MARGINAL MEDIAN LEARNING VECTOR QUANTIZER

Constantine KOTROPOULOS, Ioannis PITAS and Moncef GABBOUJ†
Department of Electrical & Computer Engineering, University of Thessaloniki, P.O. Box 463,
GR-54006 Thessaloniki, GREECE, Tel/Fax: +30 [31] 996 361 / 274 868, E-Mail: kotro@vergina.eng.auth.gr
Tel/Fax: +30 [31] 996 304 / 274 868, E-Mail: pitas@vergina.eng.auth.gr
† Signal Processing Laboratory, Tampere University of Technology P.O. Box 553
SF-33101 Tampere, FINLAND, Tel/Fax: +358 [31] 161 967 / 161 857 E-Mail: moncef@cs.tut.fi

ABSTRACT

A variant of Learning Vector Quantizer (LVQ) based on a multivariate data ordering principle, namely the marginal median Learning Vector Quantizer (MMLVQ), is proposed in order to overcome the drawback that the estimators for obtaining the reference vectors in LVQ do not have robustness either against erroneous choices for the winner vector or against the outliers that may exist in vector-valued observations. The asymptotic properties of MMLVQ are studied as well. It is shown that MMLVQ outperforms the (linear) LVQ with respect to the bias in estimating the true cluster means both for a contaminated Laplacian data model as well as for a contaminated Gaussian data model. As far as the mean-squared estimation error is concerned, it is proven that MMLVQ outperforms the (linear) LVQ in the case of a contaminated Laplacian data model.

1. INTRODUCTION

Neural networks (NN) [1, 2] is a rapidly expanding research field which attracted the attention of scientists and engineers in the last decade. A large variety of artificial neural networks has been developed based on a multitude of learning techniques and having different topologies [2, 3]. One prominent example of neural networks is the Learning Vector Quantizer (LVQ). It is an autoassociative nearest-neighbor classifier which classifies arbitrary patterns into classes using an error correction encoding procedure related to competitive learning [1, 2]. In order to make a distinction between the (standard) LVQ algorithm and the proposed variant that is based on multivariate order statistics, the LVQ algorithm will be called linear LVQ algorithm hereafter.

Let us assume a sequence of vector-valued observations $\mathbf{x}(t) \in \mathbb{R}^p$ and a set of variable reference vectors $\{\mathbf{w}_i(t); \mathbf{w}_i \in \mathbb{R}^p, i = 1, 2, \ldots, K\}$. Let $\mathbf{w}_i(0)$ be randomly initialized. Competitive learning tries to find the best-matching reference vector $\mathbf{w}_c(t)$ to $\mathbf{x}(t)$ (i.e., the winner) where $c = \arg\min_i \parallel \mathbf{x} - \mathbf{w}_i \parallel$ with $\parallel \cdot \parallel$ denoting the Euclidean distance between any two vectors. In the linear LVQ, the weight vectors are updated as blocks concentrated around the winner, i.e.,

$$\begin{aligned} \mathbf{w}_i(t+1) &= \mathbf{w}_i(t) + \alpha(t)[\mathbf{x}(t) - \mathbf{w}_i(t)] \;\; \forall i \in \mathcal{N}_c(t) \\ \mathbf{w}_i(t+1) &= \mathbf{w}_i(t) \;\; \forall i \notin \mathcal{N}_c(t) \end{aligned} \quad (1)$$

where $\alpha(t)$ is the adaptation step and $\mathcal{N}_c(t)$ denotes a neighborhood around the winner. In the following, we use the notation n instead of t to denote discrete events. It can be easily seen that the reference vector for each class $i = 1, \ldots, K$ at time $n+1$ is a linear combination of the input vectors:

$$\mathbf{w}_i(n+1) = \alpha(n)c_i(n)\mathbf{x}(n) + \sum_{k=1}^{n} \alpha(n-k) \cdot \prod_{l=n-k+1}^{n} [1 - \alpha(l)c_i(l)]\, c_i(n-k)\mathbf{x}(n-k) \quad (2)$$

