Efficient Training of RBF Networks Via the BYY Automated Model Selection Learning Algorithms

*

V
jwma@math.pku.edu.cn

Abstract.

Y Y YY

YY

1 Introduction

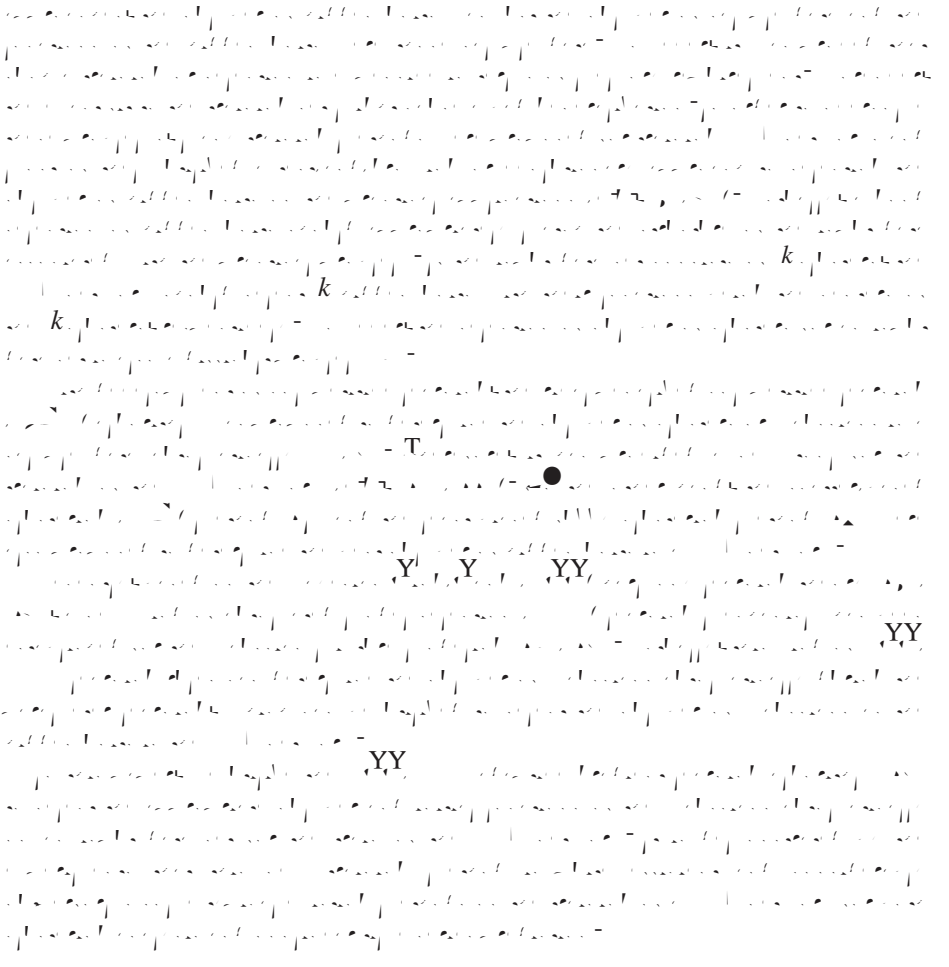
T

T

T

* T

V



2 BYY-AMS Adaptive Gradient Learning Algorithm

$p(y,x)\log p(x)+q(x,y)\log q(y)$

$$H(p,q)=\int p(y,x)\,p(x)\,q(x,y)\,q(y)\,dxdy-z_q$$

z_q

$p(y,x)$ $q(x,y)$ θ YY

YY $-q(y=j)=\alpha_j$

$\alpha_j\geq$ $\sum_{j=}^{K}\alpha_j=$ z_q z_q

$p(x)$ $p(x)=\frac{1}{N}\sum_{t=}^{N}\delta(x-x_t)$

$$p(y=j,x)=p(j,x)=\frac{\alpha_j q(x,\theta_j)}{q(x,\Theta_K)}$$

$$U_j(x) = \alpha_j q(x; m_j, \Sigma_j) \quad j = 1, \ldots, K, \quad J(\Theta_K)$$

$$J(\Theta_K) = \frac{1}{N}\sum_{t=1}^{N} J_t(\Theta_K), \quad J_t(\Theta_K) = \sum_{j=1}^{K} \frac{U_j(x_t)}{\sum_{i=1}^{K} U_i(x_t)} \log U_j(x_t)$$

