omega}) \sum_{m\in\mathcal{M}} \Pr(m|\mathbf{x},\omega),$$

$\lambda_{w_{\omega,m}}$ … $w_{\omega,m}$

$$w_{\omega,m} = \frac{\sum_{\tilde{\omega}\in\Omega} I(\tilde{\omega}=\omega) \sum_{\mathbf{x}\in\mathcal{S}} p(\omega|\mathbf{x},\tilde{\omega}) \Pr(m|\mathbf{x},\omega)}{\sum_{\tilde{\omega}\in\Omega} I(\tilde{\omega}=\omega) \sum_{\mathbf{x}\in\mathcal{S}} p(\omega|\mathbf{x},\tilde{\omega}) \sum_{m\in\mathcal{M}} \Pr(m|\mathbf{x},\omega)} \quad (12)$$

$\gamma_{\tilde{\omega},\omega}$ … $\lambda_{\gamma_{\tilde{\omega},\omega}}$ … $\sum_{\tilde{\omega}\in\Omega} \gamma_{\tilde{\omega},\omega} = 1$,

$$Q_{\gamma_{\tilde{\omega},\omega}} = Q + \lambda_{\gamma_{\tilde{\omega},\omega}} \left(1 - \sum_{\tilde{\omega}\in\Omega} \gamma_{\tilde{\omega},\omega} \right),$$

$Q_{\gamma_{\tilde{\omega},\omega}}$ w.r.t. $\gamma_{\tilde{\omega},\omega}$

$$\frac{\partial Q_{\gamma_{\tilde{\omega},\omega}}}{\partial \gamma_{\tilde{\omega},\omega}} = \sum_{\mathbf{x}\in\mathcal{S}} p(\omega|\mathbf{x},\tilde{\omega}) \frac{1}{\gamma_{\tilde{\omega},\omega}} - \lambda_{\gamma_{\tilde{\omega},\omega}} = 0$$

$$\Rightarrow \lambda_{\gamma_{\tilde{\omega},\omega}} \gamma_{\tilde{\omega},\omega} = \sum_{\mathbf{x}\in\mathcal{S}} p(\omega|\mathbf{x},\tilde{\omega}),$$

$$\overset{\text{sum over } \gamma_{\tilde{\omega},\omega}}{\Rightarrow} \lambda_{\gamma_{\tilde{\omega},\omega}} \sum_{\tilde{\omega}\in\Omega} \gamma_{\tilde{\omega},\omega} = \lambda_{\gamma_{\tilde{\omega},\omega}} = \sum_{\mathbf{x}\in\mathcal{S}} \sum_{\tilde{\omega}\in\Omega} p(\omega|\mathbf{x},\tilde{\omega})$$

$\lambda_{\gamma_{\tilde{\omega},\omega}}$ … $\gamma_{\tilde{\omega},\omega}$

$$\gamma_{\tilde{\omega},\omega} = \frac{\sum_{\mathbf{x}\in\mathcal{S}} p(\omega|\mathbf{x},\tilde{\omega})}{\sum_{\mathbf{x}\in\mathcal{S}} \sum_{\tilde{\omega}\in\Omega} p(\omega|\mathbf{x},\tilde{\omega})}. \quad (1\ )$$

λ_{π_ω} … $\sum_{\omega\in\Omega} \pi_\omega = 1$,

∂Q

$$L' - L = \sum_{\omega \in \Omega} \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} q(\omega|\mathbf{x}, \psi(\mathbf{x})) \; \frac{p'(\mathbf{x}|\omega)p'(\omega|\psi(\mathbf{x}))}{p(\mathbf{x}|\omega)p(\omega|\psi(\mathbf{x}))} + \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} \sum_{\omega \in \Omega} q(\omega|\mathbf{x}, \psi(\mathbf{x})) \; \frac{q(\omega|\mathbf{x}, \psi(\mathbf{x}))}{q'(\omega|\mathbf{x}, \psi(\mathbf{x}))}. \quad (2\ )$$

, . (2) . . - - . . K , r . f r . r.

$$I(q, q') = \sum_{\omega \in \Omega} q(\omega|\mathbf{x}, \psi(\mathbf{x})) \; \frac{q(\omega|\mathbf{x}, \psi(\mathbf{x}))}{q'(\omega|\mathbf{x}, \psi(\mathbf{x}))} \geq 0, \quad (2\ )$$

r f r , r.

$$L' - L \geq \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} \sum_{\omega \in \Omega} q(\omega|\mathbf{x}, \psi(\mathbf{x})) \; \frac{p'(\mathbf{x}|\omega)p'(\omega|\psi(\mathbf{x}))}{p(\mathbf{x}|\omega)p(\omega|\psi(\mathbf{x}))}. \quad (2\ )$$

, , f r f r r f . r , . , r ,

$$L' - L \geq \sum_{\omega \in \Omega} \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} q(\omega|\mathbf{x}, \psi(\mathbf{x})) \; \frac{p'(\mathbf{x}|\omega)}{p(\mathbf{x}|\omega)} + \sum_{\omega \in \Omega} \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} q(\omega|\mathbf{x}, \psi(\mathbf{x})) \; \frac{p'(\omega|\psi(\mathbf{x}))}{p(\omega|\psi(\mathbf{x}))} \geq 0. \quad (2\ )$$

r , , r $x \in \mathcal{S}$ r , . . r $\mathbf{x} \in \mathcal{S}_{\tilde{\omega}}$

$$L' - L = \sum_{\omega \in \Omega} \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} q(\omega|\mathbf{x}, \psi(\mathbf{x})) \; \frac{p'(\mathbf{x}|\omega)}{p(\mathbf{x}|\omega)} + \sum_{\tilde{\omega} \in \Omega} \frac{|\mathcal{S}_{\tilde{\omega}}|}{|\mathcal{S}|} \sum_{\omega \in \Omega} \left[\frac{1}{|\mathcal{S}_{\tilde{\omega}}|} \sum_{\mathbf{x} \in \mathcal{S}_{\tilde{\omega}}} q(\omega|\mathbf{x}, \tilde{\omega}) \right] \frac{p'(\omega|\tilde{\omega})}{p(\omega|\tilde{\omega})}. \quad (2\ )$$