where $c_i(j) = 1$, $i = 1, \ldots, K$, $j = 0, \ldots, n$ if the input vector $\mathbf{x}(j)$ has been assigned to class i, and 0 otherwise. It can be shown than only in the special case of a single data class (i.e., $c_1(l) = 1$, $\forall l \in [0, n]$) and for the adaptation step sequence $\alpha(n) = 1/(n+1)$ the winner vector is the arithmetic mean of the observations that have been assigned to the class (i.e., the maximum likelihood estimator of location). Neither in the case of multiple classes that are normally distributed nor in the case of non-Gaussian multivariate data distributions the linear LVQ yields optimal estimates for the cluster means. In general, linear LVQ and its variations suffer from the following drawbacks:

1. They do not use optimal estimators for obtaining the reference vectors $\mathbf{w}_i$, $i = 1, \ldots, K$ that match the pdf $f_i(\mathbf{x})$ of each class $i = 1, \ldots, K$.
2. They do not have robustness against erroneous choices for the winner vector, since it is well known that linear estimators have poor robustness properties [5, 6].
3. They do not have robustness against the outliers that may exist in the vector observations.

In order to overcome these problems, we propose a variant of Learning Vector Quantizer, the marginal median Learning Vector Quantizer (MMLVQ), that is based on a multivariate data ordering principle. Its asymptotic properties are studied as well.

2. DERIVATION OF MMLVQ AND STUDY OF ITS ASYMPTOTIC PROPERTIES

There is no unambiguous, universally agreeable total ordering of N p-variate samples $\mathbf{x}_1, \ldots, \mathbf{x}_N$ where $\mathbf{x}_i = (x_{1i}, x_{2i}, \ldots, x_{pi})^T$, $i = 1, \ldots, N$. The following so-called sub-ordering principles are discussed in [7]: *marginal ordering, reduced (aggregate) ordering, partial ordering*, and *conditional (sequential) ordering*. In the following, we shall confine ourselves to the marginal ordering principle. The marginal ordering principle implies that the multivariate samples are ordered along each one of the p-dimensions:

$$x_{i(1)} \leq x_{i(2)} \leq \ldots \leq x_{i(N)} \quad i = 1, \ldots, p \quad (3)$$

i.e., the sorting is performed in each channel of the multichannel signal independently.

The *Marginal Median Learning Vector Quantizer* (MMLVQ) relies on the notion of the marginal median order statistic that is defined by:

$$\mathbf{x}_{med} = \begin{cases} \left(x_{1(\nu+1)}, x_{2(\nu+1)}, \ldots, x_{p(\nu+1)}\right)^T & \text{for } N = 2\nu + 1 \\ \left(\frac{x_{1(\nu)} + x_{1(\nu+1)}}{2}, \ldots, \frac{x_{p(\nu)} + x_{p(\nu+1)}}{2}\right)^T & \text{for } N = 2\nu. \end{cases} \quad (4)$$

Let us denote by $\mathbf{X}_i(n)$ the set of the vector observations that have been assigned to each class i, $i = 1, \ldots, K$ until time $n-1$. We find at time n the winner vector $\mathbf{w}_c(n)$ that minimizes $\| \mathbf{x}(n) - \mathbf{w}_i(n) \|$, $i = 1, \ldots, K$. The MMLVQ updates the winner reference vector as follows:

$$\mathbf{w}_c(n+1) = \text{median}\left\{\mathbf{x}(n) \cup \mathbf{X}_c(n)\right\}. \quad (5)$$

where the median operation is given by (4). Thus, all past class assignment sets $\mathbf{X}_i(n)$, $i = 1, \ldots, K$ are needed for MMLVQ.

