LEARNING AND SELF LEARNING PROCEDURES FOR AUTOMATIC CLASSIFICATION

by

J. Aguilar-Martin

Memorandum No. UCB/ERL M78/51

1 September 1978

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

I. General Concepts Involved in a Learning Process

Learning is a very controversial concept and in human behavior its complexity is far beyond our purpose.

Several processes, as creating habits or conditioning, are directed towards repetition and, although they play an important role in natural learning, we shall not look at them in our study. Other processes that can be found in natural learning are of a more deductive kind, as extraction of relevant features, inference of relations, or partitions of the environment. We shall focus our interest in this second kind of process. Learning is then considered as discovering a hidden concept or object from an informative environment.

Perception and memorization are functions that must be present as well as, possibly, a reward function or evaluation of the quality of the decisions based on learning. The elementary situation can be described in the following way: A set of objects $X_1 \ldots . X_N$ is "percepted" sequentially and yields a sequence of observations $\left\{y_s\right\}_{s=t_o}^{t} = Y^t$ to the system.

The learning process consists in the construction of

α) an estimation of the set of objects

β) an inferred relational structure on the set of estimated objects

both based on the past information Y^t. That enables the system

α) to "recognize" the source of a new observation y_{t+1}, i.e., to decide from which object X_i this observation might come. This recognition is obviously subject to eventual error

β) to "produce" estimated objects consistent with the inferred relational structure.

Sequential learning is restricted to the situation where no more than one observation can be proceeded at the same time, and no storage of observations is available. The system must start with an á priori guess or after a previous learning period; the next observation gives a new amount of information and it must confirm or contradict the previous concepts memorized.

Two functions appear:

- modification of previous memory (estimation)
- modification of the confidence (or belief) in this previous memory

The closeness of a new observation to an expected observation and its frequence are the working materials from which these two functions can be performed. Learning should then always include a statistic or counting procedure and in many cases the essential feature will be the estimation of a probability law.

II. Classifications and Partitions

A classifier is a system that, given a set Ω, gives a class C_i to any element $\omega \in \Omega_j$. From that results a Partition $\mathcal{P} = \left\{C_i\right\}_{i=1}^{N}$

A Partition $\mathcal{P}$ is equivalent to a set of characteristic functions

$$\mathcal{X} = \left\{\chi_i(\omega)\right\}_{i=1}^{N} \quad \text{where} \quad \chi_i(\omega) = \begin{matrix} 1 \text{ if } \omega \in C_i \\ 0 \text{ if } \omega \notin C_i \end{matrix}$$

to which we add a decision rule $\mathcal{R}(\cdot|\mathcal{P})$ or classifier. A decision rule is a mapping $\Omega \rightarrow \mathcal{P}$ and $\mathcal{R}(\omega|\mathcal{P}) = C_j$ such that $\omega \in C_j$. The equivalence relation generated by this decision rule is

$$\alpha \cdot Eq \cdot \omega \Leftrightarrow \mathcal{R}(\alpha|\mathcal{P}) = \mathcal{R}(\omega|\mathcal{P}) \text{ and } \mathcal{P} \text{ is the quotient set } \Omega|\mathcal{R}(\cdot|\mathcal{P})$$

This decision rule can be described as

$$\left\{ C_j = \mathcal{R}(\omega|\mathcal{P}) \;\middle|\; \chi_j(\omega) = \max_i \chi_i(\omega) \right\}$$

We can then replace the characteristic function $\left\{ \chi_i \right\}_{i=1}^{N}$ by a set of membership functions $\mathcal{M} = \left\{ \mu_i \right\}_{i=1}^{N}$ without altering the partition, as far as the decision rule remains unchanged; that is, if

$$\mu_j(\omega) = \max_i \mu_i(\omega) \Leftrightarrow \chi_j(\omega) = \max_i \chi_i(\omega) \qquad \forall\, \omega$$

Figure 1 shows a simple case where this situation holds. In that case we shall write the decision rule as follows:

$$\mathcal{R}(\omega|\mathcal{P}) = \mathcal{R}(\omega|\mathcal{M})$$

A particularly interesting case is when all functions μ_i belong to the same parametric family such that $\mu_i(\omega) = f(\omega, \theta_i)$. Then the set $\mathcal{M}$ is completely determined by the set of parameters

$$\textcircled{H} = \left\{ \theta_i \right\}_{i=1}^{N}$$

and we may represent the classifier as $\mathcal{R}(\omega|\textcircled{H})$.

To determine the partition $\mathcal{P}$ is, therefore, the same as to find the set of functions $\mathcal{M}$ or, if parametization is possible, to find the set of parameters $\textcircled{H}$, and then to reduce to a problem of parameter estimation. To transpose the analysis of parameter estimation to partition estimation is not an easy task because a metric must be defined on partitions and convergence theorems can only be deduced in that space. An attempt is being made in a future publication.

III. Learning a Partition

Figure 2 gives the learning process that might be fully (1) or partially (2) directed by a teacher, or independent (3) in case of self learning. In (1) the learning system takes the information processed by the teacher without taking into account the classifier's decision. In (3) no teacher is available and the classifier's decision is considered correct. An intermediate situation is shown in (2) where an evaluation of the error takes place, and a great number of situations may then be generated.

