ADAPTIVE LEAST ERROR RATE ALGORITHM FOR NEURAL NETWORK CLASSIFIERS

S. Chen[†], B. Mulgrew[‡] and L. Hanzo[†]
[†] Department of Electronics and Computer Science
University of Southampton, Highfield
Southampton SO17 1BJ, U.K.
[‡] Department of Electronics and Electrical Engineering
University of Edinburgh, King's Buildings
Edinburgh EH9 3JL, U.K.

Abstract. We consider sample-by-sample adaptive training of two-class neural network classifiers. Specific applications that we have in mind are communication channel equalization and code-division multiple-access (CDMA) multiuser detection. Typically, training of such neural network classifiers is done using some stochastic gradient algorithm that tries to minimize the mean square error (MSE). Since the goal should really be minimizing the error probability, the MSE is a "wrong" criterion to use and may lead to a poor performance. We propose a stochastic gradient adaptive minimum error rate (MER) algorithm called the least error rate (LER) for training neural network classifiers.

INTRODUCTION

We study the class of neural network classifiers where pattern vectors are drawn from a finite set and corrupted by an additive noise. Specific examples include neural network equalizers and multiuser detectors [1]-[6]. We assume that a sample-by-sample adaptation is necessary to meet real-time computational constraints. In such applications, the training of neural network classifiers is usually done using some stochastic gradient algorithm based on the MSE criterion, and a classical example is the least mean square (LMS) algorithm. However, the real goal is to minimize the error probability. The bias in favour of the MSE criterion is perhaps rooted in adaptive linear filtering.

For linear classifiers, such as linear equalizers and multiuser detectors, there exists some relationship between the MSE and error probability. A small MSE is usually associated with a small error rate. However, even in the linear case, the minimum MSE (MMSE) solution in general is not the MER one. It is now well-known that the error rate gap between the MMSE solution and MER one can be large in certain situations. This has motivated

the recent research in linear adaptive MER algorithms [7]-[13]. For nonlinear classifiers, there is no relationship between the MSE and error rate, and the MMSE solution may not correspond to a small error rate. In effect, standard adaptive algorithms for training nonlinear classifiers, such as the LMS algorithm, is based on a criterion that is irrelevant to the problem.

We adopt the MER criterion and develop a stochastic gradient adaptive MER algorithm for training neural network classifiers. We employ an adaptive strategy that is very similar to the one used for deriving the LMS. The MMSE solution requires ensemble averages, which are not available in general. Sample time averages can in practice be used to provide estimates of the required statistics. Taking to extreme, using one-sample estimate leads to the LMS adaptation. The error probability of a classifier is an integration of the probability density function (p.d.f.) of the classifier decision variable, which is generally unavailable. However a sample time average of the p.d.f., called the kernel density estimate [14],[15], can be constructed. When a one-sample kernel density estimate is used, an instantaneous or stochastic gradient adaptive algorithm is derived. This approach has successfully been applied in our previous works on linear adaptive MER algorithms [9],[10].

The proposed LER algorithm is applied to equalization and multiuser detection using a radial basis function (RBF) network. The simulation results obtained show that the LER algorithm has a reasonable convergence speed and a small RBF network trained by the LER algorithm can closely match the optimal performance. The simulation study also confirms that the MSE criterion is irrelevant to this kind of applications and the RBF network trained by the LMS algorithm, although converging well in the MSE, can produce a poor error rate performance.

PROBLEM FORMULATION

We consider the class of nonlinear classifiers that are defined by

$$\hat{c}(k) = \text{sgn}(y(k)) \quad \text{with} \quad y(k) = f(\mathbf{r}(k); \mathbf{w}) \tag{1}$$

where $\mathbf{r}(k)$ is an M-dimensional pattern vector with its associated class label $c(k) \in \{\pm 1\}$, the vector $\mathbf{w}$ consists of all the adjustable parameters of the classifier f, and $\hat{c}(k)$ is the estimated class label for $\mathbf{r}(k)$. Any soft sgn function employed by a classifier is included in the function form f. The pattern vector $\mathbf{r}(k)$ is assumed to take the form: $\mathbf{r}(k) = \bar{\mathbf{r}}(k) + \mathbf{n}(k)$, where the clean part $\bar{\mathbf{r}}(k)$ takes values from a finite set with equal probability

$$\bar{\mathbf{r}}(k) \in \{\mathbf{r}_j, \quad 1 \leq j \leq N_b\} \tag{2}$$

and the noise vector $\mathbf{n}(k)$ is Gaussian with covariance matrix $E[\mathbf{n}(k)\mathbf{n}^T(k)] = \sigma_n^2 \mathbf{I}$. Each $\mathbf{r}_j$ has an associated class label $c^{(j)} \in \{\pm 1\}$.