, f r , f r

$$p'(\omega|\tilde{\omega}) = \frac{1}{|\mathcal{S}_{\tilde{\omega}}|} \sum_{\mathbf{x} \in \mathcal{S}_{\tilde{\omega}}} q(\omega|\mathbf{x}, \tilde{\omega}), \; \omega \in \Omega, \tilde{\omega} \in \Omega \quad (2\ )$$

$$p'(\cdot|\omega) = \text{r}_{p(\cdot|\omega)} \left(\frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} q(\omega|\mathbf{x}, \psi(\mathbf{x})) \; p(\mathbf{x}|\omega) \right) \omega \in \Omega, \quad (2\ )$$

r r f r . , r , , , (2), r (2) . , f . K , r f r r.

$$\sum_{\omega \in \Omega} p'(\omega|\tilde{\omega}) \; \frac{p'(\omega|\tilde{\omega})}{p(\omega|\tilde{\omega})} \geq 0, \tilde{\omega} \in \Omega, \quad (\ 0)$$

. (2) ,

$$\frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} q(\omega|\mathbf{x}, \psi(\mathbf{x})) \; p'(\mathbf{x}|\omega) \geq \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} q(\omega|\mathbf{x}, \psi(\mathbf{x})) \; p(\mathbf{x}|\omega), \omega \in \Omega,$$

r r f r

$$\frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} q(\omega|\mathbf{x}, \psi(\mathbf{x})) \; \frac{p'(\mathbf{x}|\omega)}{p(\mathbf{x}|\omega)} \geq 0, \omega \in \Omega. \quad (\ 1)$$

, , r r (2) f , (0), (1), r , r f r (2), (2), (2) r r f r [2 , 2 .

2.1.1 Gaussian Classes with Noisy Labels

, r , r , .., , - -

$$p(\mathbf{x}|\omega) = f(\mathbf{x}|\boldsymbol{\mu}_\omega, \boldsymbol{\Sigma}_\omega), \; \omega \in \Omega, \quad (\ 2)$$

r . (2) r f r

$$\left\{ \boldsymbol{\mu}'_\omega, \boldsymbol{\Sigma}'_\omega \right\} = \text{r}_{\{\mu_\omega, \Sigma_\omega\}} \left[\frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} q(\omega|\mathbf{x}, \psi(\mathbf{x})) \; f(\mathbf{x}|\boldsymbol{\mu}_\omega, \boldsymbol{\Sigma}_\omega) \right] \omega \in \Omega. \quad (\ )$$

r . ( ) f , r f [2 , r f r , . r , r, $F(\mathbf{x}|\boldsymbol{\mu})$ r r r $\boldsymbol{\mu}$ r , $\hat{\boldsymbol{\mu}}$

$$L_{\boldsymbol{\mu}} = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} f(\mathbf{x}|\boldsymbol{\mu}) \rightarrow \;\; \Rightarrow \hat{\mu} = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x} \in \mathcal{S}} \mathbf{x}. \quad (\ )$$

f $N(\mathbf{x})$, r f r , rr f $\mathbf{x} \in \mathcal{X}$ $\mathcal{S}$ $q(\mathbf{x})$ r fr , f $\mathbf{x} \in \mathcal{S}$

$$q(\mathbf{x}) = \frac{N(\mathbf{x})}{|\mathcal{S}|}, \sum_{\mathbf{x} \in \mathcal{X}} q(\mathbf{x}) = 1, (\mathbf{x} \notin \mathcal{S} \Rightarrow q(\mathbf{x}) = 0),$$

. ( ) r r , f r

$$L_{\boldsymbol{\mu}} = \sum_{\mathbf{x} \in \mathcal{X}} q(\mathbf{x}) \; F(\mathbf{x}|\boldsymbol{\mu}) \rightarrow \;\; \Rightarrow \hat{\mu} = \sum_{\mathbf{x} \in \mathcal{X}} q(\mathbf{x}) \, \mathbf{x}. \quad (\ )$$

r r f . ( ) ( ), f ( )

rr (f r r f [2]). f ( ), r

$$\mu'_\omega = \frac{1}{\sum_{\mathbf{x}\in\mathcal{S}} q(\omega|\mathbf{x},\psi(\mathbf{x}))} \sum_{\mathbf{x}\in\mathcal{S}} q(\omega|\mathbf{x},\psi(\mathbf{x}))\,\mathbf{x}, \quad \omega \in \Omega \qquad (\ )$$

$$\Sigma'_\omega = \frac{1}{\sum_{\mathbf{x}\in\mathcal{S}} q(\omega|\mathbf{x},\psi(\mathbf{x}))} \sum_{\mathbf{x}\in\mathcal{S}} q(\omega|\mathbf{x},\psi(\mathbf{x}))\mathbf{x}\mathbf{x}^T - \mu'_\omega \mu'^T_\omega, \quad \omega \in \Omega. \qquad (\ )$$

r f r r f r r r . (1), (2), ( ), ( ).

