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

[illegible]

[illegible]
jwma@math.pku.edu.cn

Abstract. [illegible]

1 Introduction

[illegible]

*[illegible]

[illegible]

[illegible]

[illegible]

2 BYY-AMS Adaptive Gradient Learning Algorithm

[illegible] YY [illegible] YY [illegible] $x \in X \subset R^d$ [illegible] $y \in Y \subset R^m$ [illegible] $p(x, y) = p(x)p(y|x)$ [illegible] $q(x, y) = q(x|y)q(y)$ [illegible] Y [illegible] Y [illegible] y [illegible]

$y \in Y = \{1, 2, \cdots, K\} \subset R$ [illegible] $D_x = \{x_t\}_{t=1}^{N}$ [illegible] YY [illegible]

$p(y,x)$ $p(x)$ $q(x,y)$ $q(y)$ [illegible]

$$H(p,q) = \int p(y,x)p(x) \; q(x,y)q(y) \, dxdy - \; z_q$$ [illegible]

[illegible] z_q [illegible]

[illegible] $p(y,x)$ [illegible] $q(x,y)$ [illegible] θ [illegible] YY [illegible] YY [illegible] $q(y=j)=\alpha_j$ [illegible] $\alpha_j \ge$ [illegible] $\sum_{j=}^{K} \alpha_j =$ [illegible] z_q [illegible] z_q [illegible] $p(x)$ [illegible] $p(x) = \frac{}{N}\sum_{t=}^{N} \delta(x-x_t)$ [illegible]

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

[illegible]

[illegible] $U_j(x) = \alpha_j q(x, m_j, \Sigma_j)$ [illegible] $j = , \cdots, K$, $J(\Theta_K)$ [illegible]

[illegible]

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

[illegible] $J(\Theta_K)$ [illegible] β_j, m_j [illegible] B_j,

[illegible]

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

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

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

[illegible] δ_{ij} [illegible] $vec[A]$ [illegible]

[illegible]

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

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

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

[illegible] $\otimes$ [illegible]

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

[illegible] Γ_{ij} [illegible] $(j, i)^{th}$ [illegible]

[illegible] $\frac{\partial (BB^T)}{\partial B}$, [illegible]

$$vec\frac{\partial J(\Theta_K)}{\partial B_j} = -p(j|x_t)\lambda_j(t)(I_{d\times d}\otimes B^T_{d\times d} + E_{d \times d}\bullet B^T_{d\times d}\otimes I_{d\times d})$$

$$\times vec[\Sigma_j^{-}(x_t-m_j)(x_t-m_j)^T\Sigma_j^{-} - \Sigma_j^{-}]$$

[illegible]

$$\Delta\beta_j = \frac{\eta}{q(x_t,\Theta_k)}\sum_{i=}^{K}\lambda_i(t)(\delta_{ij}-\alpha_j)U_i(x_t)$$

$$\Delta m_j = \eta p(j|x_t)\lambda_j(t)\Sigma_j^{-}(x_t-m_j)$$

$$\Delta vecB_j = \frac{\eta}{-}p(j|x_t)\lambda_j(t)(I_{d\times d}\otimes B^T_{d\times d} + E_{d \times d}\bullet B^T_{d\times d}\otimes I_{d\times d})$$

$$\times vec[\Sigma_j^{-}(x_t-m_j)(x_t-m_j)^T\Sigma_j^{-} - \Sigma_j^{-}]$$

[illegible] η [illegible] n [illegible] $\leq \eta(n) \leq$ [illegible]

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

[illegible] $\eta(n) = \eta \quad n$ [illegible] η [illegible] YY [illegible] K [illegible] YY [illegible] D_x [illegible]

3 Training of the RBF Network

[illegible] YY [illegible]

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

[illegible]

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

$$\phi_j(x) = \phi_j(\|x-m_j\|) = [illegible] -\frac{\|x-m_j\|}{\sigma_j}$$

[illegible] m_j, σ_j [illegible] $\phi_j(x)$ [illegible]

$$y(x) = \sum_{j=}^{n} \lambda_j \phi_j(x) = \sum_{j=}^{n} \lambda_j [illegible] -\frac{\|x-m_j\|}{\sigma_j}$$

[illegible] λ_j [illegible] j^{th} [illegible] $\lambda_j, m_j, \sigma_j >$ [illegible] $j = , , \cdots, n$ [illegible] $D(x, y) = \{(x_i, y_i) \mid i = , , \cdots, N\}$ [illegible]

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

$$= -\sum_{i=}^{N} \left(y_i - \sum_{j=}^{n} \lambda_j [illegible] -\frac{\|x_i - m_j\|}{\sigma_j}\right)^{}$$