In the sequel, we shall study the asymptotic properties of MMLVQ in comparison to those of the linear LVQ. It is well known that when the learning procedure of the linear LVQ is led to equilibrium, it results in a partition of the domain of input vector-valued observations called Voronoi tessellation [1, 4]. Let $\mathcal{V}_i(\mathbf{W})$ denote the Voronoi neighborhood of the i-th output neuron with respect to the Euclidean distance metric, i.e.,

$$\mathcal{V}_i(\mathbf{W}) = \{\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^p \mid \| \mathbf{x} - \mathbf{w}_i \| \leq \| \mathbf{x} - \mathbf{w}_l \| \quad l = 1, \ldots, K,\ l \neq i\} \quad (6)$$

where $\mathbf{W} = (\mathbf{w}_1^T \mid \ldots \mid \mathbf{w}_K^T)^T$. The expected stationary state of the linear LVQ is given by [4]:

$$\overline{\mathbf{w}}_i = \mathrm{E}[\mathbf{w}_i] = \frac{\int_{\mathcal{V}_i(\overline{\mathbf{W}})} \mathbf{x} f(\mathbf{x}) d\mathbf{x}}{\int_{\mathcal{V}_i(\overline{\mathbf{W}})} f(\mathbf{x}) d\mathbf{x}} \quad i = 1, \ldots, K. \quad (7)$$

It is seen that (7) gives an implicit definition of the stationary state of LVQ. Let $g_i(\mathbf{x})$ be the conditional pdf of $\mathbf{x}$, on the condition that $\mathbf{x}$ is restricted within the Voronoi neighborhood of class i. It is given by [10]:

$$g_i(\mathbf{x}) = \frac{f(\mathbf{x})}{\int_{\mathcal{V}_i(\mathbf{W})} f(\mathbf{x}) d\mathbf{x}}. \quad (8)$$

Let us denote by $g_{ij}(x_j)$ the marginal pdfs of $g_i(\mathbf{x})$ along each dimension j, $j = 1, \ldots, p$:

$$g_{ij}(x_j) = \int_{\mathcal{V}_i(\mathbf{W})} g_i(\mathbf{x}) d\mathbf{x}_j \quad (9)$$

where $\mathbf{x}_j = (x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_p)^T$. The stationary state of the MMLVQ will be $\tilde{\mathbf{w}}_{Mi} = (\tilde{w}_{Mi1}, \ldots, \tilde{w}_{Mip})^T$ where $\tilde{w}_{Mij}$ is the population median of the marginal distribution $g_{ij}(x_j)$, i.e.,:

$$\int_A^{\tilde{w}_{Mij}} g_{ij}(x_j) dx_j = \int_{\tilde{w}_{Mij}}^B g_{ij}(x_j) dx_j \quad (10)$$

with $[A, B]$ being the domain of $g_{ij}(x_j)$. In order to maintain simplicity in our theoretical analysis the asymptotic properties of both the linear LVQ and the MMLVQ will be considered for 1-d contaminated distribution models of the form:

$$f(x) = \epsilon\, f_1(x) + (1-\epsilon)\, f_2(x) \quad (11)$$

Two contaminated data models are studied, namely, the contaminated Laplacian data model where $f_i(x)$ are Laplacian pdfs given by:

$$f_i(x) = \frac{1}{\sigma\sqrt{2}} \exp\left[-\sqrt{2}\, \frac{|x - m_i|}{\sigma}\right] \quad (12)$$

and the contaminated Gaussian data model with $f_i(x)$ being Gaussian pdfs $N(m_i,\, \sigma)$, $i = 1, 2$. Without any loss of generality, we assume that $m_1 < m_2$. First, the expected stationary state of the MMLVQ is derived and is compared to the expected stationary state of the linear LVQ for the distribution models under study. To this end, the thresholds determined by the linear LVQ and the MMLVQ at the equilibrium for discriminating the two input data classes must be known. In addition, the thresholds determined by the linear LVQ and the MMLVQ should be compared to the threshold predicted by the statistical detection theory, i.e., the threshold T_{opt} that minimizes the probability of false classification [9].

