JB 90-11

Paper Entitled:

"Convergence of the Vectors in
Kohonen's Learning Vector Quantization"

(with T. LaVinga)

Submitted to:

INNC 90

Paris, France

July 1990

pp. 1028-1031

Convergence of the Vectors in Kohonen's Learning Vector Quantization

John S. Baras and Anthony LaVigna
Systems Research Center, University of Maryland
College Park, Maryland 20742

Abstract

Kohonen's Learning Vector Quantization is a nonparametric classification scheme which classifies observations by comparing them to k templates called Voronoi vectors. The locations of these vectors are determined from past labeled data through a learning algorithm. When learning is complete, the class of a new observation is the same as the class of the closest Voronoi vector. Hence LVQ is similar to nearest neighbors, except that instead of all of the past observations being searched only the k Voronoi vectors are searched.

In this paper, we show that the LVQ learning algorithm converges to asymptotically stable zeros of an ordinary differential equation. It is shown that the learning algorithm performs stochastic approximation. Convergence of the vectors is guaranteed under the appropriate conditions on the underlying statistics of the classification problem. We also present a modification to the learning algorithm which results in more robust convergence.

1.1 Learning Vector Quantization

The LVQ algorithm is now described. Let $\{(x_i, d_{x_i})\}_{i=1}^N$ be the training data or past observation set. This means that x_i is observed when pattern d_{x_i} is in effect. Let θ_j be a Voronoi vector and let $\Theta = \{\theta_1, \ldots, \theta_k\}$. We assume that there are many more observations than Voronoi vectors (Duda & Hart [1973]). Once the Voronoi vectors are initialized, training proceeds by taking a sample (x_j, d_{x_j}) from the training set, finding the closest Voronoi vector and adjusting its value according to equations (1) and (2). After several passes through the data, the Voronoi vectors converge and training is complete.

Suppose θ_c is the closest vector. Adjust θ_c as follows:

$$\theta_c(n+1) = \theta_c(n) + \alpha_n \left(x_j - \theta_c(n)\right) \tag{1}$$

if $d_{\theta_c} = d_{x_j}$ and

$$\theta_c(n+1) = \theta_c(n) - \alpha_n \left(x_j - \theta_c(n)\right) \tag{2}$$

if $d_{\theta_c} \neq d_{x_j}$. The other Voronoi vectors are not modified.

This update has the effect that if x_j and θ_c have the same decision then θ_c is moved closer to x_j, however if they have different decisions then θ_c is moved away from x_j. The constants $\{\alpha_n\}$ are positive and decreasing, e.g., $\alpha_n = 1/n$.

1.2 Convergence of the Learning Algorithm

The LVQ algorithm has the general form

$$\theta_i(n+1) = \theta_i(n) + \alpha_n \, \gamma(d_{x_n}, d_{\theta_i(n)}, x_n, \Theta_n) \; (x_n - \theta_i(n)) \tag{3}$$

where x_n is the currently chosen past observation. The function γ determines whether there is an update and what its sign should be. It is given by

$$\gamma(d_{x_n}, d_{\theta_i}, x_n, \Theta_n) = 1_{\{x_n \in V_{\theta_i}\}}(1_{\{d_{x_n}=d_{\theta_i}\}} - 1_{\{d_{x_n}\neq d_{\theta_i}\}}). \tag{4}$$

Here $1_{\{\}}$ represents the indicator function and V_{θ_j} represents the set of points closest to θ_j.

The update in (3) is a stochastic approximation algorithm (Benveniste, Metivier & Priouret [1987]). It has the form

$$\Theta_{n+1} = \Theta_n + \alpha_n \; H(\Theta_n, z_n) \tag{5}$$

where Θ is the vector with components θ_i; $H(\Theta, z)$ is the vector with components defined in the obvious manner from (3) and z_n is the random pair consisting of the observation and the associated *true* pattern number. If the appropriate conditions are satisfied by α_n, H, and z_n, then Θ_n approaches the solution of

$$\frac{d}{dt}\bar{\Theta}(t) = h(\bar{\Theta}(t)) \tag{6}$$

for the appropriate choice of $h(\Theta)$.

Let $p_i(x)$ represent the pattern density for pattern i and let π_i represent its prior. Suppose there are ℓ patterns. It can be shown (Kohonen [1986]) that

$$h_i(\Theta) = \int_{V_{\theta_i}} (x - \Theta_i)\, p_i(x)\, \pi_i \, dx - \sum_{\substack{j=1 \\ j\neq d_{\theta_i}}}^{\ell} \int_{V_{\theta_j}} (x - \Theta_j)\, p_j(x)\, \pi_j \, dx \tag{7}$$

The following hypotheses are assumed:

[H.1] $\{\alpha_n\}$ is a nonincreasing sequence of positive reals such that $\sum_n \alpha_n = \infty$, $\sum_n \alpha_n^\lambda < \infty$.

