

# OPTIMAL LEARNING ALGORITHMS FOR AUTOMATIC SYSTEMS IN THE CASE OF A NONIDEAL TEACHER

Corresponding Member of the Academy of Sciences of the USSR V.  
S. PUGACHEV

1967

SovietRxiv

---

View the original and related papers at <https://sovietrxiv.org/items/ru-196701.04288>

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.

**Abstract**

**Full Text**

UDC 621.398+621.52[016.3]

## CYBERNETICS AND CONTROL THEORY

Corresponding Member of the Academy of Sciences of the USSR V. S. PUGACHEV

# OPTIMAL LEARNING ALGORITHMS FOR AUTOMATIC SYSTEMS IN THE CASE OF A NONIDEAL TEACHER

In the literature on the theory of learning systems, usually only the cases of learning with an ideal teacher and of self-learning are considered. Here, by an ideal teacher is meant a teacher (a human operator or an automaton) who inputs into the system absolutely exact values of the required output signal, corresponding to the given values of the input signal. In reality this is possible only in an artificially organized setting, for example, when training an automatic system under laboratory conditions with the use of modeling devices and computing equipment. When learning under natural conditions, and sometimes also under laboratory conditions (for example, when teaching a system a control process by means of a human operator), the input into the system of absolutely exact values of the required output signal often proves impossible. Instead of exact values of the required output signal, only the corresponding values of the teacher's output signal are input into the system; these are merely statistical estimates of the values of the required output signal, produced by the teacher from the results of processing the accepted values of the input signal. The same situation also occurs in the case of teaching a system by a teacher who does not show the system the desired values of the output signal, but only observes the operation of the system itself and gives it an evaluation (in the simplest case, "reward" or "punishment"). In this case the teacher communicates to the system only his own, more or less subjective, evaluation of the system's actions.

Let  $Z$  be the input signal;  $W$ ,  $\hat{W}$ , and  $\tilde{W}$  respectively the required output signal, the output signal of the system being trained, and the teacher's output signal. In the general case, the teacher's operating algorithm is determined by the conditional probability distribution of his output signal  $\tilde{W}$  for given values of  $Z$ ,  $W$ , and  $\hat{W}$ . Regarding  $W$ ,  $\hat{W}$ , and  $\tilde{W}$  as finite-dimensional random vectors, we may define the teacher's operating algorithm by the conditional probability density  $\delta_y(\tilde{w} | z, w, \hat{w})$  of the vector  $\tilde{W}$  for any given values  $z$ ,  $w$ ,  $\hat{w}$  of the quantities  $Z$ ,  $W$ , and  $\hat{W}$ . The function  $\delta_y(\tilde{w} | z, w, \hat{w})$  will henceforth be called the teacher's **deciding function**.

In the particular case of an **ideal** teacher, who inputs into the system being trained absolutely exact values of  $W$ ,  $\delta_y(\tilde{w} | z, w, \hat{w}) = \delta(\tilde{w} - w)$  and does not depend on  $z$  and  $\hat{w}$ , where  $\delta(x)$  is the usual impulse Dirac  $\delta$ -function (in the general case, multidimensional).

In contrast to the ideal teacher, we shall call any teacher whose output signal  $\tilde{W}$  does not coincide with the required output signal  $W$  a **real** one. For any real teacher, the value  $w$  of the vector  $W$  is always unknown, and he can produce the value  $\tilde{w}$  of his output signal  $\tilde{W}$  only from the observed value  $z$  of the input signal  $Z$  (learning by demonstration), or from the observed values  $z$ ,  $\hat{w}$  of the input and output signals  $Z$ ,  $\hat{W}$  of the system being trained (learning by evaluating the system's actions).

Thus, the conditional probability density  $\delta_y(\tilde{w} | z, w, \hat{w}) = \delta_y(\tilde{w} | z)$  does not depend on  $w$  and  $\hat{w}$  in the case when the real teacher trains the system by demonstration, and  $\delta_y(\tilde{w} | z, w, \hat{w}) = \delta_y(\tilde{w} | z, w)$  does not depend on  $\hat{w}$  in the case when the real teacher trains the system by evaluating its actions.

The required operating algorithm of the system being trained, which must be developed as a result of training, is in the general case also characterized by a decision function, i.e., by the conditional probability density  $\delta(\hat{w} | z)$  of the output signal of the system  $\hat{W}$  for a given value  $z$  of the input signal  $Z$ . In the particular case where there exists a deterministic Bayes-optimal system corresponding to the given criterion, and it is required that the system being trained after training be as close as possible to this optimal system,  $\delta(\hat{w} | z) = \delta(\hat{w} - w^*)$ , where  $w^*$  is the value of the output signal  $W^*$  of the Bayes-optimal system corresponding to the value  $z$  of the input signal  $Z$ .