It has to be noticed that the estimated partition that stays in memory is not necessarily the partition of the already processed observations, nor is it the teacher's partition. This estimated partition is a set of estimated functions $\hat{\mathcal{M}}_t = \left\{\hat{\mu}_i(t)\right\}_{i=1}^{N}$ or estimated parameters $\hat{\Theta}_t = \left\{\hat{\theta}(t)\right\}_{i=1}^{N}$ based on past information Y^t. They are measurable functions in the σ algebra induced by Y^t.

We shall call historical partition the record of the past decision rules applied only to the elements of Y^t. As is normal in natural learning, an element can be misclassified in the early stages of the process and, therefore, is likely to change its class as the quality of the estimated partition improves.

IV. Self Learning Algorithms for Automatic Classification

We shall give three examples of learning algorithms in this section:

a) classification of strings of characters of a given length;

b) classification (or clustering) of points in an euclidian space distributed with a multi-gaussian law;

c) classification of sets (strings, arrays...) of imperfectly perceived dichotomies with application to image processing.

For the first two examples, only a quick description will be given here as they will appear wholly developed in [7]. The last algorithm will be given more attention as it involves the concept of probability of membership and is presented here for the first time.

V. Classification of Strings of Characters of a Given Length

Let us suppose that the alphabet is composed of n_v characters and that the length of the strings to be classified is n_p. If we order our alphabet arbitrarily, we may construct the possible array that represents the set $\mathcal{X}$ of all the possible strings or events X. (Figure 3)

A particular event (or string) can be characterized by such an array where one and only one case is darkened in each column. An example is given in Figure 4. Assuming that strings can come from different classes $C_1, C_2 \ldots C_k \ldots C_N$, we shall characterize each class by $P_k = P(C_k)$ matrix $n_v \times n_p$ whose elements are:

$$p^k(i,j) = \text{prob}\left[j^{th} \text{ character in string} = i^{th} \text{ in alphabet/given that the string comes from } C_k\right]$$

If no dependence between characters of a same string is known, we shall treat them as independent and then

$$\text{prob}\left[\text{string } X | \text{comes from } C_k\right] = \prod_{j=1}^{n_p} p^k(x_j,j) = p(X|k)$$

We may now define a decision rule or classifier $\mathcal{R}\left[X|\left\{P_k\right\}_{k=1}^{N}\right] = C_j$

where C_j is such that $p(X|j) = \max_k p(X|k)$. The set of matrices P and this rule are a partition of the set $\mathcal{X}$.

Estimation of p

If probability is defined as number of favorable events/total number of events, we have at time $t \in \left[t_0, t_1 \ldots t_N\right]$

$$p^k(i,j) = \frac{\text{Number of times } x_j = i}{\text{Number of elements in } C_k} = \frac{n^k_{ij}}{n_k}$$

Each time a string is classified in class C_k, either by self learning or by a teacher, n_k becomes n_k+1 and for the i's such that $i = x_j$ we shall have $n^k_{x_j,j} = n^k_{x_j,j} + 1$

Initialization

When a teacher is available, he can start with $n_k = 0$ and $n^k_{ij} = 0$ as no decision depends on those values. But in self learning, the classifier's decision is used to update the partition and the number of classes is not known in advance, so we must define an "empty" class C_o that gives an equal probability to all strings. This class is such that

$$p^o_{ij} = \frac{1}{n_v} = \frac{n^o_{ij}}{n_o}$$

An element will be classified in this class when $\left(p^o_{ij}\right)^{n_p} = \left(\frac{1}{n_v}\right)^{n_p}$ is greater than any $p(X|k)$. Then a new class C_{N+1} will be created such that

$$p^{N+1}_{ij} = \frac{n^o_{ij}+1}{n_o+1} \quad \text{if} \quad i = x_j$$

$$= \frac{n^o_{ij}}{n_o+1} \quad \text{if} \quad i \neq x_j$$

The choice of n^o_{ij} is subjective and determines the "intensity" of the influence of the first element classified in the new class.

VI. Classification of Points in an Euclidian Space

We shall give here a schematic view of an algorithm based on estimation of gaussian probabilities using filtering equations. Let us take the space $\Omega = \mathcal{R}^n$. A fixed point of this space is $x \in \Omega$ and if $C_i \in \mathcal{P}$, we shall write $x^i \in C_i$. We consider at each instant t a noisy observation in the form $y(t) = x^j(t) + v^j(t)$. $v^j(t)$ is a gaussian white noise of given covariance R^j, where j is to be determined. Moreover, we suppose that all x^i for $i = 1, 2...N$ are elements of gaussian populations, the means of which are stationary and we modelize that by a pseudo-dynamic equation $x^i(t+1) = x^i(t) + w^i(t)$

where $w^i(t)$ are gaussian white noises with generally unknown covariances Q^i. In this particular case those covariances shall be zero. The set of probability density functions

$$\left\{p\left[x^i | y(0), y(1)...y(t)\right]\right\}_{i=1}^{N}$$

and the maximum likelihood decision rule gives an estimated partition at time t and, as all those densities are gaussian, the set of parameters

$\left\{\bar{x}^i, \; R^i\right\}_{i=1}^{N}$ is sufficient, where

$$\bar{x}^i = E\left[x^i | y(0)...y(t)\right] \quad \text{and} \quad R^i = \text{cov}\left[x^i | y(0)...y(t)\right]$$

Adaptive estimation of those parameters replaces the self learning classification problem. For this purpose Kalman-Bucy filter equations are used where recursive estimators for covariances are added. The overall equations give a set or "battery" of suboptimal filters and asymptotic unbiasedness is proved in [2]. A threshold must be defined, if self learning is to be achieved, that prevents making decisions based on too small values of probability densities, and enables us to create a new class or cluster when needed.