[H.2] Given d_{x_n}, x_n are independent and distributed according to $p_{d_{x_n}}(x)$.

[H.3] The pattern densities, $p_i(x)$, are continuous.

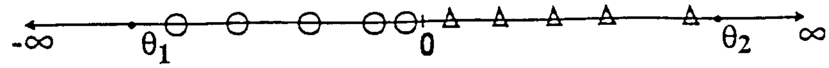

Figure 1: *A possible distribution of observations and two Voronoi vectors.*

With these assumptions it is possible, using techniques from (Benveniste, Metivier & Priouret [1987]) or (Kushner & Clark [1978]), to prove the following theorem.

Theorem 1 *Assume that [H.1]–[H.3] hold. Let $\bar{\Theta}^*$ be a locally asymptotic stable equilibrium point of* (6) *with domain of attraction D^*. Let Q be a compact subset of D^*. If $\Theta_n \in Q$ for infinitely many n then*

$$\lim_{n\to\infty} \Theta_n = \bar{\Theta}^* \tag{8}$$

Proof: (see (LaVigna [1989]))

Hence if the initial locations and decisions of the Voronoi vectors are close to a locally asymptotic stable equilibrium of (6) and if they do not move too much then the vectors converge.

1.3 Modified LVQ Algorithm

The convergence results above require that the initial conditions are close to the stable points of (6) in order for the algorithm to converge. In this section we present a modification to the LVQ algorithm which increases the number of stable equilibrium for equation (6) and hence increases the chances of convergence. First we present a simple example which emphasizes a defect of LVQ and suggests an appropriate modification to the algorithm.

Let ◯ represent an observation from pattern 2 and let △ represent an observation from pattern 1. We assume that the observations are scalar. Figure 1 shows a possible distribution of observations. Suppose there are two Voronoi vectors θ_1 and θ_2 with decisions 1 and 2, respectively, initialized as shown in Figure 1. At each update of the LVQ algorithm, a point is picked at random from the observation set and the closest Voronoi vector is modified. We see that during this update, $\theta_2(n)$ is pushed towards ∞ and $\theta_1(n)$ is pushed towards $-\infty$, hence the Voronoi vectors do not converge.

This divergence happens because the decisions of the Voronoi vectors do not agree with the majority vote of the observations closest to each vector. As a result, the Voronoi vectors are pushed away from the origin. This phenomena occurs even though the observation data is bounded. The point here is that, if the decision associated with a Voronoi vector does not agree with the majority vote of the observations closest to that vector then it is possible for the vector to diverge. A simple solution to this problem is to correct the decisions of all the

Voronoi vectors after every adjustment so that their decisions correspond to the majority vote. In practice this correction would only be done during the beginning iterations of the learning algorithm since that is when α is large and the Voronoi vectors are moving around significantly. With this modification it is possible to show convergence to the Bayes optimal classifier (LaVigna [1989]) as the number of Voronoi vectors become large.

1.4 Conclusions

We have shown convergence of the Voronoi vectors in the LVQ algorithm. We have also presented the majority vote modification of the LVQ algorithm. This modification prevents divergence of the Voronoi vectors and results in convergence for a larger set of initial conditions. In addition, with this modification it is possible to show that as the appropriate parameters go to infinity the decision regions associated with the modified LVQ algorithm approach the Bayesian optimal (LaVigna [1989]).

1.5 Acknowledgements

This work was supported by the National Science Foundation through grant CDR-8803012, Texas Instruments through a TI/SRC Fellowship and the Office of Naval Research through an ONR Fellowship.

1.6 References

A. Benveniste, M. Metivier & P. Priouret [1987], *Algorithmes Adaptatifs et Approximations Stochastiques*, Mason, Paris.

R. O. Duda & P. E. Hart [1973], *Pattern Classification and Scene Analysis*, John Wiley & Sons, New York, NY.

T. Kohonen [1986], "Learning Vector Quantization for Pattern Recognition," Technical Report TKK-F-A601, Helsinki University of Technology.

H. J. Kushner & D. S. Clark [1978], *Stochastic Approximation Methods for Constrained and Unconstrained Systems* , Springer-Verlag, New York–Heidelberg–Berlin.

A. LaVigna [1989], "Nonparametric Classification using Learning Vector Quantization," Ph.D. Dissertation, Department of Electrical Engineering, University of Maryland.