[illegible]

$$\begin{cases} \Delta\lambda_j = \eta_\lambda \sum_{i=}^{N} \left(y_i - \sum_{l=}^{n} \lambda_l \phi_l(x_i)\right) \phi_j(x_i) \\ \Delta m_j = \eta_m \sum_{i=}^{N} \left(y_i - \sum_{l=}^{n} \lambda_l \phi_l(x_i)\right) \phi_j(x_i)(x_i - m_j)\lambda_j \quad \sigma_j \\ \Delta\sigma_j = \eta_\sigma \sum_{i=}^{N} \left(y_i - \sum_{l=}^{n} \lambda_l \phi_l(x_l)\right) \phi_j(x_i)(x_i - m_j)^T (x_i - m_j)\lambda_j \quad \sigma_j \end{cases}$$

[illegible] η_λ, η_m, η_σ [illegible] λ_j, m_j, σ_j [illegible] j [illegible]

[illegible]

[illegible] n [illegible] $n = K$ [illegible] K [illegible] n [illegible] σ_j [illegible]

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

[illegible] C_j [illegible] x_t [illegible] $p(j|x_t)$ [illegible] N_j [illegible] C_j [illegible] m_j [illegible]

4 Experiment Results

[illegible]

4.1 On the Noisy XOR Problem

[illegible]

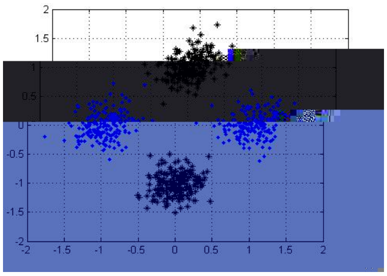

Fig. 1. [illegible] X [illegible]

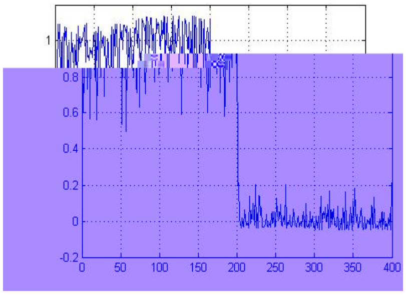

Fig. 2. [illegible]

[illegible] S [illegible]

4.2 On the Mackey-Glass Time Series Prediction

[illegible] YY S [illegible]

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

[illegible] $a =$ [illegible] $b =$ [illegible] $\tau =$ [illegible] $x(t-\,), x(t-\,), x(t-\,), x(t), x(t+\,)$ [illegible] $\le t \le$ [illegible]

[illegible] $y_i = f(x_i)$ [illegible] $x_i = (x(t-\,), x(t-\,), x(t-\,), x(t))^T$ [illegible] $y_i = x(t+\,)$ [illegible] $i = \,\cdots, N$ [illegible]

[illegible] SE YY S [illegible]

[illegible] YY S [illegible]

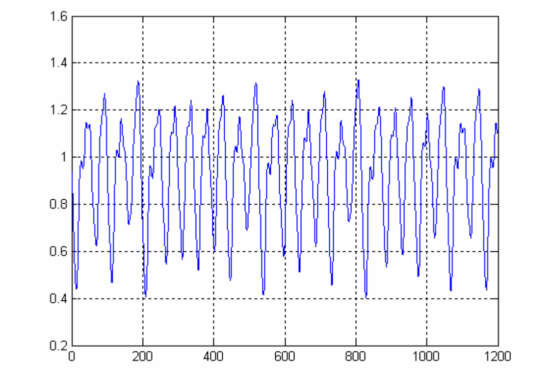

Fig. 3. [illegible]

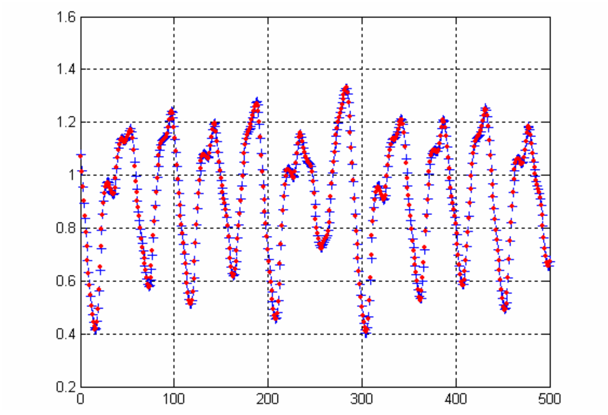

Fig. 4. [illegible]

5 Conclusions

[illegible]

References

[illegible]