Journal of Intelligent Learning Systems and Applications, 2026, 18(2), 123-131
https://www.scirp.org/journal/jilsa
ISSN Online: 2150-8410
ISSN Print: 2150-8402

Reducing Data Chaos and Partitioning the Training Sample into Macro-Features in Classification Problem

Vladimir Shats

St. Petersburg, Russia
Email: vlshats@hotmail.com

How to cite this paper: Shats, V. (2026) Reducing Data Chaos and Partitioning the Training Sample into Macro-Features in Classification Problem. *Journal of Intelligent Learning Systems and Applications*, **18**, 123-131.
https://doi.org/10.4236/jilsa.2026.182008

Received: March 27, 2026
Accepted: May 23, 2026
Published: May 26, 2026

Abstract

This paper is devoted to revealing some features of machine learning problems and developing a new approach to solving them. It is based on the application of information processing technology by animal sensory systems, each of which perceives information of only a certain type. Therefore, computational operations for solving the classification problem are performed mainly for individual features of objects, although they are usually carried out for objects as a whole. This approach ensures the simplicity of the algorithm and the ability to order the features by sorting in non-decreasing order their values, which leads to a decrease in the entropy and chaos of the data. It has been established that ordered features are hidden variables that allow us to detect the functional relationship "feature-class" and to partition any training sample into macro-features, which are ordered features of objects of a certain class. Classification of any object in test sample comes down to calculating the frequency of occurrence of its feature values in the nearest neighborhood of ordered feature values of the corresponding macro-feature of a certain class. The object class corresponds to the maximum of the average value of this frequency. Applying of ordered features opens up the prospect of new types of neural networks. The article also discusses the application of an ordered data matrix to solve problems of partitioning a set into clusters of objects with common properties.

Keywords

Reducing Data Chaos, Hidden Variables, Functional Relationship "Feature-Class"

1. Introduction

The paper proposes a new approach to solving machine learning problems in

which it is necessary to divide objects into non-overlapping classes or to determine the composition of a subset of objects that meet given requirements [1]. These problems are characterized by a high level of uncertainty, since the feature values are measured with random errors, and the set of features cannot take into account all the object features. At the same time, the concept of class is not defined precisely enough. Such problems are considered difficult to formalize, in which the dependencies between data elements are difficult to express in formulas or words.

Their solution requires a cognitive approach based on the analysis of experience in solving similar problems in nature, which has shown the commonality of information processing mechanisms in humans and animals [2]. Then, the set of task data is interpreted as a model of a complex system formed by a set of interconnected elements in which general patterns of information processing exist and the abovementioned uncertainty factors operate. Research carried out according to this approach led to the development of a method for reducing data entropy, the application of which ensured formalization of the classification problem.

The obtained results are found on a new concept of similarity, according to which for objects of the same class, it is assessed not by the distance between objects in metric space, but by the proximity of individual feature values. In [3], it is shown that a hierarchically organized system, the elements of which are features, objects and classes, allows us to describe the existing mechanism of information processing by individual sensory systems of an animal [4]. Its receptors perceive information from the external and internal environment, which, after processing, is transmitted to the brain, where it is compared with similar information already accumulated there.

This scheme of the information processing process was implemented in the classification problem [5] as follows: the set of each feature value was divided into an equal number of intervals, within which it was considered possible to neglect the difference of the feature values. For each interval number, lists of training sample (TS) objects of the same class were determined, forming the corresponding subsets called granules, as well as the frequency of granules. Given the concept of proximity for granule objects, the classes of test sample objects were calculated by combining the obtained results based on the simplest formula for total probability.

According to the given algorithm, the main volume of calculations falls on operations with the individual feature values of objects or a set of objects that form granules. Its simplicity makes it qualitatively different from existing algorithms [6], where, as a rule, objects are considered as a multidimensional set of all their features. It is obvious that the simplification of the algorithm was caused by the indicated change in the data structure.

A number of questions about the properties of distributions of ordered features were considered in [7]. It turned out that these distributions differ significantly for each feature and class. In another series of studies, the influence of a class on the frequency of occurrence of one, two, and so on objects of any other class among

its objects was examined for ordered features. Calculations for several databases showed a stable nature of the joint influence of class and feature on these frequencies.

In analyzing the obtained results, an approach that has been repeatedly used in physics was used: "guessing the patterns" of processes by comparing known research data [8]. The main focus was on the theory of patterns [9]. The consecutive numbers of intervals were considered as variables, with the help of which it was possible to identify relationships and optimize the structure of the TS data. Since the corresponding the TS feature values were assumed to be equal on each interval, they played the role of closest neighbors in terms of the feature value. In fact, it was assumed that the distribution of any feature values has the form of a step curve, increasing at each interval.