Classically, adaptive training of such a nonlinear classifier is done by adjusting $\mathbf{w}$ so that the MSE, $E[(c(k) - y(k))^2]$, is minimized, and is typically implemented using the LMS algorithm which has a simple form:

$$y(k) = f(\mathbf{r}(k); \mathbf{w}(k-1))$$
$$\left. \mathbf{w}(k) = \mathbf{w}(k-1) + \mu(c(k) - y(k))\frac{\partial f(\mathbf{r}(k); \mathbf{w}(k-1))}{\partial \mathbf{w}} \right\} \qquad (3)$$

where μ is an adaptive gain. However, the true performance criterion should be the error rate. We consider how to construct an adaptive training algorithm based on the MER criterion. Note that the MER is with respect to the chosen classifier structure.

ADAPTIVE MINIMUM ERROR RATE TRAINING

The error probability of the classifier (1) is

$$P_E(\mathbf{w}) = \mathrm{Prob}\{\mathrm{sgn}(c(k))y(k) < 0\} \qquad (4)$$

Define the signed variable $y_s(k) = \mathrm{sgn}(c(k))y(k)$ and its p.d.f. $p_y(y_s)$. Then

$$P_E(\mathbf{w}) = \int_{-\infty}^{0} p_y(y_s)\,dy_s \qquad (5)$$

By linearizing the classifier around $\bar{\mathbf{r}}(k)$, it can be approximated as

$$y(k) = f(\bar{\mathbf{r}}(k)+\mathbf{n}(k); \mathbf{w}) \approx f(\bar{\mathbf{r}}(k); \mathbf{w}) + \left[\frac{\partial f(\bar{\mathbf{r}}(k); \mathbf{w})}{\partial \mathbf{r}}\right]^{T} \mathbf{n}(k) = f(\bar{\mathbf{r}}(k); \mathbf{w}) + e(k) \qquad (6)$$

where $e(k)$ is Gaussian with zero mean and variance

$$\rho^2 = \mathrm{E}\left[\sigma_n^2 \left[\frac{\partial f(\bar{\mathbf{r}}(k); \mathbf{w})}{\partial \mathbf{r}}\right]^{T} \frac{\partial f(\bar{\mathbf{r}}(k); \mathbf{w})}{\partial \mathbf{r}}\right] = \frac{\sigma_n^2}{N_b} \sum_{j=1}^{N_b} \left[\frac{\partial f(\mathbf{r}_j; \mathbf{w})}{\partial \mathbf{r}}\right]^{T} \frac{\partial f(\mathbf{r}_j; \mathbf{w})}{\partial \mathbf{r}} \qquad (7)$$

Thus, the classifier is approximated as an additive Gaussian noise model $y(k) \approx \bar{y}(k) + e(k)$, with $\bar{y}(k)$ taking values from the finite set

$$\bar{y}(k) \in \{y_j = f(\mathbf{r}_j; \mathbf{w}), \quad 1 \le j \le N_b\} \qquad (8)$$

The p.d.f. of $y_s(k)$ can then be approximated by

$$p_y(y_s) \approx \frac{1}{N_b\sqrt{2\pi}\rho} \sum_{j=1}^{N_b} \exp\left(-\frac{(y_s - \mathrm{sgn}(c^{(j)})y_j)^2}{2\rho^2}\right) \qquad (9)$$

and the error probability of the classifier is approximately

$$P_E(\mathbf{w}) \approx \frac{1}{N_b\sqrt{2\pi}} \sum_{j=1}^{N_b} \int_{g_j(\mathbf{w})}^{\infty} \exp\left(-\frac{x_j^2}{2}\right) dx_j = \frac{1}{N_b} \sum_{j=1}^{N_b} Q(g_j(\mathbf{w})) \qquad (10)$$

where

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} \exp\left(-\frac{y^2}{2}\right) dy \qquad (11)$$

and

$$g_j(\mathbf{w}) = \frac{\operatorname{sgn}(c^{(j)})y_j}{\rho} = \frac{\operatorname{sgn}(c^{(j)})f(\mathbf{r}_j; \mathbf{w})}{\rho} \qquad (12)$$