2.1.2 Class-conditional Gaussian Mixtures with Noisy Labels

(2) r r r r r r $p(\mathbf{x}|\omega)$ r . r r,

$$p(\mathbf{x}|\omega) = \sum_{m\in\mathcal{M}_\omega} w_{m\omega} f(\mathbf{x}|\mu_{m\omega}, \Sigma_{m\omega}), \quad \sum_{m\in\mathcal{M}_\omega} w_{m\omega} = 1, \ \omega \in \Omega \qquad (\ )$$

r $m \in \mathcal{M}_\omega$ f r $P(\mathbf{x}|\omega)$. ( ) (1), f f r r f r

$$L = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x}\in\mathcal{S}} \left[\sum_{\omega\in\Omega} \sum_{m\in\mathcal{M}_\omega} p(\omega|\psi(\mathbf{x}))\, w_{m\omega} f(\mathbf{x}|\mu_{m\omega}, \Sigma_{m\omega}) \right].$$

f r

$$h(m,\omega|\mathbf{x},\psi(\mathbf{x})) = \frac{p(\omega|\psi(\mathbf{x}))w_{m\omega} f(\mathbf{x}|\mu_{m\omega},\Sigma_{m\omega})}{\sum_{\omega\in\Omega}\sum_{m\in\mathcal{M}_\omega} p(\omega|\psi(\mathbf{x}))\, w_{m\omega} f(\mathbf{x}|\mu_{m\omega}}$$

$$p'(\omega|\tilde{\omega}) = \frac{1}{|\mathcal{S}_{\tilde{\omega}}|} \sum_{\mathbf{x}\in\mathcal{S}_{\tilde{\omega}}} \sum_{m\in\mathcal{M}_{\omega}} h(m,\omega|\mathbf{x},\tilde{\omega}), \omega \in \Omega,\ \tilde{\omega} \in \Omega. \quad (\ )$$

$$w'_{m\omega} = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x}\in\mathcal{S}} h(m|\omega,\mathbf{x},\psi(\mathbf{x})) = \frac{1}{|\mathcal{S}|} \sum_{\mathbf{x}\in\mathcal{S}} \frac{h(m,\omega|\mathbf{x},\psi(\mathbf{x}))}{\sum_{m\in\mathcal{M}_{\omega}} h(m,\omega|\mathbf{x},\psi(\mathbf{x}))} \quad m \in \mathcal{M}_{\omega}, \omega \in \Omega, \quad (\ )$$

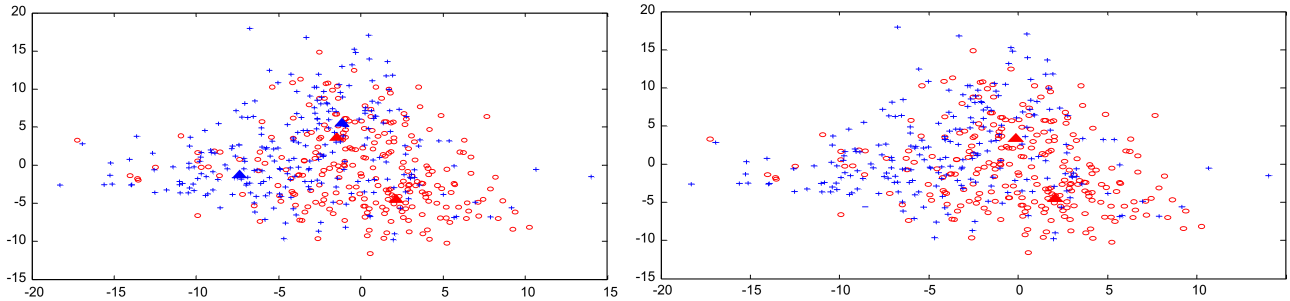
3 Related work

Table 1

1	2000	0	
2	1000	2	2
r	10		
r	1 0		
	1	1	
r	2	22	2
	0	1	2
f r	000	21	

[illegible]

[illegible]



4 Experiments and discussion

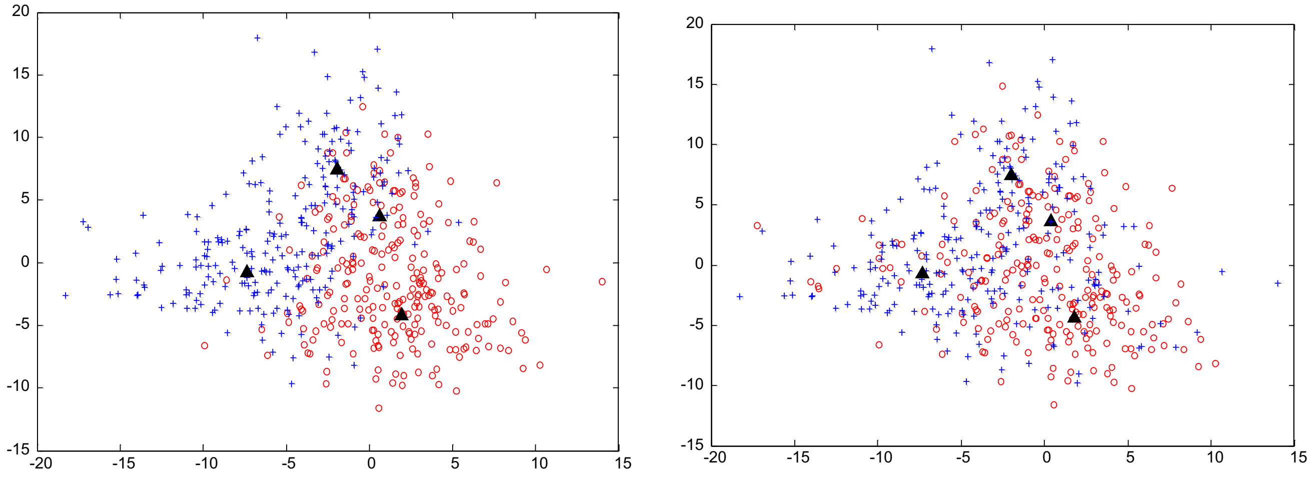


Fig. 5 2 r r r

Table 2 2 r

	r r	rr r r (%)
r 2	1 2. $e-0$ .0 -0 1	[0. 0. 00] 1 .00
2 20% rr r	0. 2 0.21 0.1 0. 02	[0. 101 0.] 12. 0