VII. Classification of Sets of Imperfectly Perceived Dichotomies (fuzzy)

Probabilized dichotomies

Let us define a probabilized set $(\Omega \mathcal{A} P)$, a subset $A \subset \Omega$, $A \in \mathcal{A}$, defines a dichotomy in Ω as for any element $\omega \in \Omega$, $\{\omega \in A \text{ or } \omega \notin A\}$ is true. The characteristic function $\chi_A(\omega) = \begin{matrix} 1 \text{ if } \omega \in A \\ 0 \text{ if } \omega \notin A \end{matrix}$ is a measurable function $\Omega \to \{0,1\}$ and defines a random variable $x = \chi_A(\omega)$ whose probability law is

$$\text{prob}\,[x=1] = P(A)$$

$$\text{prob}\,[x=0] = 1-P(A)$$

Let us call $\rho = P(A)$ and then we can write the probability distribution of x in the following way:

$$p[x] = \rho^x(1-\rho)^{(1-x)}, \quad \rho \in \left]0,1\right[\text{ and } x \in \{0,1\}$$

Imperfectly perceived dichotomies

An imperfect observer is one that does not perceive the exact position of the boundaries of the set A but who can point to at least one element $\omega_{in} \in A$ and at least one element $\omega_{ex} \notin A$. We shall, therefore, associate a function such that $\mu_A(\omega_{in}) = 1$

$$\mu_A(\omega_{ex}) = 0$$

$$\text{and } 0 \leq \mu_A(\omega) \leq 1 \quad \forall \omega \in \Omega$$

This is a membership function and defines a fuzzy set $\underset{\sim}{A}$; $\mu_A(\omega)$ is a measurable function $\Omega \to [0,1]$ where the dichotomy space $\{0,1\}$ has been replaced by the unit interval. A particular practical case is when only a finite number of points of this interval are considered such that the dichotomy is replaced by a discrete membership $\mu_A(\omega)$, measurable function $\Omega \rightarrow \left\{0, v_1, v_2 \ldots v_\alpha, 1\right\}$ where $0 \leq v_i \leq v_j \leq 1$ if $j > i$

We shall first study the continuous interval $\left[0,1\right]$. In any of those cases $x = \mu_A(\omega)$ is a random variable. To keep the same type of probability low as in the perfect dichotomy, we shall define the following probability density for x:

$$p[x] = K(\rho)\ \rho^x (1-\rho)^{1-x}$$

where the constant factor $K(\rho)$ stands for normalization; that is:

$$\frac{1}{K(\rho)} = \int_o^1 \rho^x (1-\rho)^{1-x}\ dx = \frac{2\rho-1}{\text{Log}\left(\frac{\rho}{1-\rho}\right)}$$

Any other function $f(x)$ such that $f(0)=(1-\rho)$ and $f(1)=\rho$ (and particularly the linear function $f(x) = x\rho + (1-x)(1-\rho)$) could be chosen. The interest of this particular function will be justified later when studying the discrete space of membership.

Estimation of the probability law of the membership grade

The chosen probability law depends on the parameter ρ, and can be written $p[x|\rho]$. In the case where this parameter is unknown, we shall be interested in its estimation based on a number of available observations, $\left\{y_1, \ldots y_M\right\}$. If we choose the maximum likelihood estimator $\hat{\rho}^M$ such that

$$p\left[y_1, \ldots y_M | \hat{\rho}^M\right] = \max_\rho\ p\left[y_1, \ldots y_M | \rho\right]$$

It appears that

$$\hat{\rho}^M = \frac{1}{N} \sum_{i=1}^{M} y_i$$

because

$$p\left[y_1 \ldots y_M | \rho\right] = \rho^{\overset{M}{\Sigma} y_i} (1-\rho)^{\left(M - \overset{M}{\Sigma} y_i\right)} K^M(\rho)$$

This is the main justification of our choice because it extends, in the simplest way, the algorithm for the classification of strings of characters, replacing the counting process by recursive calculation in the following way: $\hat{\rho}^{M+1} = \hat{\rho}^{M} + \frac{1}{M+1}\left(y_{M+1} - \hat{\rho}^{M}\right)$

Similar to the recursive counting, a modification can be introduced to avoid steady state insensibility to new data, replacing this statistical equation by

$$\hat{\rho}^{M+1} = \hat{\rho}^{M} + \alpha(M)\left(y_{M+1} - \hat{\rho}^{M}\right)$$

where $\alpha(M)$ can be a constant or any decreasing function of M.