Further analysis showed that it is advisable to use the method of ordering feature values by sorting in non-decreasing order, in which their feature values also form a sequence of nearest neighbors. (The concept of ordering will be clarified below). Then, the number of data placement options will be minimized, which will lead to a decrease in data entropy and, accordingly, the level of uncertainty and chaos [10]. This ordering serves as a way to optimize the data structure and opens up new possibilities for revealing patterns of the system. It is no coincidence that for over 50 years now, issues of chaos have attracted increased attention and solutions have been obtained to problems that have important theoretical and practical significance in physics, mechanics, chemistry, medicine, ecology, telecommunications, control of mechanical and electronic systems, technological processes, etc. [11].

It was shown that for any TS, ordered features are hidden variables that reveal the relationships between the observed, given values of objects' features of the same class. From now on, the term "feature" will be used to refer only to unordered features. It turned out that the set of the TS data can be considered as a union of its subsets, on each of which a deterministic function of ordered values of objects' feature of a certain class is defined.

Note that there is no functional dependence between the objects' features of the same class, since the distribution of the values feature has many jumps of the same order as the range of the feature values. By ordering features, complex chaotic relationships between features of objects of the same class are transformed into deterministic functions.

2. Properties of Ordered Features

Let us consider the classification problem TS. Let $G = \|x_{sk}\|_{M\times N}$ be the matrix of quantitative data of the TS, $s = 1,\cdots,M$ are the numbers of objects, $X^k = (x_{1k},\cdots,x_{Mk})^{\mathrm{T}}$ is the feature vector $k = 1,\cdots,N$, $i = g(s)$ is the class of object s, $i = 1,\cdots,C$.

A set of elements of a vector X^k will be called ordered [12] if they have been renumbered and given new numbers $(s)^k = (1)^k,(2)^k,\cdots,(M)^k$ such that the

corresponding values of this vector form a non-decreasing sequence $x_{(1)^k} \le x_{(2)^k} \le \cdots \le x_{(M)^k}$. The process of ordering the feature k is reduced to the simplest sorting of the vector X^k. In the case where objects have equal feature values, the one-to-one correspondence between their numbers and ordered numbers may be violated. However, this circumstance will not affect subsequent results, since the numbers and ordered numbers of objects correspond to the same attribute value. Therefore, the mapping of features onto ordered features of the TS objects can be considered as one-to-one [13].

Structure of the ordered vector X^k has important features. It is obvious that if the feature values $x_{s2k} > x_{s1k}$ for objects $s1$ and $s2$, then the ordered numbers of these objects are $(s2)^k > (s1)^k$, and this relationship is preserved for objects of the same class. Let us denote by $\tilde{s}_i^k = 1, \cdots, l_i$ ordinal numbers objects of class i, arranged in non-decreasing order of feature values, where l_i is the length of class i. These numbers determine the values of x_{sk} on the same set $\{\tilde{s}_i^k\}$ for features or ordered features.

Let us illustrate the features of structuring using the example of the vector X^3, all of whose objects, except $s = 5$ and $s = 7$, have class 1:

$$X^3 = (0.23, 0.11, 0.73, 0.05, 0.42, 0.421, 0.065)^{\mathrm{T}}.$$

Here, the set $\{(s^3)\} = \{4^3, 7^3, 2^3, 1^3, 5^3, 6^3, 3^3\}$ is the union of the subsets $\{(s^3)\} = \{7^3, 5^3\}$ and $\{(s^3)\} = \{4^3, 2^3, 1^3, 6^3, 3^3\}$ for objects of class $i = 1$ and $i = 2$, respectively. In ordinal scales, these subsets have the form $\{\tilde{s}_1^3\} = \{1, 2\}$ and $\{\tilde{s}_2^3\} = \{1, 2, 3, 4, 5\}$. Then, the vectors $(0.065, 0.42)^{\mathrm{T}}$ and $(0.05, 0.11, 0.23, 0.421, 0.73)^{\mathrm{T}}$ will describe in these scales the objects' features of classes $i = 1$ and $i = 2$, respectively.

Note that in the case where two objects have the same value of the feature $x_{s2k} = x_{s1k}$, we will get an ambiguous relation $(s2)^k = (s1)^k \pm 1$ when sorting. But this circumstance will not affect subsequent results, since both object numbers correspond to the same attribute value. This conclusion extends to the case where several objects have equal feature values.

As shown above, the values of $x_{(s)^k}$ on the set $\{\tilde{s}_i^k\}$ form a non-decreasing sequence. This result means that there is some discrete monotone function that describes the values feature k for objects of class i:

$$f(\tilde{s}_i^k) = x_{(s)k}^i, \text{ where } \tilde{s}_i^k = 1, \cdots, l_i.$$

In an ordered feature space, these functions are visualized into clear "chains" of feature values for objects of the same class. On the plane, we obtain a graph of point values feature k for objects of class i. For example, let class i consist of three objects, the feature values k are $x_{sk} = 0.29, 0.08, 0.62$. According to these data, the ordered numbers $(s)^k = (2), (1), (3)$ and the values $\tilde{s}_i^k = 1, 2, 3$ correspond to the values $x_{(s)k}^i = 0.08, 0.29, 0.62$.