ff . f r r r f 0. . r ff r r r . . r 0. . r r r r r , , rr r r f r r , r . r r 1 .00% 12. 0%, r ff r r , r r r fr 20 0% r r r r ff f f r rf r f r . r ff f r f r rr r r f r f r r r ff r f r f r , rf r , , . r r 0% . r f r , r r rr r r . 0% . r r 2 ff rr r r r r rr r r . r r r r 20% rr r 2 rf r . r r r r . 2 rf r rr r r r rr r r f r rf r r r r 2 2 r . r r r 20% rr r . $\gamma_{ij}(i=j)$ r f r r f r . r . 2 r f r r , . r 2 r r 1, r f rf r r 20% rr r r 0. , r 1 r r r

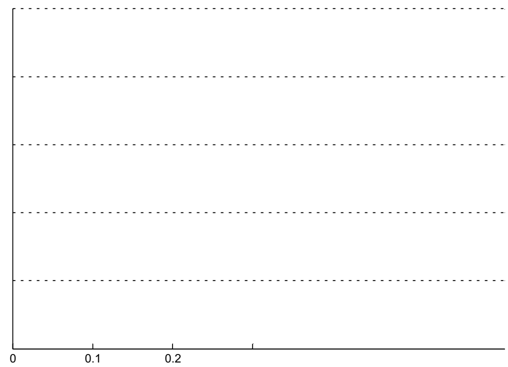

0
0.1
0.2

Table 3 r rr r r r - r ff r r r

		r r					
		0	0.1	0.2	0.	0.	0.
1		0.0	0.0	0.0 1	0.0	0.11	0.22
	r	0.0	0.0	0.0	0.0	0.0	0.0
		0.005	**0.003**	**0.003**	**0.003**	**0.030**	**0.003**
2	r	0.1 2	0.1	0.1	0.1	0. 0	0. 0
	r	0.1	0.1	0.1	0.1	**0.186**	0. 0
		0.1 2	0.1	0.1 2	0.1	0.22	0. 1
	r	0.1	0.1	0.1	0.2	0. 2	0. 2
		0.128	**0.120**	**0.126**	**0.166**	0.1	**0.182**
,		0.	0.	0. 1	0. 2	0. 1	0. 1
r	r	**0.305**	0. 0	**0.357**	0.	0. 0	0. 2
		0.	**0.238**	**0.357**	**0.357**	**0.500**	**0.504**
r		0.02	0.0	0.1	0.200	0.	0.2
	r	0.02	0.0	**0.033**	0.0	0.100	**0.08**
		0.013	**0.016**	**0.033**	**0.05**	**0.083**	**0.08**
		0.0	0.12	0.112	0.1	0.2	0. 0
	r	0.0	**0.011**	**0.044**	0.0	**0.044**	0.0
		0.033	0.022	**0.044**	**0.033**	0.0	**0.076**
r	r	0.	0.	0. 1	0.	0.	0.
	r	0.	0.	0. 2	0.	0.	0.
		0.2 0	**0.261**	0.2	0.	0. 2	0.
	r	0. 0	0. 2	0. 0	0.	0.	0.
		0.248	**0.261**	**0.261**	**0.289**	**0.287**	**0.271**
	r	**0.130**	0.1	0.12	0.1	0.	0. 1
	r	**0.130**	**0.183**	**0.113**	0.1	0. 1	0.
		0.1	0.1	0.1	**0.183**	0.2 2	0.
	r	0.2	0.1	0.1	0.1	0.	0. 1
		0.20	**0.183**	0.1	**0.183**	**0.188**	**0.185**
		0.188	**0.201**	**0.223**	**0.223**	0.2 0	0. 12
f r	r	0.2	0.2	0.2	0.2 1	0.2 2	0.2
		0.222	0.22	0.2 1	0.2 2	**0.244**	**0.244**

rf r r

Table 4 r rr r r r - r ff r r r

		r r					
		0	0.1	0.2	0.	0.	0.
1		0.0	0.0	0.0	0.1	0.211	0.221
	r	0.0	0.0	0.0	0.0	0.0	0.
		0.005	**0.005**	**0.005**	**0.005**	**0.026**	**0.191**
2	r	0.1 2	0.1	0.1	0.1	0.1 2	–
	r	0.1	0.1	0.1 2	0.1	0.2	–
		0.1 2	0.1 2	0.2	0.	0.	–
	r	0.1	**0.128**	0.1	0.1	0.1	**0.504**
		0.128	**0.128**	**0.132**	**0.13**	**0.144**	0. 2
r		0.	0.	0.	0.	0. 2	0. 0
,	r	**0.305**	**0.238**	0.2	0.	0. 2	0.
		0.	0.2 1	**0.214**	**0.25**	**0.404**	**0.415**
r		0.02	0.01	0.1	0.1	0.	–
	r	0.02	**0.016**	**0.016**	**0.022**	**0.033**	–
		0.013	**0.016**	**0.016**	**0.022**	**0.033**	–
		0.0	0.1 0	0.112	0.1	0. 0	–
	r	0.0	0.0	0.0	**0.042**	**0.056**	–
		0.033	**0.042**	**0.042**	**0.042**	**0.056**	–
r	r	0.	0.	0.	0.	0. 00	–
	r	0.	0. 20	0. 11	0.	0. 2	–
		0.2 0	0.2	0. 0	0. 2	0. 1	–
	r	0. 0	0.	0.2 1	0.	0. 2	–
		0.248	**0.261**	**0.261**	**0.261**	**0.327**	–
	r	**0.130**	**0.163**	**0.173**	**0.178**	0.1	0.
	r	**0.130**	**0.163**	0.1	0.1	0.1	0.
		0.1	0.1	0.2 2	0. 1	0.	0. 2
	r	0.2	0.1	0.1	0.1	0.	0.2
		0.20	0.1	0.1	0.1	**0.163**	**0.221**
		0.188	**0.196**	0.2	0.	0. 0	–
f r	r	0.2	0.2	0.2 1	0.	0. 00	–
		0.222	0.22	**0.226**	**0.280**	**0.296**	–