Learning procedure

When no observations are available, we may consider that a non-characterized class would have the equiprobability of all values in $[0,1]$. That happens when $\rho = 0.5$ and, therefore, $\hat{\rho}^{o} = 0.5$ will be the starting value for any class. The partition is, therefore, a set of probabilities of membership for each class.

$$\left\{p\left[x|x \in C_i\right]\right\}_{i=1}^{N} = \left\{p\left[x|\rho_i\right]\right\}_{i=1}^{N} \qquad N = \text{number of classes}$$

or equivalently a set of parameters $\left\{\rho_i\right\}_{i=1}^{N}$

The learning process consists in the estimation of this set of parameters. If a teacher is available, the algorithm updates the estimates of each $\hat{\rho}_j^{M}$ each time an observation is forced to belong to class C_j. In self learning the decision rule is applied to select C_j and, therefore, to enable the system to create a new class when needed, the partition must include an "indifferent class" C_o associated with parameter $\rho_o = 0.5$. Whenever this class is selected by the decision rule, it becomes a "non-indifferent" class whose estimated parameter is

$$\hat{\rho}_{N+1} = \rho_o + 1\left(y_t - \rho_o\right) = y_t$$

Discrete space of membership

We shall consider here the situation when $\mu_A(\omega)$ takes a finite number V of equidistant values $\left\{v_i\right\}_{i=0}^{V}$ in the interval $\left[0,1\right]$. $\mu_A(\omega)$ is a measurable function $\Omega \rightarrow \left\{v_i\right\}_{i=1}^{V}$. Let us define a set of V variables $x_i(\alpha)$ in $\left\{0,1\right\}$ such that $x_i(\alpha)=1 \Rightarrow x_{i-\ell}(\alpha)=1$

$$\forall \ell>0$$

and where α is a positive integer such that

$$x_\alpha(\alpha)=0 \text{ and } x_{\alpha-1}(\alpha)=1$$

We have then $v_\alpha = \frac{1}{V} \sum_{i=0}^{V} x_i(\alpha) = \frac{\alpha}{V}$

Let us consider the event $<<\alpha>>$ as the set $\left\{x_i(\alpha)\right\}_{i=0}^{V}$. It can be associated to the following sequence of subsets such that

$$\Omega \supset A_o \supset A_1 \ldots \supset A_\alpha \supset \ldots \supset A_V$$

or to the sequence of characteristic functions $\left\{\chi_i\right\}_{i=0}^{V}$. The set of values taken by the last sequence is $\left\{\left\{x_i(\alpha)\right\}_{i=0}^{V}\right\}_{\alpha=0}^{V}$

Then the event $<<\alpha>>$ can be represented by any element $\omega \in \Omega$ such that $\omega \in A_i \ \ \forall_i \geq \alpha$. If the probability measure of $A_i = \rho_i$ the probability of $<<\alpha>>$ is

$$\text{pr}\left[<<\alpha>>\right] = \lambda \prod_{i=0}^{\alpha-1} \rho_i \cdot \prod_{i=\alpha}^{V} \left(1-\rho_i\right)$$

where λ is the normalizing constant such that

$$\sum_{\alpha=0}^{V} p\left[<<\alpha>>\right] = 1$$

As we know that there is at least one ω_{in} such that $\mu_A\left(\omega_{in}\right) = 1$ and as we defined $\rho = P(A)$, we may define the event represented by ω_{in}

and $$pr\left[<<\omega_{in}>>\right] = P(A) = \rho = \lambda \prod_{i=0}^{V} \rho_i \leq 1$$

Similarly for $<<\omega_{ex}>>$ we have $(1-\rho) = \lambda \prod_{i=0}^{V} \left(1-\rho_i\right)$

In the particular case where the ρ_i's are all equal, we have

$$\rho_i = r = \frac{1}{\lambda} \rho^{\frac{1}{V}} \text{ and } \left(1-r\right) = \frac{1}{\lambda} \left(1-\rho\right)^{\frac{1}{V}}$$

then the probability of event $<<\alpha>>$ becomes

$$pr\left[<<\alpha>>\right] = r^{\alpha}\left(1-r\right)^{V-\alpha} = \lambda\rho^{\frac{\alpha}{V}}\left(1-\rho\right)^{\frac{V-\alpha}{V}}$$

and as $\frac{\alpha}{V} = v_\alpha$

$$pr \left[<<\alpha>>\right] = \lambda \rho^{v_\alpha}\left(1-\rho\right)^{(1-v_\alpha)}$$

and λ is calculated by

$$\frac{1}{\lambda} = \sum_{\alpha=0}^{V} \rho^{V\alpha}(1-\rho)^{(1-V_\alpha)}$$

Making $V \to \infty$ we may give a justification of the form of the probability density of the continuous case.

The condition that all ρ_i are equal, i.e., that in a monotonous sequence of sets with respect to inclusion, each of them has the same measure, may seem abnormal. Nevertheless, that gives an equal probability to all intermediate grades of membership and, depending on the value of ρ, enhances the cumulative effect of inclusion.