$J(\Theta_K)$ β_j, m_j, B_j

$$\frac{\partial J_t(\Theta_K)}{\partial \beta_j} = \frac{1}{q(x_t; \Theta_k)} \sum_{i=1}^{K} \lambda_i(t)(\delta_{ij} - \alpha_j U_i(x_t))$$

$$\frac{\partial J_t(\Theta_K)}{\partial m_j} = p(j \mid x_t) \lambda_j(t) \Sigma_j^{-1}(x_t - m_j)$$

$$vec\left(\frac{\partial J_t(\Theta_K)}{\partial B_j}\right) = \frac{\partial (B_j B_j^T)}{\partial B_j} vec\left(\frac{\partial J_t(\Theta_k)}{\partial \Sigma_j}\right)$$

δ_{ij} $vec\ A$

$$\lambda_i(t) = -\sum_{l=1}^{K} p(l \mid x_t)(\delta_{il} - \alpha_l q(x_t; m_l, \Sigma_l))$$

$$\frac{\partial J_t(\Theta_K)}{\partial \Sigma_j} = -\frac{1}{2} p(j \mid x_t) \lambda_j(t) \left(\Sigma_j^{-1}(x_t - m_j)(x_t - m_j)^T \Sigma_j^{-1} - \Sigma_j^{-1}\right)$$

$$\frac{\partial (BB^T)}{\partial B} = I_{d\times d} \otimes B_{d\times d}^T + E_{d\times d} \bullet B_{d\times d}^T \otimes I_{d\times d}$$

$\otimes$

$$E_{d\times d} = \frac{\partial B^T}{\partial B} = (\Gamma_{ij})_{d\times d} = \begin{pmatrix} \Gamma_{11} & \cdots & \Gamma_{1d} \\ \vdots & \ddots & \vdots \\ \Gamma_{d1} & \cdots & \Gamma_{dd} \end{pmatrix}_{d\times d}$$

Γ_{ij} j,i^{th}

$\frac{\partial (BB^T)}{\partial B}$

$$vec\left(\frac{\partial J(\Theta_K)}{\partial B_j}\right) = -\sum_{t=1}^{T} p(j \mid x_t, \lambda_j, t)\left(I_{d\times d} \otimes B_{d\times d}^T + E_{d\times d} \bullet B_{d\times d}^T \otimes I_{d\times d}\right)$$
$$\times vec\left(\Sigma_j^{-1}(x_t - m_j)(x_t - m_j)^T \Sigma_j^{-1} - \Sigma_j^{-1}\right) \quad (\quad)$$

$$\Delta\beta_j = \frac{\eta}{q(x_t \mid \Theta_k)} \sum_{i=1}^{K} \lambda(i, t)(\delta_{ij} - \alpha_j) U(i, x_t) \quad (\quad)$$

$$\Delta m_j = \eta\, p(j \mid x_t, \lambda_j, t)\Sigma_j^{-1}(x_t - m_j) \quad (\quad)$$

$$\Delta vecB_j = \frac{\eta}{\quad} p(j \mid x_t, \lambda_j, t)\left(I_{d\times d} \otimes B_{d\times d}^T + E_{d\times d} \bullet B_{d\times d}^T \otimes I_{d\times d}\right)$$
$$\times vec\left(\Sigma_j^{-1}(x_t - m_j)(x_t - m_j)^T \Sigma_j^{-1} - \Sigma_j^{-1}\right) \quad (\quad)$$

η n $\leq \eta(n) \leq$

$$\lim_{n\to\infty} \eta(n) = \quad \sum_{n=1}^{\infty} \eta(n) = \infty \quad (\quad)$$

T $\eta(n) = \eta \quad n$

η

T YY K

YY

D_x -

3 Training of the RBF Network

YY

$$y_j(x) = \sum_{j=1}^{n} w_{ij}\phi_i(x) \quad (\quad)$$

n w_{ij} i^{th} j^{th} $\phi_i(x)$ i^{th}

$$\phi_j(x) = \phi_j(\|x - m_j\|) = \exp\left(-\frac{\|x - m_j\|}{\sigma_j}\right)$$

m_j, σ_j $\phi_j(x)$

$$y(x) = \sum_{j=1}^{n} \lambda_j \phi_j(x) = \sum_{j=1}^{n} \lambda_j \exp\left(-\frac{\|x - m_j\|}{\sigma_j}\right)$$