Approximate minimum error rate solution

If the set (2) is known, an approximate MER solution can be obtained by minimizing the error rate expression (10). Noting the gradient of $P_E(\mathbf{w})$

$$\nabla P_E(\mathbf{w}) \approx -\frac{1}{N_b\sqrt{2\pi}} \sum_{j=1}^{N_b} \exp\left(-\frac{y_j^2}{2\rho^2}\right) \frac{\partial g_j(\mathbf{w})}{\partial \mathbf{w}}$$

$$= -\frac{1}{N_b\sqrt{2\pi}\rho} \sum_{j=1}^{N_b} \exp\left(-\frac{y_j^2}{2\rho^2}\right) \operatorname{sgn}(c^{(j)}) \frac{\partial f(\mathbf{r}_j; \mathbf{w})}{\partial \mathbf{w}} \qquad (13)$$

the following iterative gradient algorithm can be used to arrive at an approximate MER solution. Given an initial $\mathbf{w}(0)$, at lth iteration:

$$\left.\begin{array}{l} y_j(l) = f(\mathbf{r}_j; \mathbf{w}(l-1)), \quad 1 \le j \le N_b \\ \nabla P_E(\mathbf{w}(l)) = -\frac{1}{N_b\sqrt{2\pi}\rho} \sum_{j=1}^{N_b} \exp\left(-\frac{y_j^2(l)}{2\rho^2}\right) \operatorname{sgn}(c^{(j)}) \frac{\partial f(\mathbf{r}_j; \mathbf{w}(l-1))}{\partial \mathbf{w}} \\ \mathbf{w}(l) = \mathbf{w}(l-1) - \mu\nabla P_E(\mathbf{w}(l)) \end{array}\right\} \quad (14)$$

Notice that the variance ρ^2 could iteratively be calculated using (7). However, for numerical and convergence considerations, it is preferred to fix ρ^2 to an appropriately chosen constant, that is, to consider ρ^2 as an algorithm parameter that requires tuning. If the classifier (1) is linear, all the approximations becomes exact, and we arrive at the exact MER solution [9],[10].

Block-data gradient adaptive algorithm

We will adopting a sample time average for estimating the p.d.f.. This is known as kernel density estimation [14],[15]. Given a block of K training samples $\{\mathbf{r}(k), c(k)\}$, a kernel density estimate of $p_y(y_s)$ is

$$\hat{p}_y(y_s) = \frac{1}{K\sqrt{2\pi}\rho} \sum_{k=1}^{K} \exp\left(-\frac{(y_s - \operatorname{sgn}(c(k))y(k))^2}{2\rho^2}\right) \qquad (15)$$

From the estimated error probability

$$\hat{P}_E(\mathbf{w}) = \int_{-\infty}^{0} \hat{p}_y(y_s)\, dy_s \qquad (16)$$

$\nabla\hat{P}_E(\mathbf{w})$ can be calculated

$$\nabla\hat{P}_E(\mathbf{w}) = -\frac{1}{K\sqrt{2\pi}\rho} \sum_{k=1}^{K} \exp\left(-\frac{y^2(k)}{2\rho^2}\right) \operatorname{sgn}(c(k)) \frac{\partial f(\mathbf{r}(k); \mathbf{w})}{\partial \mathbf{w}} \qquad (17)$$

Thus a block adaptive gradient algorithm similar to (14) can be derived.

Stochastic gradient adaptive algorithm

To derive the LER algorithm, consider a single-sample estimate of $p_y(y_s)$:

$$\hat{p}_y(y_s, k) = \frac{1}{\sqrt{2\pi}\rho} \exp\left(-\frac{(y_s - \text{sgn}(c(k))y(k))^2}{2\rho^2}\right) \qquad (18)$$

Using the instantaneous or stochastic gradient

$$\nabla \hat{P}_E(k; \mathbf{w}) = -\frac{1}{\sqrt{2\pi}\rho} \exp\left(-\frac{y^2(k)}{2\rho^2}\right) \text{sgn}(c(k)) \frac{\partial f(\mathbf{r}(k); \mathbf{w})}{\partial \mathbf{w}} \qquad (19)$$

a stochastic gradient algorithm is given by

$$\left.\begin{array}{l} y(k) = f(\mathbf{r}(k); \mathbf{w}(k-1)) \\ \mathbf{w}(k) = \mathbf{w}(k-1) + \frac{\mu}{\sqrt{2\pi}\rho} \exp\left(-\frac{y^2(k)}{2\rho^2}\right) \text{sgn}(c(k)) \frac{\partial f(\mathbf{r}(k); \mathbf{w}(k-1))}{\partial \mathbf{w}} \end{array}\right\} \qquad (20)$$

where μ and ρ are the adaptive gain and width parameter.