<u>Application to image classification</u>

Let an image be a two-dimensional array of points $x_{m,n}$

$m = 1,2...M$

$n = 1,2...N$

A point $x_{m,n}$ can be black ($x_{m,n} = 0$) or white ($x_{m,n} = 1$). This image will be called the <u>object</u>. This object is perceived as an array of points $y_{m,n}$ whose normalized light intensity is $y_{m,n} \in [0,1]$. This value is taken as a measure of the membership function of this point to the set of white points. The source object being x and the <u>observed image</u> y, we want to calculate the probability of getting the image y from the object x.

If for each point this probability is

$$p\left[y_{m,n} | x_{m,n} \in X\right] = \rho_{m,n}^{\;y_{m,n}} \left(1-\rho_{m,n}\right)^{\left(1-y_{m,n}\right)} K\left(\rho_{m,n}\right)$$

and if all points are considered independent, we have

$$p\left[y | x\right] = \prod_{m,n=1}^{M,N} \rho_{m,n}^{\;y_{m,n}} \left(1-\rho_{m,n}\right)^{\left(1-y_{m,n}\right)} K\left(\rho_{m,n}\right)$$

So we can see that to an object $X = \left\{x_{m,n}\right\}$, we may associate an image $R = \left\{\rho_{m,n}\right\}$. If we do not know from which object $x^1, x^2...x^T$ the image y comes, we shall calculate $p\left[y|x^i\right]$ and choose x^j such that

$$p\left[y|x^j\right] = \max p\left[y|x^i\right].$$

Updating of parameters R^i shall be made using recursive averaging as shown.

Figure 5 gives the computer program for the classifier including three options:

- learning with teacher
- self learning
- recognition without learning

Figure 6 gives an example of results given by this program for one-dimension images. The vertical position of X's gives the light intensity of the point. Images to be classified are on the left of the star line. Average image for the class to which it has been classified is on the right. That gives the parameters that represent this class.

Acknowledgment

Research sponsored by the National Science Foundation Grant INT75-04371-A01, U.S.-France ERL/LAAS Cooperative Program.

The author is on leave from: Laboratoire d'Automatique et d'Analyse des Systemes
C.N.R.S. Toulouse 31055 France

REFERENCES

1. Lopez de Mantaras, R., "Autoapprentissage d'une Partition: Application au Classement Itératif de Données Multidimensionnelles," These de Doctorat de Specialite, Universite Paul Sabatier, Toulouse, 1977.

2. Lopez de Mantaras, R. and Aguilar-Martin, J., "Classification par Autoapprentissage á l'aide de Filtres Numeriques Adaptatifs," Colloque AFCET-IRIA, Feb. 1978.

3. Zadeh, L. A., "Probability Measures of Fuzzy Events," Journal of Mathematical Analysis and Applications, Vol. 23, pp. 421-427, 1968.

4. Diday, E. and Govaert, G., "Apprentissage et Mesures de Ressemblance Adaptatives," Computer Oriented Learning Processes, editor J. C. Simon. Nordhoff-Leyden, 1976.