2.1. Contaminated Laplacian model

The probability of false classification is given by:

$$P_e(T) = \epsilon \frac{1}{\sigma\sqrt{2}} \int_T^{+\infty} \exp\left[-\sqrt{2}\frac{|x-m_1|}{\sigma}\right] dx + (1-\epsilon) \frac{1}{\sigma\sqrt{2}} \int_{-\infty}^T \exp\left[-\sqrt{2}\frac{|x-m_2|}{\sigma}\right] dx. \quad (13)$$

In order to differentiate (13) with respect to T, we have to make certain assumptions about T. It can be shown that this problem (i.e., to find T_{opt}) is well defined if and only if $m_1 < T < m_2$, or equivalently if $\frac{1}{\gamma+1} \leq \epsilon \leq \frac{\gamma}{\gamma+1}$ with $\gamma = \exp\left[\frac{\sqrt{2}}{\sigma}(m_2 - m_1)\right]$. In this case, the optimal threshold is given by:

$$T_{\text{opt}} = \frac{m_1 + m_2}{2} + \frac{\sigma}{2\sqrt{2}} \ln\left(\frac{\epsilon}{1-\epsilon}\right). \quad (14)$$

For $\epsilon = 0.5$, we verify that (14) yields $T_{\text{opt}} = (m_1 + m_2)/2 = T_{\text{mid}}$. The Voronoi neighborhoods (6) determined by the linear LVQ are given by:

$$\mathcal{V}_1(\mathbf{W}) = \{x \leq T_{\text{LVQ}}\} \quad \mathcal{V}_2(\mathbf{W}) = \{x > T_{\text{LVQ}}\} \quad (15)$$

with $T_{\text{LVQ}} = \frac{\overline{w}_1 + \overline{w}_2}{2}$. Let us assume that $T_{\text{LVQ}} = T$ is known. Then, by using (7) we obtain:

$$\overline{w}_1 = \frac{1}{F(T)} \left\{\epsilon\, m_1 - \frac{\epsilon}{2}\left(T + \frac{\sigma}{\sqrt{2}}\right) \exp\left[-\sqrt{2}\left(\frac{T-m_1}{\sigma}\right)\right] + \frac{1-\epsilon}{2}\left(T - \frac{\sigma}{\sqrt{2}}\right) \exp\left[\sqrt{2}\left(\frac{T-m_2}{\sigma}\right)\right]\right\} \quad (16)$$

where

$$F(T) = \epsilon\, \left\{1 - \frac{1}{2} \exp\left[-\sqrt{2}\left(\frac{T-m_1}{\sigma}\right)\right]\right\} + \frac{1-\epsilon}{2} \exp\left[\sqrt{2}\left(\frac{T-m_2}{\sigma}\right)\right]. \quad (17)$$

For the second neuron, we have:

$$\overline{w}_2 = \frac{1}{1 - F(T)} \left\{[\epsilon\, m_1 + (1-\epsilon)\, m_2] - F(T)\, \overline{w}_1\right\}. \quad (18)$$

It can easily be shown that $\overline{w}_1 \simeq m_1$ and $\overline{w}_2 \simeq m_2$, if and only if $(m_2 - m_1) \gg \sigma$. Furthermore, for $\epsilon = 0.5$, we have $T_{\text{LVQ}} = T_{\text{mid}}$.

In the case of MMLVQ, it can be proven that its stationary weights are expressed in closed-form formulae. Let $\zeta = \exp\left[\sqrt{2}\left(\frac{m_1 - m_2}{\sigma}\right)\right]$. If $\Phi_1 = \epsilon + (1-\epsilon)\,\zeta$ and $\Delta_1 = [F(T) - 2\epsilon]^2 + 4\,\epsilon\,(1-\epsilon)\,\zeta$, then the stationary weight of the first neuron of MMLVQ at the equilibrium is given by:

$$\tilde{w}_{M1} = m_1 + \frac{\sigma}{\sqrt{2}} \ln\left[\frac{F(T)}{\Phi_1}\right] \quad \text{if } F(T) \leq \Phi_1$$

$$\tilde{w}_{M1} = m_1 + \frac{\sigma}{\sqrt{2}} \ln\left[\frac{F(T) - 2\epsilon + \sqrt{\Delta_1}}{2\,(1-\epsilon)\,\zeta}\right] \quad \text{if } F(T) > \Phi_1. \quad (19)$$