APPLICATIONS

The first application considers equalization in the presence of channel inter-symbol interference (ISI), co-channel interference and additive white Gaussian noise. For simplification, it is assumed that there exists only one interfering co-channel, and the received signal at sample k is given by

$$r(k) = \bar{r}(k) + n(k) = \sum_{i=0}^{n_0-1} a_{0,i} b_0(k-i) + \sum_{i=0}^{n_1-1} a_{1,i} b_1(k-i) + n(k) \qquad (21)$$

where $n(k)$ is a Gaussian white noise with variance σ_n^2, $a_{0,i}$ are the channel taps and $a_{1,i}$ the co-channel taps, the desired and interfering data $b_0(k)$ and $b_1(k)$ take value from the set $\{\pm 1\}$, and they are uncorrelated. The equalizer uses the received signal vector, $\mathbf{r}(k) = [r(k) \ r(k-1) \cdots r(k-M+1)]^T$, to produce an estimate of $b_0(k-d)$. Thus, $b_0(k-d)$ serves as the class label, and the number of states is $N_b = 2^{2M+n_0+n_1-2}$. The optimal equaliser, known as the Bayesian equaliser, is given in [16]. A RBF network with n_c Gaussian kernels is used as the adaptive equalizer. The weight vector of the equalizer thus consists of all the RBF weights, centers and widths.

In the simulation, the channel and co-channel are respectively $A_0(z) = 0.5 + 1.0z^{-1}$ and $A_1(z) = \lambda(1.0 + 0.5z^{-1})$, with λ set to give a signal to interference ratio SIR = 12 dB. The equalizer order is set to $M = 2$ and the decision delay $d = 1$. The number of states is $N_b = 64$. The first $n_c/2$ data points that belong to the class $+1$ and the first $n_c/2$ data points that belong to the class -1 are used as initial centers. The initial weights are set to $\pm\frac{1}{n_c(2\pi\sigma_n^2)^{M/2}}$ accordingly. All the widths are initially set to $8\sigma_n^2$. Given signal to noise ratio SNR = 20 dB (signal to interference and noise

ratio SINR = 11.36 dB), RBF equalizers with 4 and 6 centers are trained by the LMS and LER algorithms, respectively. Fig. 1 (a) shows the learning rates in terms of the estimated bit error rate (BER) for the four adaptive RBF equalizers, where the results are averaged over 100 runs. For the LMS training, the learning rates in terms of the MSE are depicted in Fig. 1 (b).
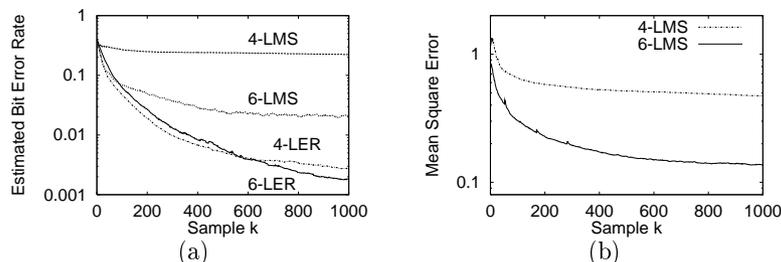

Figure 1: Convergence rates in terms of (a) the estimated BER for various adaptive RBF equalizers, and (b) the MSE for LMS adaptive RBF equalizers. The results are averaged over 100 runs.
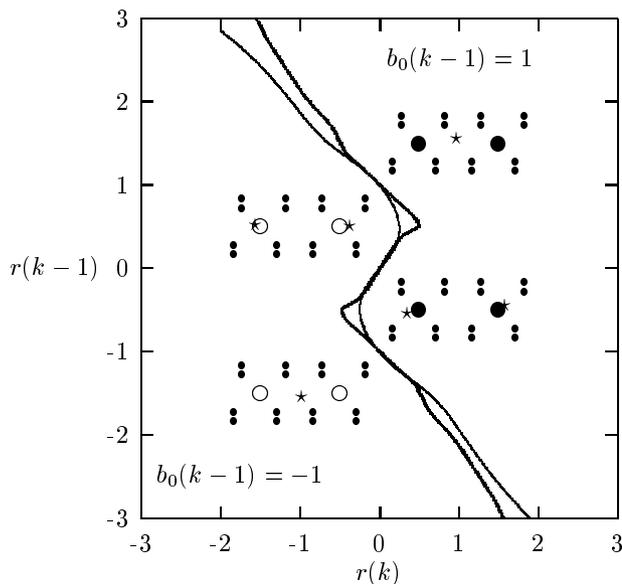