Let us first consider the case of a discrete system. In this case the input signal  $Z$  is a finite-dimensional random vector. Let  $h(z, w)$  be the joint probability density of the random vectors  $Z, W$ , which in the general case may contain a linear combination of  $\delta$ -functions with singularities at points  $(z, w)$  to which nonzero probabilities correspond.

In essence, in the problem under consideration the functions  $\delta_y$ ,  $\delta$ , and  $h$  are unknown. If the teacher's decision function  $\delta_y$  were completely known, then one could create a system with  $\delta = \delta_y$ , which without any training would operate with the same quality as the teacher. Given a decision function  $\delta$ , one can create a system with this decision function that does not require training. Finally, given the probability density  $h$ , one can create a system close to the Bayes-optimal system, also not requiring training. For practical purposes it is sufficient to assume that the functions  $h$ ,  $\delta$ , and  $\delta_y$  are known functions of their arguments depending on a finite number of unknown parameters forming a vector  $\lambda$ :

$$\begin{aligned} h(z, w) &= H(z, w | \lambda), & \delta(\hat{w} | z) &= \Delta(\hat{w} | z, \lambda), \\ \delta_y(\tilde{w} | z, w, \hat{w}) &= \Delta_y(\tilde{w} | z, w, \hat{w}, \lambda). \end{aligned} \tag{1}$$

Put

$$G(z | \lambda) = \int H(z, w | \lambda) dw, \quad (2)$$

$$R(z, \hat{w}, \tilde{w} | \lambda) = \int H(z, w | \lambda) \Delta_y(\tilde{w} | z, w, \hat{w}, \lambda) dw, \quad (3)$$

where the integration extends over the domain of all possible values  $w$  of the random vector  $W$ .

Suppose that in the course of training the values  $z_1, \dots, z_N$  of the input signal  $Z$  and the corresponding values  $\tilde{w}_1, \dots, \tilde{w}_N$  of the teacher's output signal  $\tilde{W}$  have been introduced into the system. Suppose that the system responded to the introduced values of the input signal respectively with the values  $\hat{w}_1, \dots, \hat{w}_N$  of the output signal  $\hat{W}$ . Then, regarding all possible values of the vector  $\lambda$  as values that can be assumed by a certain random vector  $\Lambda$  with prior probability density  $\alpha(\lambda)$ , and applying the known Bayes formula, we find the posterior probability density of the vector  $\Lambda$ :

$$\omega(\lambda) = c\alpha(\lambda) \prod_{i=1}^N R^i(z_i, \hat{w}_i, \tilde{w}_i | \lambda), \quad (4)$$

where  $c$  is a normalizing factor. We have marked the function  $R$  with the superscript  $i$ , assuming that the functions  $h$  and  $\delta_y$  may change slowly with time, as a result of which the functions  $H, \Delta_y, G, R$  may be different in different cycles of system operation.

In the case of a continuous system the input signal  $Z$  is a random function of time. In this case the problem of determining  $\omega(\lambda)$  is easily solved if  $Z$  is the sum of a useful signal, depending on a finite-dimensional random vector  $U$ , and normally distributed noise independent of  $U$ . Representing the noise approximately by a finite segment of the canonical expansion, we replace the random function  $Z$  by a finite-dimensional random vector of coefficients. After this, having determined the posterior probability density of the vector  $\Lambda$  and passing to the limit, we again obtain formula (4), but the function  $h(z, w) = H(z, w | \lambda)$  in this case will not be the probability density of  $Z, W$  <sup>(1)</sup>.

In the special case of an ideal teacher  $R(z, \hat{w}, \tilde{w} | \lambda) = H(z, \tilde{w} | \lambda)$ , and

$$\omega(\lambda) = c\alpha(\lambda) \prod_{i=1}^N H^i(z_i, \tilde{w}_i | \lambda). \quad (5)$$

In the case of a real teacher teaching by demonstration,  $R(z, \hat{w}, \tilde{w} | \lambda) = G(z | \lambda) \Delta_y(\tilde{w} | z, \lambda)$ , and

$$\omega(\lambda) = ca(\lambda) \prod_{i=1}^N G^i(z_i|\lambda) \Delta_y^i(\tilde{w}_i|z_i, \lambda). \quad (6)$$

From comparison of (5) and (6) it is clear that a real teacher whose decision function  $\delta_y$  coincides with the conditional probability density of the required output signal  $W$  for a given value  $z$  of the input signal  $Z$  is equivalent to an ideal teacher, since for such a teacher  $\Delta_y(\tilde{w}|z, \lambda) = H(z, \tilde{w}|\lambda)/G(z|\lambda)$ .