rf r r

r r r ff f -
f . r f .
f 1000 , r , fr 0.
, 20% rr r r r
.
, r r , ff r
ff r , ff r . r
. r , r rr r r .
r rf r , r r
.
f , ff
f , r , r f f

, r f .
1000 r r
, , r fr 1 f r .
, r r
, r r rr r r . ,
r , rf r r
.
r . , , , r
r rf r , -
- . r r r
ff f , r f . , r
r r 1000

Table 5 / r /

		/ r /	
		r	r
1		0/0/	0/0/
	r	0/0/	0/0/
		/0/0	/0/0
2	r	0/0/	0/0/
	r	1/0/	0/0/
		0/0/	0/0/
	r	0/0/	1/1/
		/0/1	/1/1
r		0/0/	0/0/
,	r	1/1/	2/0/
		/1/1	/0/2
r		0/0/	0/0/
	r	0/ /	0/ /1
		/ /0	1/ /0
		0/0/	0/0/
	r	1/2/	0/2/
		/2/1	/2/0
r	r	0/0/	0/0/
	r	0/0/	0/0/
		0/1/	0/0/
	r	0/0/	0/0/
		/1/0	/0/0
	r	1/1/	1/2/
	r	1/1/	1/2/
		0/1/	0/0/
	r	0/0/	0/0/
		2/1/	2/0/
f r		/0/2	2/0/
	r	0/0/	0/0/
		2/0/	/0/2

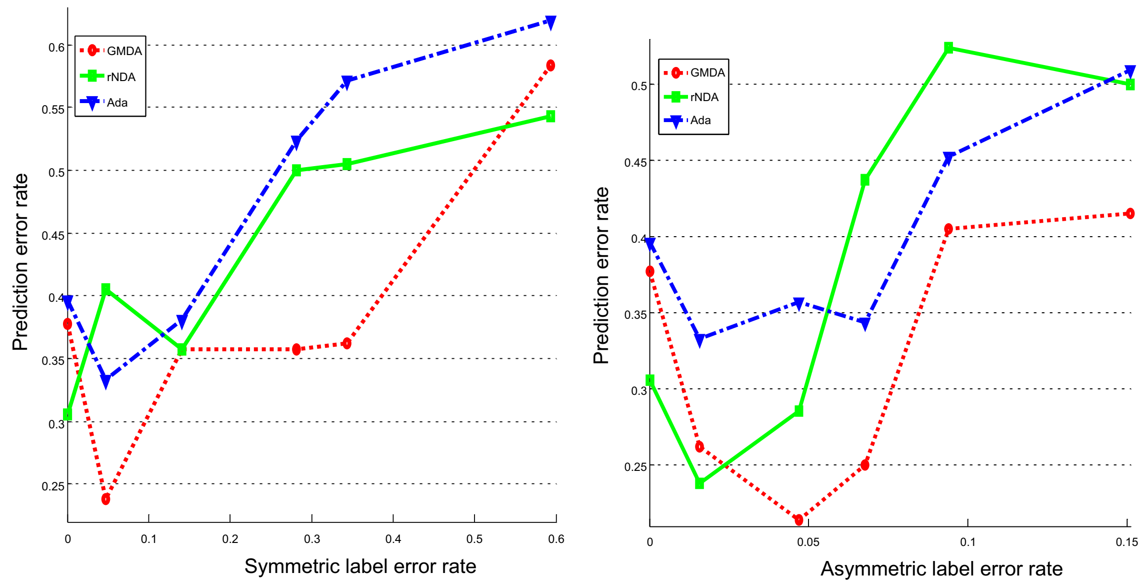
. , r , r r ff , r . , r r , r . r f rr - , r , r. r, , r . r - r r , r f fr 1 . r . 10, r , r , r r rr r r , r r rr r r f 0. 0 0 0. 0, r . , r r- f r r r r , r , r , . r rf r , r f 1.

, , r r . , , r r - , r , r r . , r r r , r rf r ff rr r r f r . r r, ff r , , r , , . r r , f r , r f, , . , r r r , r f , r r , . , , f , r f , r , rr , .