λ_j j^{th}

$\lambda_j, m_j, \sigma_j >$

$j = 1, \cdots, n$

$D = \{(x_i, y_i) \mid i = 1, \cdots, N\}$

$$E = \frac{1}{2}\sum_{i=1}^{N} (y_i - f(x_i))^2 = \frac{1}{2}\sum_{i=1}^{N} \left(y_i - \sum_{j=1}^{n} \lambda_j \phi_j(x_i)\right)^2$$

$$= \frac{1}{2}\sum_{i=1}^{N} \left(y_i - \sum_{j=1}^{n} \lambda_j \exp\left(-\frac{\|x_i - m_j\|^2}{\sigma_j}\right)\right)^2$$

$$\begin{cases} \Delta\lambda_j = \eta_\lambda \sum_{i=1}^{N} (y_i - \sum_{l=1}^{n} \lambda_l \phi_l(x_i)) \phi_j(x_i) \\ \Delta m_j = \eta_m \sum_{i=1}^{N} (y_i - \sum_{l=1}^{n} \lambda_l \phi_l(x_i)) \phi_j(x_i) (x_i - m_j) \lambda_j / \sigma_j^2 \\ \Delta\sigma_j = \eta_\sigma \sum_{i=1}^{N} (y_i - \sum_{l=1}^{n} \lambda_l \phi_l(x_i)) \phi_j(x_i) (x_i - m_j)^T (x_i - m_j) \lambda_j / \sigma_j^3 \end{cases}$$

η_λ, η_m, η_σ

λ_j, m_j, σ_j j

$$\sigma_j = \frac{1}{N_j} \sum_{x_t \in C_j} (x_t - m_j)(x_t - m_j)^T$$

4 Experiment Results

4.1 On the Noisy XOR Problem

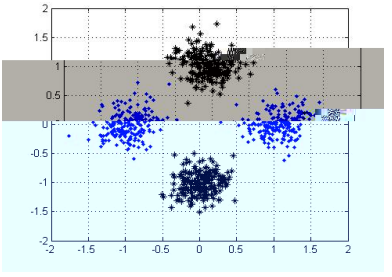


Fig. 1. T X

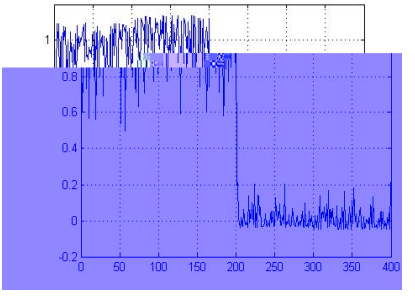


Fig. 2. T

4.2 On the Mackey-Glass Time Series Prediction

YY

$$x(t+1) = (1-b)x(t) + \frac{ax(t-\tau)}{1+x(t-\tau)},$$

$a = \quad b = \quad \tau =$ $x(t-\quad x(t-\quad x(t-\quad x(t) \quad x(t+\quad \le t \le$

$y_i = f(x_i)$ $x_i = [x(t-\quad x(t-\quad x(t-\quad x(t)]^T$ $y_i = x(t+$

$i = \cdots N$

T

YY

YY

T

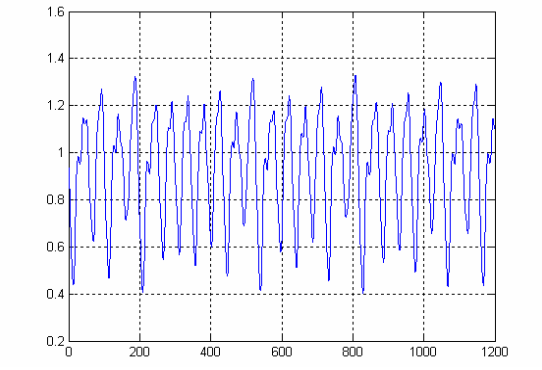


Fig. 3. T

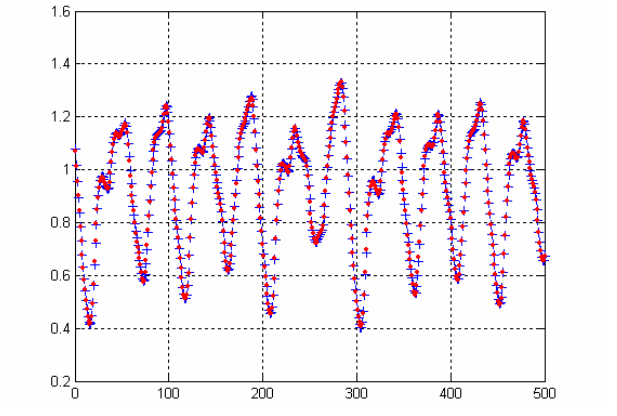


Fig. 4. T YY

5 Conclusions

YY

References