A real teacher for which the conditional variance  $\tilde{W}$  for a given value  $z$  of the input signal  $Z$  is less than the conditional variance of  $W$  will, as a rule, be better than an ideal teacher in the sense that for him the posterior variances of the parameters  $\Lambda$  will be smaller than for an ideal teacher. In particular, a teacher that is a Bayes-optimal system (in the case when it is deterministic) is practically always better than an ideal teacher, since for him  $\Delta_y(\tilde{w}|z, \lambda) = \delta(\tilde{w} - A(\lambda)z)$ , where  $A(\lambda)$  is the operator of the Bayes-optimal system depending on  $\lambda$ , and, consequently, the distribution of the vector  $\Lambda$  is entirely concentrated on the subset of values  $\lambda$  determined by the equations

$$\tilde{w}_i = A(\lambda)z_i \quad (i = 1, \dots, N). \quad (7)$$

If there exist  $r$  such equations having a unique solution with respect to  $\lambda$ , then for any  $N \geq r$  the distribution of the vector  $\Lambda$  is concentrated at a single point corresponding to the unknown true values of the parameters  $\lambda$ . In this case the system will be completely trained after input into it of  $r$  pairs of training realizations  $z_i, \tilde{w}_i$ .

In the case of a real teacher teaching by evaluating the actions of the system,

$$R(z, w, \tilde{w}|\lambda) = G(z|\lambda) \Delta_y(\tilde{w}|z, w, \lambda),$$

and

$$\omega(\lambda) = ca(\lambda) \prod_{i=1}^N G^i(z_i|\lambda) \Delta_y^i(\tilde{w}_i|z_i, \hat{w}_i, \lambda). \quad (8)$$

Let us note, finally, that in the case of self-training the formula for the posterior probability density of the vector  $\Lambda$  has the form

$$\omega(\lambda) = ca(\lambda) \prod_{i=1}^N G^i(z_i|\lambda). \quad (9)$$

This formula may be regarded as a special case of formula (6) or formula (8), when the variance of the teacher's output signal is infinite or when the teacher generates his output signal independently of

of the vector  $\lambda$ . Thus, self-learning is equivalent to learning with a very poor teacher, who gives the system no useful information.

After  $\omega(\lambda)$  has been found, the optimal Bayesian estimate of the decision function  $\delta$  by the criterion of the minimum mean square error is determined by formula (1)

$$\delta^*(\hat{w} | z) = \int \Delta(\hat{w} | z, \lambda) \omega(\lambda) d\lambda. \quad (10)$$

In an analogous way one can find Bayesian optimal estimates corresponding to other criteria.

Formulas (4) and (10) define a general Bayesian optimal learning algorithm for an automatic system, which gives, as particular cases, the corresponding optimal algorithms for various types of learning. The general algorithm defined by formulas (4) and (10) does not depend on how the information introduced into the system in the course of learning is used: all at once after the completion of learning, or in parts, each time using the posterior distribution of the vector  $\Lambda$ , obtained from all preceding training realizations, as the prior distribution for a new portion of training realizations.

The algorithm defined by formulas (4) and (10) fully uses the information about the unknown characteristics contained in the training realizations of the signals, but makes absolutely no use of the information contained in the realizations of the input signal obtained by the system after learning. To construct an optimal learning system that fully uses the information about the unknown characteristics obtained both in the course of learning and after the completion of learning, it is necessary to apply the Bayesian approach consistently. As a result, one obtains a Bayesian optimal learning system which, after the completion of learning, switches to self-learning during operation.

As a special case, formulas (4) and (10) imply the Bayesian optimal algorithms for learning and self-learning of pattern-recognition systems, given in (2).

Received  
19 II 1966

## CITED LITERATURE

1. V. S. Pugachev, *Theory of Random Functions and Its Application to Problems of Automatic Control*, Moscow, 1962.
2. V. S. Pugachev, "Statistical Problems in the Theory of Pattern Recognition," report at the Third All-Union Conference on Automatic Control and Technical Cybernetics, Odessa, 20-26 IX 1965.

*Note: Figure translations are in progress. See original paper for figures.*

*Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.*