5 Experimental results on large-scale datasets

r f rf r f , r r r r. - , , ff f , r , r f f , r f . 1 ,000 200 - r , , r fr 1 f r . r r , r f f , r r . r - , r , r r f r r . , r f f , r . fr 0 2 0, rr f fr , rr , rr . r , f r f r , rr , f , f , r f r - , r . , 0%, 10%, 20%, 0%, 0%, 0% f , r f . rr r r , 10, r - f r r . , (1), (1), (1) , r rr r r rr . r , , (2), (2), (2) , r rr r r , rr . r , . , r 11 , r . rf r f r r . r , . . 11, (1), (1), (1) r r . rf r rr . r , , (2), (2), (2) r r . rf r- , rr . r , . r . . 11,

( ) , r f f . r r. . r



(g) **(h)**

Fig. 6

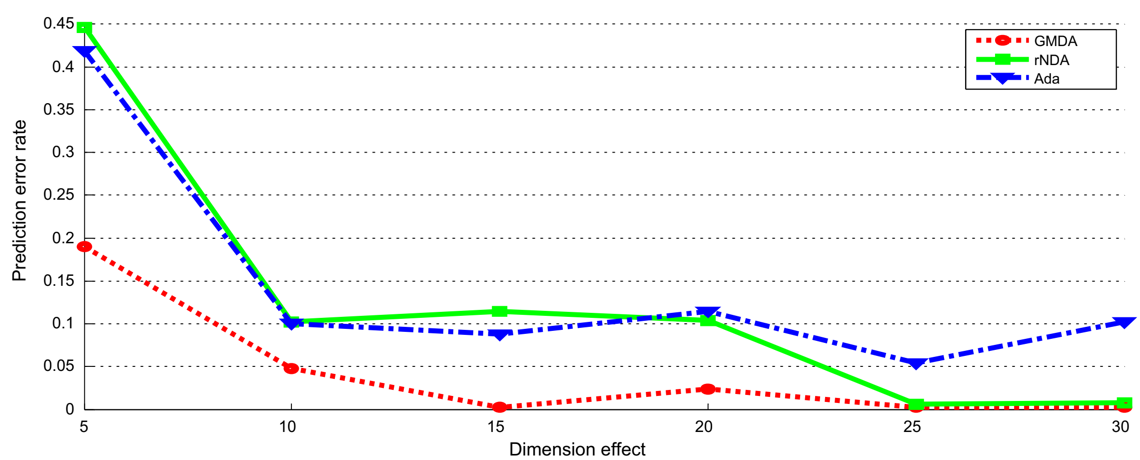


Fig. 7 ff f f

Fig. 8 ff f , r f f (, r f , r)

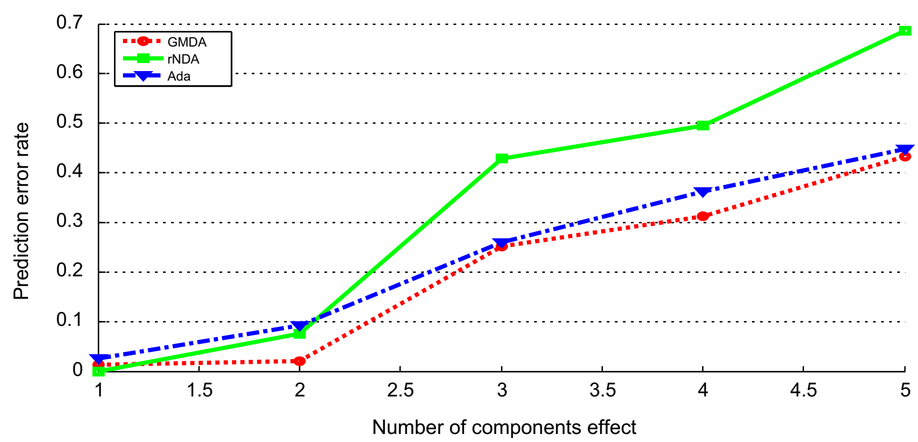


Fig. 9 ff f , r f

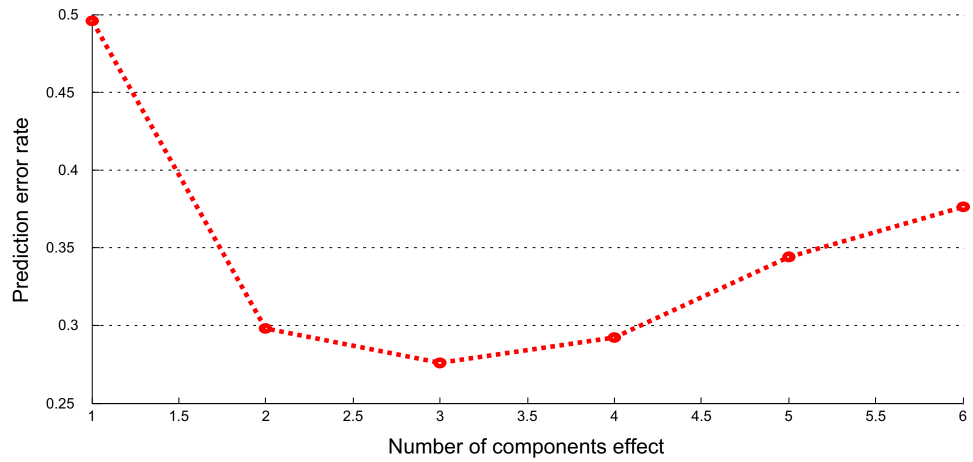


Fig. 10 ff f , r f

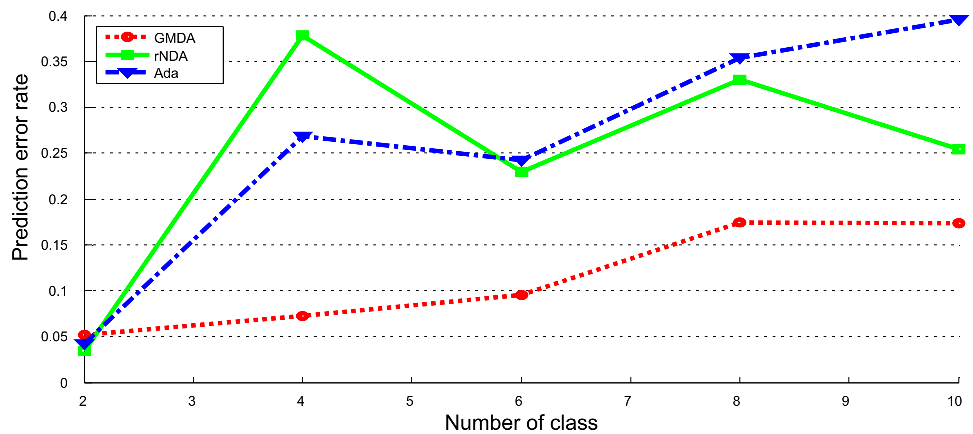


Table 6 r r f r -

	r r					
	N	*D*	r	r -	, r f 1	, r f 2
1	1 ,000	200	2 0		2	0
2	1 ,000	200	0		1	1
	1 ,000	200	2 0	10		
	1 ,000	200	0	10	1	1
	1 ,000	200	2 0		0	2
	1 ,000	200	0			20