Figure 2: Comparison of the optimal decision boundary with that of the 6-center LER RBF equalizer (thin solid: RBF, thick solid: optimal). SNR = 20 dB and SIR = 12 dB. The dots indicate the noise-free states and stars the final centers.

It can be seen that, although the LMS RBF equalizers converge well in the MSE, they have poor BER performance. Typical decision boundary of the 6-center LER RBF equalizer is compared with the optimal boundary in Fig. 2. The true BERs of the 4-center LER RBF equalizer together with those of the linear MMSE and optimal Bayesian equalizers are depicted in Fig. 3

as functions of SNRs. The BERs of the 6-center LER RBF equalizer are not shown, as they are almost indistinguishable from the optimal performance. The BERs of the 6-center LMS RBF equalizer, not shown, are not better than those of the linear MMSE equaliser.
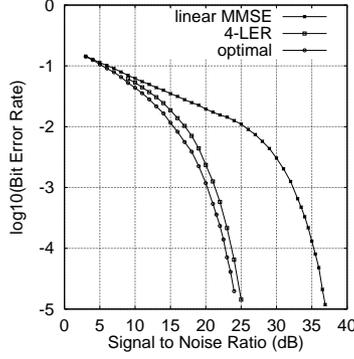

Figure 3: Performance comparison of three equalizers in terms of BER versus SNR. SIR = 12 dB. The adaptive LER RBF equalizer has 4 centers.

The second application considers the multiuser detection for the synchronous CDMA downlink with N users and M chips per symbol. The bit vector of N users is $\mathbf{b}(k) = [b_1(k) \cdots b_N(k)]^T$, with $b_i(k) \in \{\pm 1\}$ denoting the kth symbol of user i, the unit-length signature code sequence for user i is $\bar{\mathbf{s}}_i = [\bar{s}_{i,1} \cdots \bar{s}_{i,M}]^T$, and the channel (at chip rate) is $H(z) = \sum_{i=0}^{n_h-1} h_i z^{-i}$. It can be shown that the received signal vector sampled at chip rate, $\mathbf{r}(k) = [r_1(k) \cdots r_M(k)]^T$, is:

$$\mathbf{r}(k) = \mathbf{P} \begin{bmatrix} \mathbf{b}(k) \\ \mathbf{b}(k-1) \\ \vdots \\ \mathbf{b}(k-L+1) \end{bmatrix} + \mathbf{n}(k) = \bar{\mathbf{r}}(k) + \mathbf{n}(k) \tag{22}$$

where $\bar{\mathbf{r}}(k)$ denotes the noise-free received signal, the Gaussian noise vector $\mathbf{n}(k) = [n_1(k) \cdots n_M(k)]^T$ with $E[\mathbf{n}(k)\mathbf{n}^T(k)] = \sigma_n^2 \mathbf{I}$, the $M \times LN$ system matrix $\mathbf{P}$ is given by

$$\mathbf{P} = \mathbf{H} \begin{bmatrix} \bar{\mathbf{S}}\mathbf{A} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{S}}\mathbf{A} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \bar{\mathbf{S}}\mathbf{A} \end{bmatrix} \tag{23}$$

the $M \times LM$ channel matrix $\mathbf{H}$ has the form

$$\mathbf{H} = \begin{bmatrix} h_0 & h_1 & \cdots & h_{n_h-1} & & & \\ & h_0 & h_1 & \cdots & h_{n_h-1} & & \\ & & \ddots & \ddots & \cdots & \ddots & \\ & & & h_0 & h_1 & \cdots & h_{n_h-1} \end{bmatrix} \tag{24}$$