In an analogous fashion, it can be proven that if $\Phi_2 = \epsilon(1-\zeta)$ and $\Delta_2 = [1 + F(T) - 2\epsilon]^2 + 4\epsilon(1-\epsilon)\,\zeta$ then, the stationary weight of the second neuron of the MMLVQ at the equilibrium is given by:

$$\begin{aligned} \tilde{w}_{M2} &= m_2 - \frac{\sigma}{\sqrt{2}} \ln\left[\frac{[2\epsilon - 1 - F(T)] + \sqrt{\Delta_2}}{2\,\epsilon\,\zeta}\right] \\ & \qquad \text{if } F(T) < \Phi_2 \\ \tilde{w}_{M2} &= m_2 - \frac{\sigma}{\sqrt{2}} \ln\left[\frac{1 - F(T)}{1 - \Phi_2}\right] \\ & \qquad \text{if } F(T) \geq \Phi_2. \quad (20) \end{aligned}$$

The sets of equations (16), (18) and (19), (20) define implicitly the stationary state of both the linear LVQ and the MMLVQ in terms of T. The following simple algorithm for solving the above-mentioned sets of equations is proposed:

1. Begin with arbitrary $\overline{w}_1$ and $\overline{w}_2$
2. Set T equal to the midpoint of $\overline{w}_1$ and $\overline{w}_2$.
3. Re-evaluate $\overline{w}_1$ and $\overline{w}_2$ for the specific T of Step 2.
4. If the absolute error in $\overline{w}_1$ and $\overline{w}_2$ between two successive iterations exceeds a proper threshold, go to Step 2.

(For the MMLVQ, $\overline{w}_1$ and $\overline{w}_2$ are replaced by $\tilde{w}_{M1}$ and $\tilde{w}_{M2}$ respectively.) Figure 1 depicts the thresholds determined by the linear LVQ, the MMLVQ as well as the optimal threshold predicted by the statistical detection theory for a data set having pdf (11) with $m_1 = 5$, $m_2 = 10$ and $\sigma = 3$ for several $\epsilon \in [0.2,\ 0.8]$. The bias in estimating the mean of the dominating cluster, i.e., the quantity

$$\begin{array}{lll} |\overline{w}_2 - m_2| & (\text{or, } |\tilde{w}_{M2} - m_2|\ ) & \text{for } \epsilon \leq 0.5, \text{ and} \\ |\overline{w}_1 - m_1| & (\text{or, } |\tilde{w}_{M1} - m_1|\ ) & \text{for } \epsilon > 0.5 \end{array} \quad (21)$$

is plotted in Figure 2. In the same figure, we have also included the bias that the conditional means which correspond to the decision regions predicted by the statistical detection theory yield. It is clear that the MMLVQ produces asymptotically less bias than the linear LVQ.

From Figure 2, it is evident that the linear LVQ and the MMLVQ are not unbiased estimators of the data cluster means. Accordingly, the asymptotic variance $V(\mathcal{T}, F)$, $\mathcal{T} =$ LVQ, MMLVQ defined by $V(\mathcal{T}, F) = \int \text{IF}(\mathbf{x}; \mathcal{T}, F)^2 f(\mathbf{x}) d\mathbf{x}$ where $\text{IF}(\mathbf{x}; \mathcal{T}, F)$ is the influence function of $\mathcal{T}$ at F [5, 8] does not take into account the bias introduced by each estimator, since it is simply the variance of the random variable $\sqrt{n}(\mathcal{T}_n - \mathcal{T}(F))$ that is normally distributed with zero-mean as $n \to \infty$ [6]. Observe that the asymptotic variance of the estimator $\mathcal{T}$ at model F is essentially the upper bound of its variance, i.e., $V(\mathcal{T}, F) = \max_n \text{E}\left[(\mathcal{T}_n - \mathcal{T}(F))^2\right] = \text{E}\left[(\mathcal{T}_n - \mathcal{T}(F))^2\right]|_{n=1}$. Therefore, the asymptotic relative efficiency (ARE) of LVQ and MMLVQ defined by:

$$\text{ARE(MMLVQ, LVQ)} = \frac{V(\text{LVQ}, F)}{V(\text{MMLVQ}, F)} \quad (22)$$

is not appropriate for comparing the performance of the two estimators. We propose the following modified ARE:

$$\widetilde{\text{ARE}}(\text{MMLVQ, LVQ}) = \frac{\max_n \text{E}\left[(\text{LVQ}_n - \mathbf{M})^2\right]}{\max_n \text{E}\left[(\text{MMLVQ}_n - \mathbf{M})^2\right]} \quad (23)$$

where $\mathbf{M} = (\mathbf{m}_1 \mid \ldots \mid \mathbf{m}_K)^T$ is the vector of the unconditional means to be estimated. The modified ARE (23) has been evaluated for the contaminated Laplacian distribution model under study. In Figure 3, the modified ARE is plotted for several $\epsilon \in [0.2,\ 0.8]$ and σ. It can be seen that MMLVQ outperforms the linear LVQ with respect to the mean-squared estimation error as well.

2.2. Contaminated Gaussian model

For a 1-d contaminated Gaussian model, it can be shown that the optimal threshold is given by:

$$T_{\text{opt}} = \frac{m_1 + m_2}{2} - \frac{\sigma^2}{m_1 - m_2} \ln\left(\frac{\epsilon}{1-\epsilon}\right). \quad (24)$$

Let $\text{erf}(a)$ be the error function $\text{erf}(a) = \frac{1}{\sqrt{2\pi}} \int_0^a \exp(-\frac{t^2}{2}) dt$ [10]. In an analogous fashion, the stationary weight of the first neuron of the linear LVQ can be expressed in terms of $T_{\text{LVQ}} = T$ as follows:

$$\begin{aligned} \overline{w}_1 = \tfrac{1}{F(T)} \Big\{ & \tfrac{1}{2}\left[\epsilon\, m_1 + (1-\epsilon)\, m_2\right] + \Big[\epsilon\, m_1 \text{erf}\left(\tfrac{T-m_1}{\sigma}\right) + \\ & +(1-\epsilon)\, m_2\, \text{erf}\left(\tfrac{T-m_2}{\sigma}\right)\Big] - \tfrac{\sigma}{\sqrt{2\pi}} \Big[\epsilon\, \exp\left[-\tfrac{1}{2}\left(\tfrac{T-m_1}{\sigma}\right)^2\right] \\ & +(1-\epsilon)\, \exp\left[-\tfrac{1}{2}\left(\tfrac{T-m_2}{\sigma}\right)^2\right]\Big]\Big\} \end{aligned} \quad (25)$$

where

$$F(T) = \frac{1}{2} + \epsilon\, \text{erf}\left(\frac{T - m_1}{\sigma}\right) + (1-\epsilon)\, \text{erf}\left(\frac{T - m_2}{\sigma}\right). \quad (26)$$

The stationary weight for the second neuron is given by (18). Unlike the contaminated Laplacian model examined above, the stationary state of MMLVQ is now defined implicitly by:

$$\begin{aligned} & \frac{1}{F(T)} \left[\frac{1}{2} + \epsilon\, \text{erf}\left(\frac{\tilde{w}_{M1} - m_1}{\sigma}\right) + (1-\epsilon) \cdot \right. \\ & \left. \cdot\ \text{erf}\left(\frac{\tilde{w}_{M1} - m_2}{\sigma}\right)\right] = 0.5 \quad (27) \\ & \frac{1}{1 - F(T)} \left[\frac{1}{2} - \epsilon\, \text{erf}\left(\frac{\tilde{w}_{M2} - m_1}{\sigma}\right) - (1-\epsilon) \cdot \right. \\ & \left. \cdot\ \text{erf}\left(\frac{\tilde{w}_{M2} - m_2}{\sigma}\right)\right] = 0.5. \quad (28) \end{aligned}$$