r , r r , , r
r r . r r r , ,
( ) r r ff r , . , ,
, , r r r ,
, 2 0, ff f r , , ff , f
r , . r r rf r f r ,
r , r f ( ) r r ff r
r , rr r r , r , , r
0%. , r , , 0, f
rr r r f r , 1%. ff r r f r .

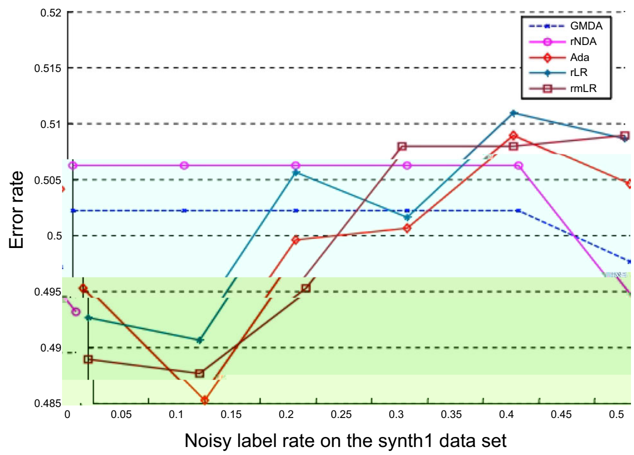
Table 7 r r rr r r ff r rr ff r r - r

	r					
	0.0	0.1	0.2	0.	0.	0.
(a1) error rate with 5-cross-validation on correlated synth1 dataset						
	0. 02	0. 02	0. 02	**0.5023**	**0.5023**	0.
r	0. 0	0. 0	0. 0	0. 0	0. 0	**0.4937**
	0.	**0.4853**	0.	0. 00	0. 0 0	0. 0
r	0. 2	0. 0	0. 0	0. 01	0. 110	0. 0
r	**0.4890**	0.	**0.4953**	0. 0 0	0. 0 0	0. 0 0
(b1) error rate with 10-cross-validation on correlated synth3 dataset						
	0. 0	0. 0	0. 2	0. 2	0. 2	**0.4927**
r	0. 0	0. 0	0.	0.	0.	0.
	0.4520	**0.4520**	**0.4520**	**0.4520**	0. 0	0. 0
r	0. 0	0. 0	0.	0. 0	0.	0.
r	0. 120	0.	0.	0. 0	**0.4827**	0.
(c1) error rate with 3-cross-validation on correlated synth5 dataset						
	0. 0 0	0. 0 0	0. 0 0	0. 0 0	0. 0 0	**0.4938**
r	0.	0. 0	0. 0	0. 0	0. 0	0. 0
	0.4812	**0.4812**	**0.4812**	**0.4812**	0. 1	0. 1
r	0.	0.	0.	0. 1	0.	0. 1
r	0.	0. 0	0. 0	0. 0	**0.4904**	0. 1 2
(a2) error rate with 5-cross-validation on uncorrelated synth2 dataset						
	0. 0	0. 0	0. 0	0.	0. 0	0. 0
r	0. 2	0. 2	0. 2	**0.4427**	**0.4427**	0. 2
	0. 2	0. 2	0. 2	0.	0.	0. 2
r	**0.1653**	**0.1103**	**0.3853**	0.	0.	**0.1653**
r	0.2 0	0.2 0	0. 0	0.	0. 20	0.2 0
(b2) error rate with 10-cross-validation on uncorrelated synth4 dataset						
	0.0420	0. 0	0.	0. 0	0.	0. 1
r	0. 0	0. 0	0. 0	0. 0	0. 0	**0.4560**
	0. 0	0. 0	0. 0	0. 0	0. 0	0. 0
r	0.0	**0.2313**	**0.2320**	**0.2133**	**0.4267**	0.
r	0.0 00	0.2 1	0.	0.	0.	0. 00
(c2) error rate with 3-cross-validation on uncorrelated synth6 dataset						
	0.0424	0. 2	0. 2	0. 0	0.	0. 2
r	0. 22	0. 22	0. 22	0. 22	0. 22	**0.4422**
	0. 22	0. 22	0. 22	0. 22	0.	0.
r	0.0	**0.1754**	**0.1562**	**0.1902**	**0.4370**	0. 2
r	0.0 0	0.2	0.2	0. 2	0.	0. 112

rf r r

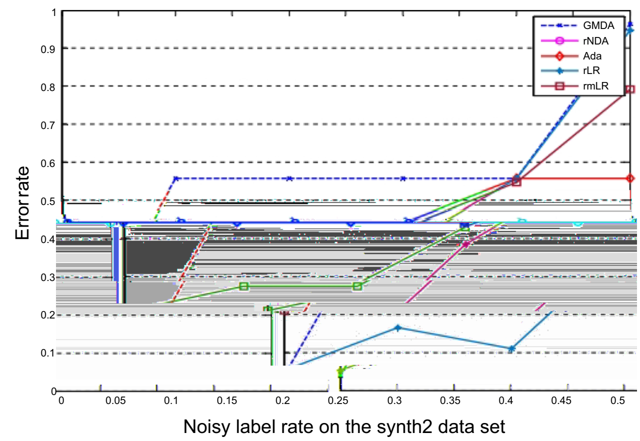
r , r r , rr r r . [0.2, 0. , r r - rf r- f . f r ff r fr r r f r r . ff r f , r . ff r f r ,

r r r r r . ' . ' ' , r r ff r r ' , - f r r r ' . r rr r r , ff f r . r , r rr r r . ' .