5. Csibi, S., Stochastic Processes with Learning Properties (book).

6. Image Object and Illusion. In Readings from Scientific American, 1976.

7. Lopez de Mantaras, R., "On Self Learning Pattern Classification," Memorandum, U. C. Berkeley (to appear).

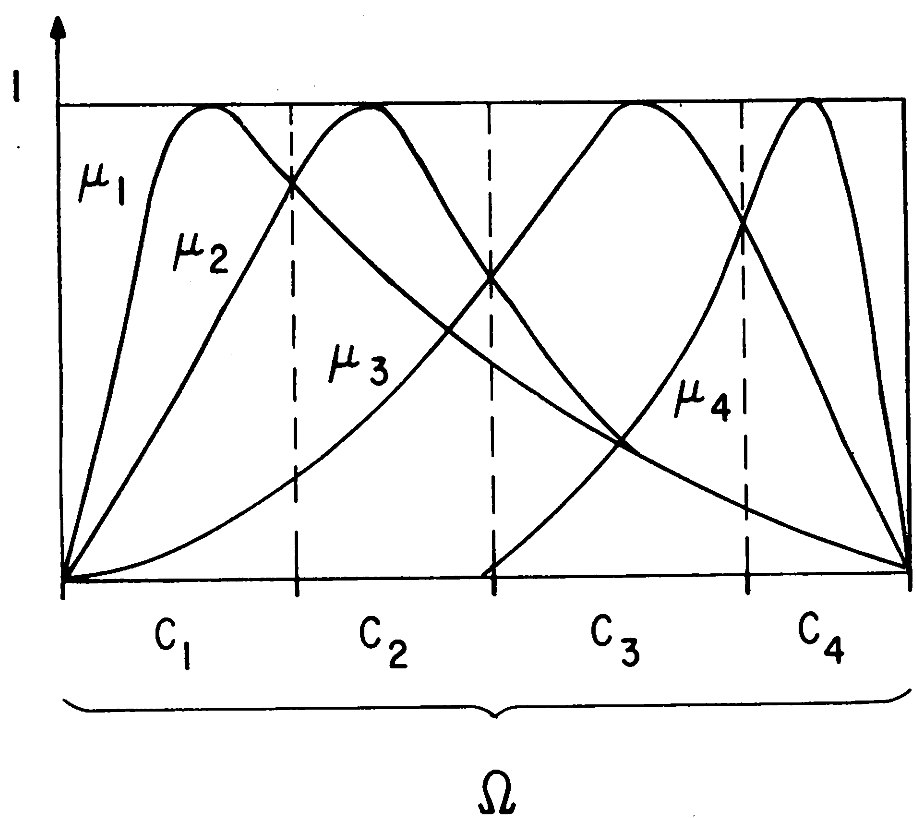


Figure 1

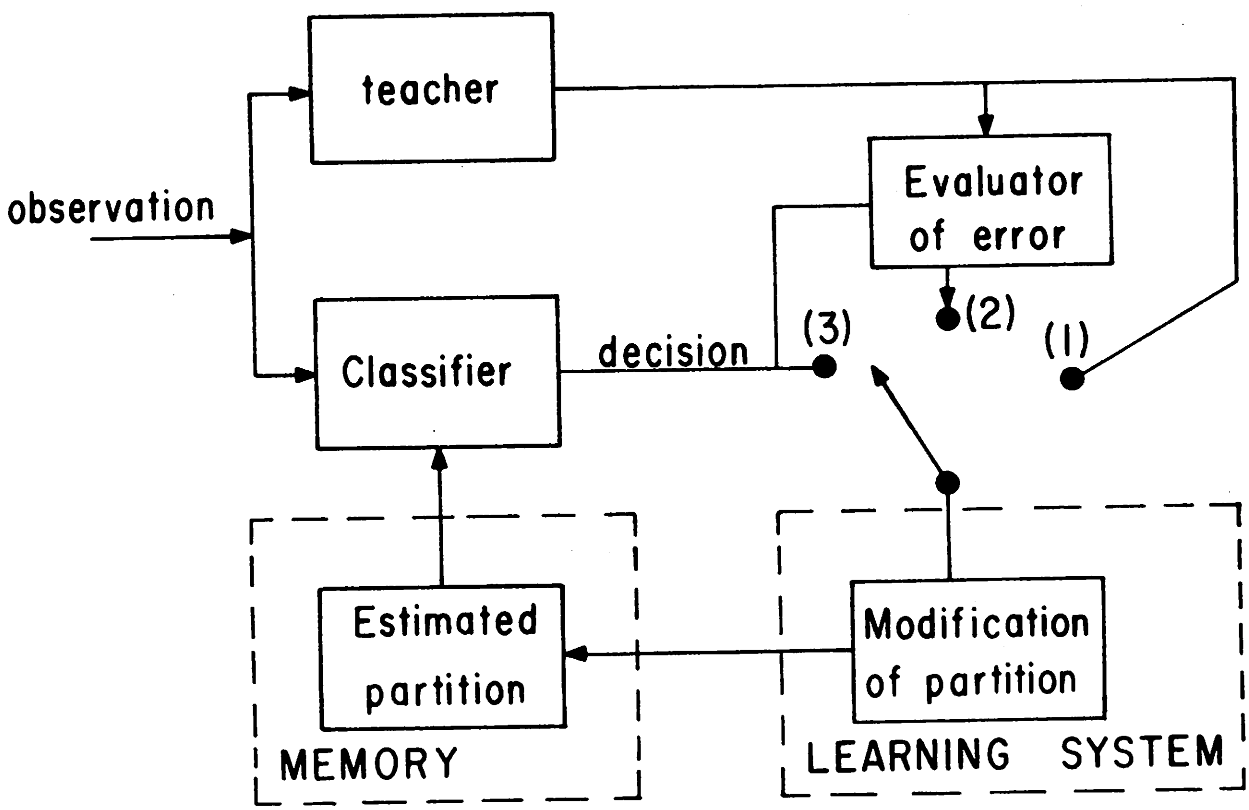

teacher
observation
Evaluator
of error
(2)
(3)
(1)
Classifier
decision
Estimated
partition
MEMORY
Modification
of partition
LEARNING SYSTEM