The algorithm proposed above for deriving the stationary weights of LVQ/MMLVQ in the case of a contaminated Laplacian model can easily be modified to solve the set of equations (27), (28) as well. It has been found that MMLVQ introduces smaller bias in estimating the mean value of the dominating cluster than the linear LVQ for any ϵ. For $\epsilon < 0.5$ (or, $\epsilon > 0.5$) and Gaussian pdfs with little overlap, the smaller bias of MMLVQ is attributed to a threshold shift so that $T_{\text{MMLVQ}} < T_{\text{LVQ}}$ ($T_{\text{MMLVQ}} > T_{\text{LVQ}}$). When the Gaussian pdfs have significant overlap, the smaller bias of MMLVQ is due to the replacement of the conditional means by the conditional medians. For $\epsilon = 0.5$, both the statistical detection theory as well as the LVQ and the MMLVQ result in the same threshold. Therefore, they attain the same probability of false classification. However, MMLVQ introduces smaller bias in estimating the unconditional means of both clusters than the LVQ. However, the linear LVQ outperforms the MMLVQ with respect to the mean-squared estimation error. Although, the performance of the MMLVQ is improved as σ increases, even in this case, the linear LVQ is better than the MMLVQ.

3. EXPERIMENTAL RESULTS

We have tested the performance of the proposed MMLVQ against the one of the linear LVQ in the case of an 1-d data set that is described by the pdf

$$f(x) = pU(-5,\ 20) + (1-p)\left[\epsilon\, f_1(x) + (1-\epsilon)\, f_2(x)\right] \quad (29)$$

where $U(-5, 10)$ denotes a uniform pdf in the interval $[-5,\ 20]$ and $f_i(x)$ are 1-d Laplacian pdfs (12) for $p = 0.2$, $\epsilon = 0.5$, $m_1 = 5$, $m_2 = 10$ and $\sigma = 2$. A data set having pdf (29) is a severely corrupted one due to the presence of the uniformly distributed outliers with probability of occurrence $p = 0.2$. Figure 4 depicts the output bias of the neuron that attempts to estimate the unconditional mean of the first data class described by $f_1(x)$ both for the MMLVQ (i.e., $|\mathrm{E}[w_{M1}(n)] - 5|$) as well as for the linear LVQ (i.e., $|\mathrm{E}[w_1(n)] - 5|$) versus the iteration number n. It is clear that the outliers produce more output bias in the case of the linear LVQ than in the case of the MMLVQ. The expected values of the LVQ and MMLVQ outputs have been found by averaging the results on 1000 independent runs.

REFERENCES

[1] **T.K. Kohonen (1989)**: Self-Organization and Associative Memory, 3rd ed. *Springer-Verlag*.

[2] **P.K. Simpson (1990)**: Artificial Neural Systems, *Pergamon Press*.

[3] Special issue on neural networks, I: Theory & modeling; II: Analysis, techniques, & applications, *Proceedings of the IEEE*, **78** (9-10), 1409–1680.

[4] **H. Ritter and K. Schulten (1989)**, Convergence properties of Kohonen's topology conserving maps: Fluctuations, stability and dimension selection, *Biological Cybernetics*, **60**, 59–71.

[5] **P.J. Huber (1981)**: Robust Statistics, *J. Wiley*.

[6] **E.L. Lehman (1983)**: Theory of point estimation, *J. Wiley*.

[7] **V. Barnett (1976)**, The ordering of multivariate data, *J. R. Statist. Soc. A*, **139** (3), 318–354.