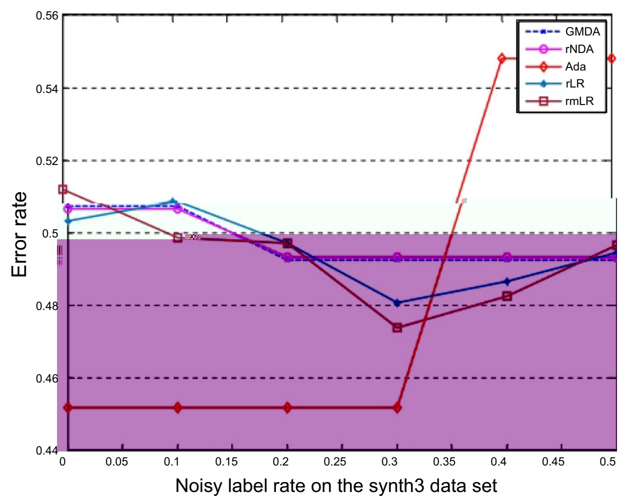
GMDA
rNDA
Ada
rLR
rmLR
Error rate
Noisy label rate on the synth1 data set

(a1)



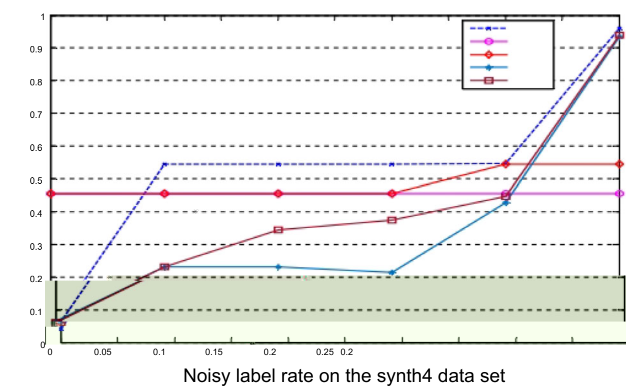
GMDA
rNDA
Ada
rLR
rmLR
Error rate
Noisy label rate on the synth2 data set

(a2)

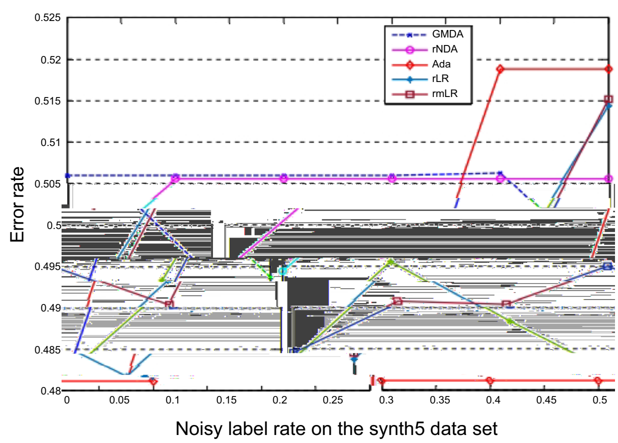


GMDA
rNDA
Ada
rLR
rmLR
Error rate
Noisy label rate on the synth3 data set

(b1)

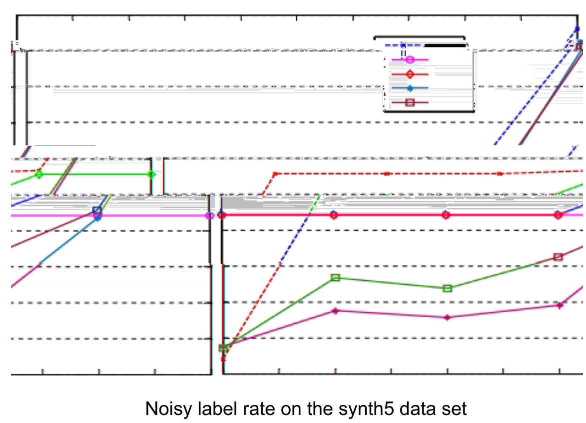


Noisy label rate on the synth4 data set



GMDA
rNDA
Ada
rLR
rmLR
Error rate
Noisy label rate on the synth5 data set

(c1)



Noisy label rate on the synth5 data set

6 Conclusion

Acknowledgements

Compliance with ethical standards

Conflict of interest

References

1. (2010)
2. (201)
(1) 11(1) 1 1 1
(200)
(2010)
K 2010, 2 2 (201)
1 11.01 1,201
(2001)
(2012)
K 2012, 1 -1 , 2012
(1) 100
10. (200) 0(12)
11. (2010) ( ) 2 0
12. K , (201) 1 0 .0
(201) 2 2
(201) 2 11
(201) 1 0 .120
(201) ( ) 11 11 1
(1) 1 (1) 0
K (200) 22( ) 1 2
K (2012)
20. K (2001) 2( ) 2 20
21. (2000) ( ) 2 2
22. (1) 0
(201)
K (1) 1 ( ) 0
(1) (1) 1
(1 2) K) 1 ( ) 1 1 0

Publisher's Note