Figure 2

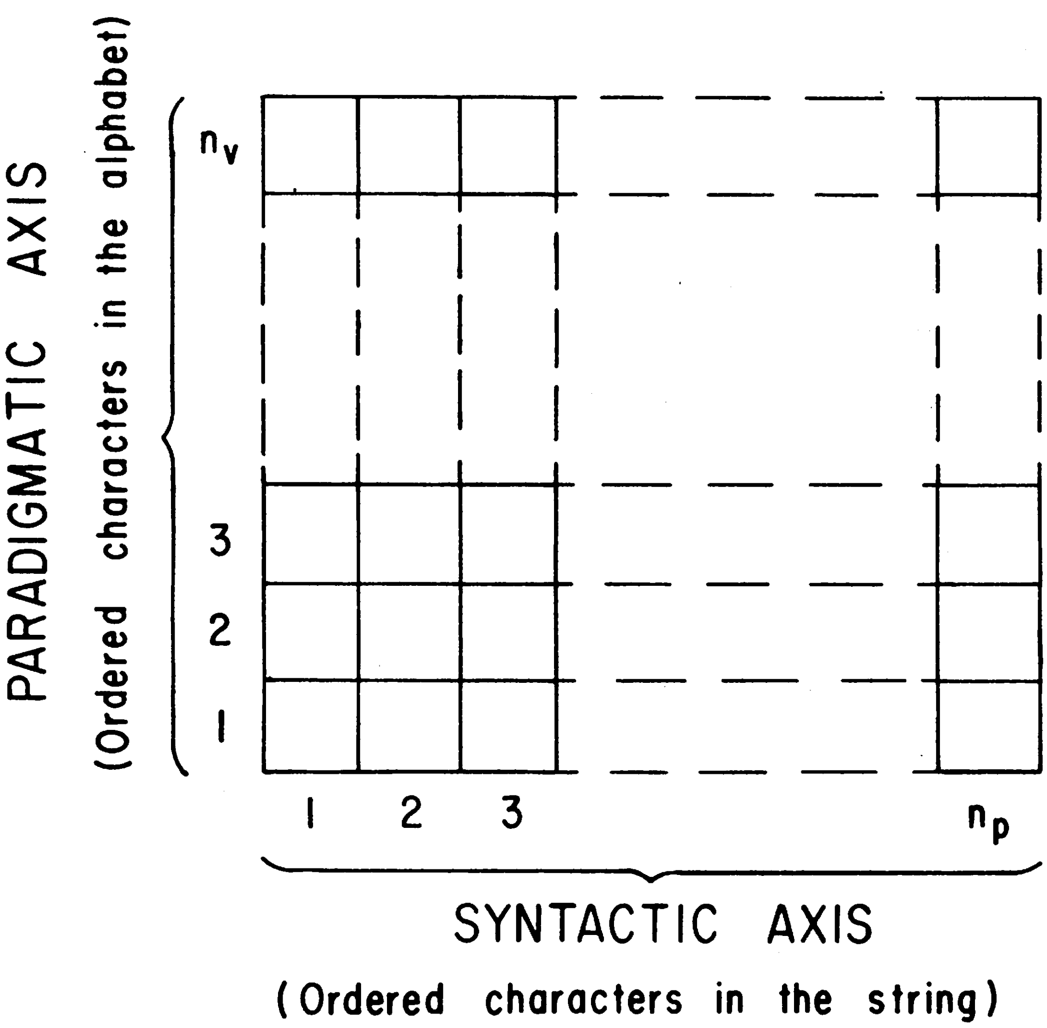

PARADIGMATIC AXIS
(Ordered characters in the alphabet)
n_v
3
2
1
1
2
3
n_p
SYNTACTIC AXIS
(Ordered characters in the string)