the user signature sequence matrix $\bar{\mathbf{S}} = [\bar{\mathbf{s}}_1 \cdots \bar{\mathbf{s}}_N]$, and the diagonal user signal amplitude matrix $\mathbf{A} = \text{diag}\{A_1 \cdots A_N\}$. The channel ISI span L depends on the channel length, n_h, related to the chip length, M. $L = 1$ for $n_h = 1$, $L = 2$ for $1 < n_h \le M$, $L = 3$ for $M < n_h \le 2M$, and so on. The detector at the receiver for user i detects the transmitted bit $b_i(k)$ based on the received signal $\mathbf{r}(k)$. Thus $b_i(k)$ serves as class label, and the number of states is $N_b = 2^{LN}$. The optimal detector, called the Bayesian detector, is given in [6]. Adaptive detector employed is a RBF network with n_c Gaussian kernel functions.
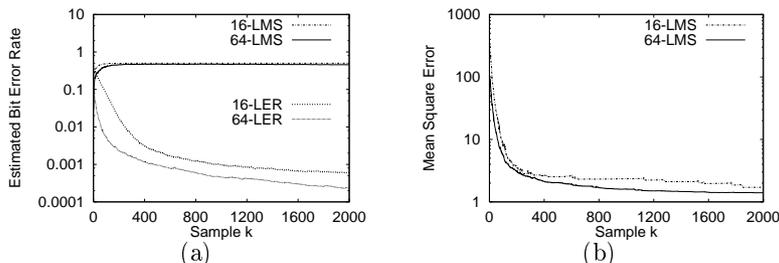

Figure 4: Convergence rates in terms of (a) the estimated BER for various user-3 adaptive RBF detectors and (b) the MSE for user-3 LMS RBF detectors. The results are averaged over 100 runs.

A 3-user system with 8 chips per symbol is used in the simulation. The chip codes for the three users are $(+1, +1, +1, +1, -1, -1, -1, -1)$, $(+1, -1, +1, -1, -1, +1, -1, +1)$ and $(+1, -1, -1, +1, -1, +1, +1, -1)$, respectively, and the channel is $H(z) = 0.8 + 0.6z^{-1} + 0.5z^{-2}$. The three users have equal signal power. For this example, the number of states is $N_b = 64$. The detector for user 3 is considered. Weights, centers and widths of RBF detectors are initialized in a similar manner as for the previous equalization application. Given SNR for user 3 $\text{SNR}_3 = 15$ dB (SINR for user 3 is $\text{SINR}_3 = -3.08$ dB), RBF detectors with 16 and 64 centers are trained by the LMS and LER algorithms, respectively. The learning rates in terms of the estimated BER are depicted in Fig. 4 (a) for the 4 adaptive RBF detectors. Fig. 4 (b) shows the learning rates in terms of the MSE for the 2 LMS RBF detectors. It can be seen that the LER algorithm produces consistent results and, in particular, the 64-center LER RBF detector is able to achieve the optimal performance. For the LMS training, the algorithm converges very well in the MSE and there is an almost 30 dB reduction in the MSE, as can be seen in Fig. 4 (b). However, the BERs of the two LMS RBF detectors both approach to 0.5. In fact, on average, the initial 16-center RBF detector has a BER= 0.2 and the initial 64-center RBF detector has a BER= 0.008. Yet, after training using the LMS, both yield almost 1 in 2 errors. This is nothing to do with local minima. As far as the LMS algorithm is concerned, it does a good job in what it supposes to do: getting the MSE down. The BERs of the 16-center LER RBF detector are compared with the optimal detector in Fig. 5 for a range of SNR, where it can be seen that its performance is very close to the
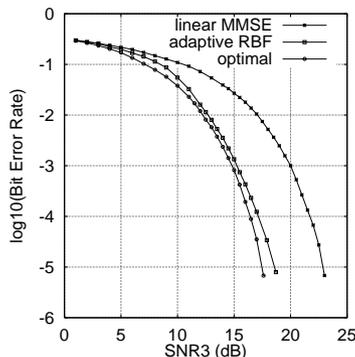
Bayesian detector of 64 states.

Figure 5: Performance comparison of three detectors for user 3. SNR_i, $1 \leq i \leq 3$, are identical. The adaptive RBF detector has 16 centers and is trained by the LER algorithm.

CONCLUSIONS

Adaptive or sample-by-sample training of nonlinear classifiers is classically done using some stochastic gradient algorithms, such as the LMS, that are linked to the MMSE criterion. This is despite the fact that the MMSE criterion bears no relationship to the MER criterion. It is widely acknowledged that adaptive training of neural network classifiers often encounters difficulties. These difficulties are typically attributed to "local minima". Rarely a more fundamental question is asked: whether the underlying criterion adopted, the MMSE, is a relevant one for the problem.