[8] **I. Pitas and A. N. Venetsanopoulos (1990)**: Nonlinear Digital Filters: Principles and Applications, *Kluwer Academic Publishers*.

[9] **H.L. Van Trees (1968)**: Detection, Estimation and Modulation Theory, *J. Wiley*.

[10] **A. Papoulis (1984)**: Probability, Random Variables and Stochastic Processes, *McGraw-Hill*.

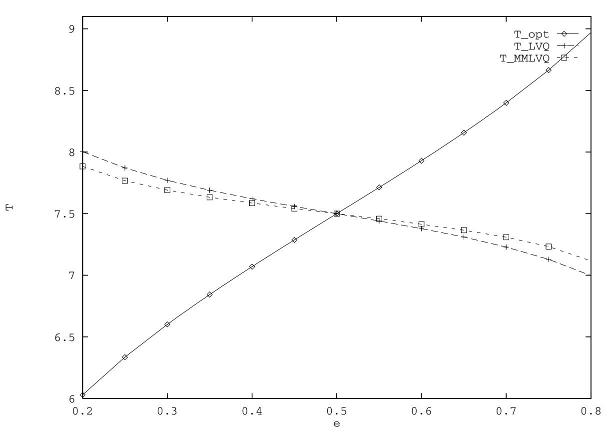

Figure 1

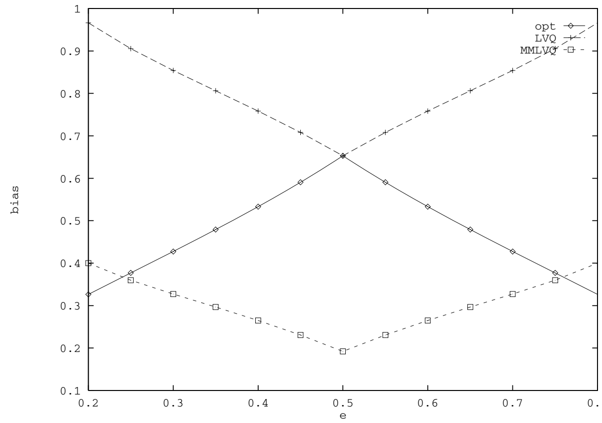

Figure 2

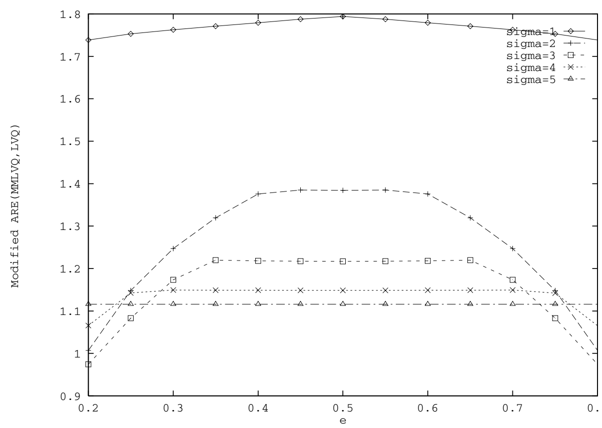

Figure 3

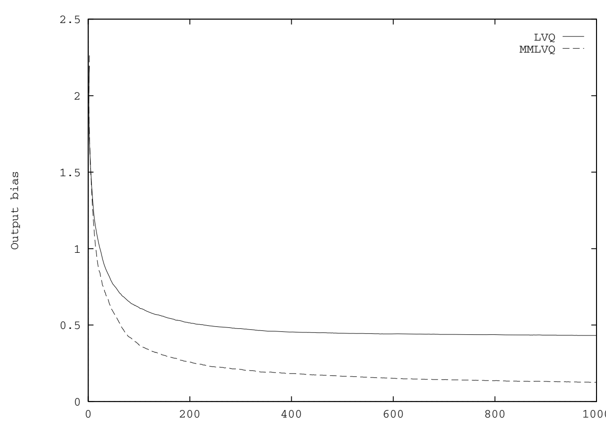

Figure 4