Figure 3

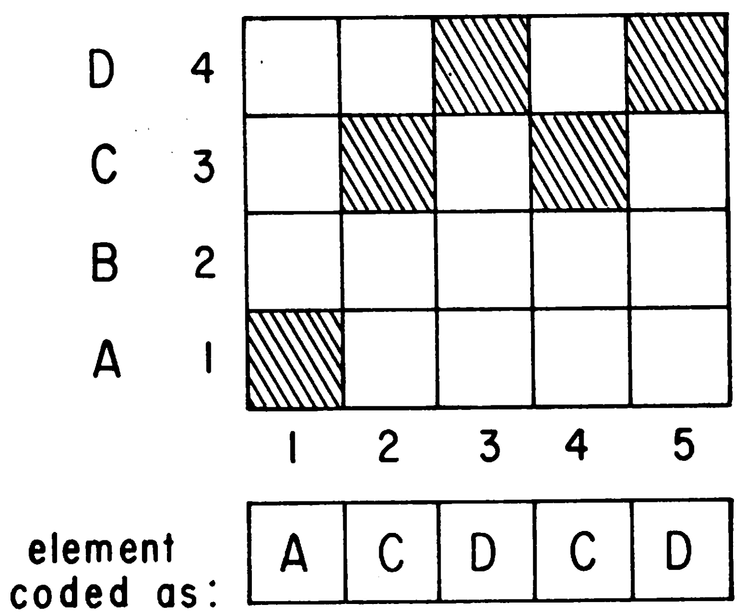


Figure 4

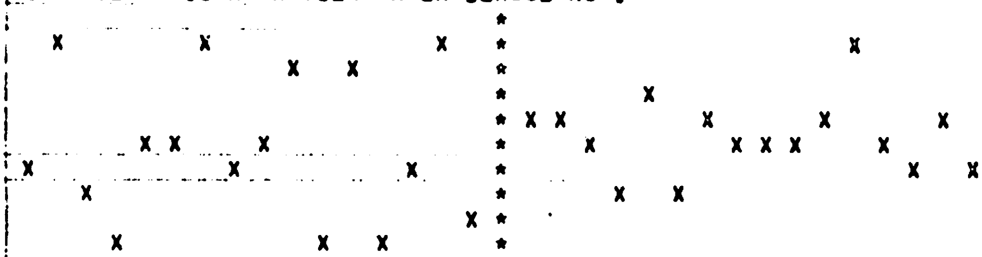
```
C
C    SOUS-PROGRAMME DE CLASSIFICATION DE DONNEES CONTINUES ENTRE 0 ET 1
C
C

      SUBROUTINE CLRO(X,NP,KV,IV,MMAX,RO,M,NK,KC,NO,MU)
      REAL LAMBDA,MUMAX
      REAL MU(MMAX)
      DIMENSION X(NP),RO(NP,MMAX),NK(MMAX)
      MUMAX=0
C
C     CAS PARTICULIER:LA PREMIERE IMAGE SERT A INITIALISER C1
      IF (M.EQ.0) GO TO 3
C
C      CALCUL DE LA PROXIMITE MU DE X A LA CLASSE CK
C
      DO 1 K=1,M
      MU(K)=1
      DO 2 I=1,NP
      LAMBDA=ALOG(RO(I,K)/(1-RO(I,K)))
      P=LAMBDA*((1-RO(I,K))/(2*RO(I,K)-1))*EXP(LAMBDA*X(I))
2     MU(K)=MU(K)*P
C
C     DETERMINATION DE LA CLASSE CKC DONT L'IMAGE X EST LA PLUS PROCHE

      IF(MU(K).GT.MUMAX) KC=K;MUMAX=MU(KC)
1     CONTINUE
      IF(IV-1) 4,5,6
C
C     IV=0 CORRESPOND A L'APPRENTISSAGE AVEC PROFESSEUR
C     IV=1 CORRESPOND A L'AUTO-APPRENTISSAGE
C     IV>1 SUPPOSE L'APPRENTISSAGE FINI:ON CLASSE
C
C
C     X EST PLACE D'AUTORITE DANS LA CLASSE KV
C     NO EST LA NOTE ATTRIBUEE AU SOUS-PROGRAMME
4     NO=NO+1
      IF(KC.NE.KV) NO=NO-2;KC=KV
      GO TO 7
C
C     COMPARAISON DE LA PROXIMITE MU(KC) AVEC UN SEUIL
5     IF((MUMAX.LT.1).AND.((M+1).LE.MMAX)) GO TO 3
C
C     MODIFICATION DU REPRESENTANT DE CKC
7     NK(KC)=NK(KC)+1
      DO 11 I=1,NP
      DELTA=(X(I)-RO(I,KC))/(NK(KC)+1)
11    RO(I,KC)=RO(I,KC)+DELTA
      RETURN
6     NK(KC)=NK(KC)+1
      RETURN
C
C     ON CREE UNE NOUVELLE CLASSE
3     M=M+1;KC=M;NK(KC)=0
      DO 8 I=1,NP
8     RO(I,KC)=0.5
      GOTO 7
      END
```

Figure 5

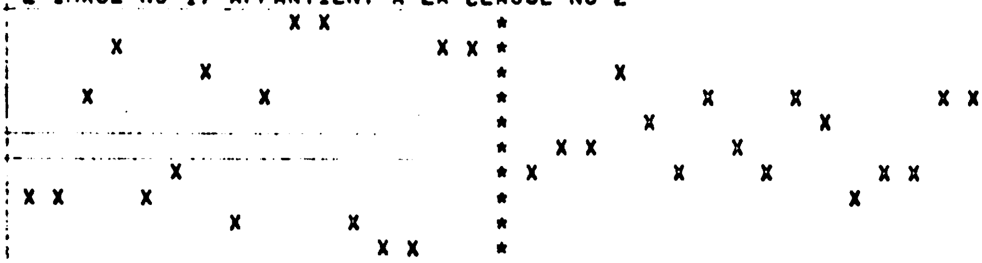
LA REPARTITION DES IMAGES EST LA SUIVANTE APRES 3 PROCEDURES DE CLASSIFICATION

L'IMAGE NO 1 APPARTIENT A LA CLASSE NO 1

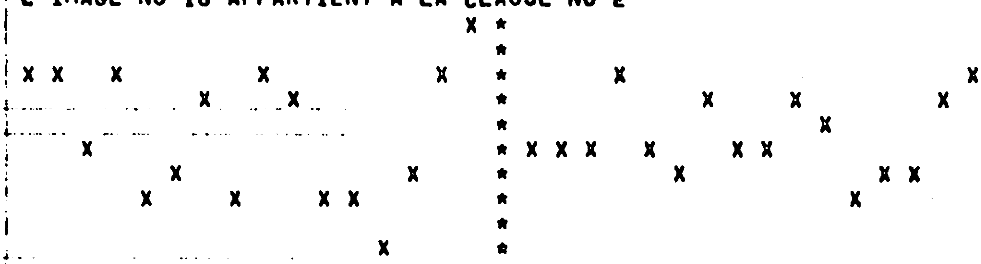
L'IMAGE NO 16 APPARTIENT A LA CLASSE NO 1



L'IMAGE NO 17 APPARTIENT A LA CLASSE NO 2



L'IMAGE NO 18 APPARTIENT A LA CLASSE NO 2



L'IMAGE NO 19 APPARTIENT A LA CLASSE NO 4

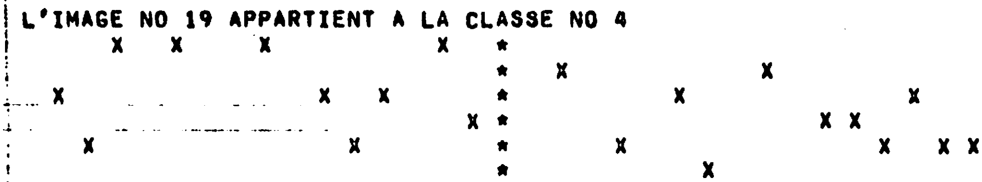


Figure 6