The main contribution of this paper is to derive an adaptive MER algorithm called the LER for training a class of neural network classifiers that include adaptive nonlinear equalisers and multiuser detectors. Our approach has been motivated from a kernel density estimation of the error rate as a smooth function of the training data and an adoption of stochastic gradient of the estimated error probability. This LER algorithm has been applied to channel equalization and CDMA multiuser detection using a RBF network. Simulation results have demonstrated that the LER algorithm performs consistently and the algorithm has a good convergence speed. A small-size RBF network trained by the LER algorithm can closely match the optimal Bayesian performance. The results also confirms that the standard LMS algorithm performs poorly in terms of error rate.

REFERENCES

[1] G.J. Gibson, S. Siu, S. Chen, C.F.N. Cowan and P.M. Grant, "The application of nonlinear architectures to adaptive channel equalisation," in *Proc. ICC'90* (Atlanta, USA), 1990, pp.312.8.1.–312.8.5.

[2] S. Chen, G.J. Gibson, C.F.N. Cowan and P.M. Grant, "Adaptive equalization of finite non-linear channels using multilayer perceptrons," *Signal Processing*, Vol.20, No.2, pp.107–119, 1990.

[3] S. Chen, G.J. Gibson, C.F.N. Cowan and P.M. Grant, "Reconstruction of binary signals using an adaptive radial-basis-function equalizer," *Signal Processing*, Vol.22, No.1, pp.77–93, 1991.

[4] B. Aazhang, B.P. Paris and G.C. Orsak, "Neural networks for multiuser detection in code-division multiple-access communications," *IEEE Trans. Communications*, Vol.40, No.7, pp.1212–1222, 1992.

[5] U. Mitra and H.V. Poor, "Neural network techniques for adaptive multiuser demodulation," *IEEE J. Selected Areas in Communications*, Vol.12, No.9, pp.1460-1470, 1994.

[6] S. Chen, A.K. Samingan and L. Hanzo, "Support vector machine multiuser receiver for DS-CDMA signals in multipath channels," *IEEE Trans. Neural Networks*, to appear, May 2001.

[7] S. Chen, E.S. Chng, B. Mulgrew and G. Gibson, "Minimum-BER linear-combiner DFE," in *Proc. ICC'96* (Dallas, Texas), 1996, Vol.2, pp.1173–1177.

[8] S. Chen, B. Mulgrew, E.S. Chng and G. Gibson, "Space translation properties and the minimum-BER linear-combiner DFE," *IEE Proc. Communications*, Vol.145, No.5, 1998, pp.316–322.

[9] B. Mulgrew and S. Chen, "Stochastic gradient minimum-BER decision feedback equalisers," in *Proc. IEEE Symposium on Adaptive Systems for Signal Processing, Communication and Control* (Lake Louise, Alberta, Canada), Oct.1-4, 2000, pp.93–98.

[10] S. Chen, A.K. Samingan, B. Mulgrew and L. Hanzo, "Adaptive minimum-BER linear multiuser detection for DS-CDMA signals in multipath channels," *IEEE Trans. Signal Processing*, Vol.49, No.6, pp.1240–1247, 2001.

[11] C.C. Yeh and J.R. Barry, "Adaptive minimum bit-error rate equalization for binary signaling," *IEEE Trans. Communications*, Vol.48, No.7, pp.1226–1235, 2000.

[12] C.C. Yeh, R.R. Lopes and J.R. Barry, "Approximate minimum bit-error rate multiuser detection," in *Proc. Globecom'98* (Sydney, Australia), Nov. 1998, pp.3590–3595.

[13] I.N. Psaromiligkos, S.N. Batalama and D.A. Pados, "On adaptive minimum probability of error linear filter receivers for DS-CDMA channels," *IEEE Trans. Communications*, Vol.47, No.7, pp.1092–1102, 1999.

[14] B.W. Silverman, *Density Estimation*. London: Chapman Hall, 1996.

[15] A.W. Bowman and A. Azzalini, *Applied Smoothing Techniques for Data Analysis*. Oxford University Press, 1997.

[16] S. Chen and B. Mulgrew, "Overcoming co-channel interference using an adaptive radial basis function equaliser," *Signal Processing*, Vol.28, No.1, pp.91–107, 1992.