Let us consider the subset of the feature values k for the TS objects of class i

They are defined for all k and i on the subset

$$W_{ki} = \left\{ x_{(s)^k} \mid (s) = 1, \cdots, M \cap g\left((s)\right) = i \right\}.$$

which will call a "macro-feature". Any TS consists of $N * C$ macro-features that generalize the information contained in it and map it onto a set of functions

$$\left\{ f\left((\tilde{s})_i^k \right) \right\} = \bigcup_{k=1}^{N} \bigcup_{i=1}^{C} f\left((\tilde{s})_i^k \right).$$

It is obvious that macro-features play the role of patterns.

3. Classification of Ordered Data

The classes of objects of the test sample are determined based on the generally accepted assumption that the training and test samples belong to a single general population. To make the results of problem-solving clearer, we will assume that the values of each feature of the combined sample objects were previously normalized by bringing them to the interval [0, 1] using the formula $\frac{x_{sk} - x_{\min}}{x_{\max} - x_{\min}}$.

The class of an arbitrary object of the test sample t is calculated on the basis of the simplest formula of total probability by estimating the frequency γ_{ki} of its feature values z_{tk} falling into the nearest neighborhood of the ordered feature values for each class. In the paper, two variants of the value γ_{ki} were used, corresponding to the application of formulas for estimating the proximity condition $x_{\tilde{s}k} < z_{tk} \le x_{(\tilde{s}+1)k}$ and $\left| x_{\tilde{s}k} - z_{tk} \right| \le \delta$, where δ is the proximity parameter. These relations allow us to determine the class of objects in the test sample whose feature values are closest in value to z_{tk} or fall within δ is the neighborhood of the value to z_{tk}, respectively.

However, given that monotone functions $f\left(\tilde{s}_i^k\right)$ determine the magnitude of features, and not the estimates of their probability, it is advisable to implement option two also for the approximating function. Using the linear regression equation, we obtain the relation $f\left(\tilde{s}_i^k\right) \cong a + b * \tilde{s}_i^k$. Considering that b is equal to the tangent of the angle of inclination of the regression line to the horizontal axis, we can normalize the length of the normal segment to this line according to the relation $\left| z_{tk} - f\left(\tilde{s}_i^k\right) \right| \le \delta\sqrt{1 + b^2}$.

One of the options for determining the class of an arbitrary object t of a test sample is illustrated in **Figure 1**. It shows a graph of the values $x_{(s)k}^i$ for the TS with $\tilde{s}_i^k = 1, \cdots, 16$, linear regression line and two straight lines removed from this line at distance of $\delta = \pm 0.05$, and the straight line $z_{tk} = 0.18$, which corresponds to the feature value k for some object t of the test sample. It follows from the drawing that $r_{ki} = 1$, since the value of z_{tk} is inside the rectangle ABCD. Here, r_{ki} is a binary value equal to 1 if the value of z_{tk} corresponds to class i, and 0 otherwise.

The average frequency of cases across all features in which the value z_{tk} will correspond to class i equal to

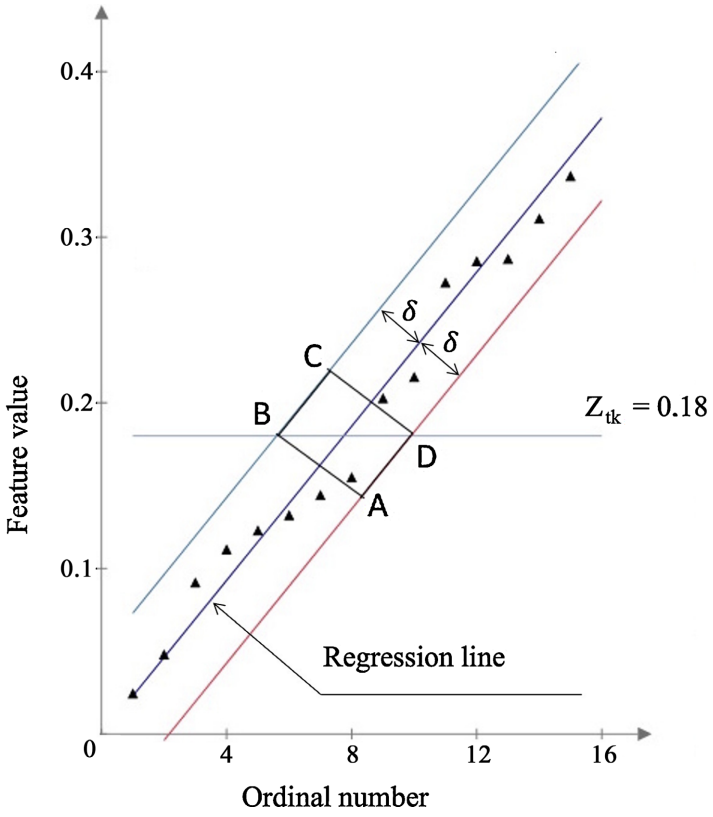

Figure 1. Scheme for assessing the proximity z_{ik} and class i.

$$\gamma(i)=\frac{1}{N}\sum_{k=1}^{N} r_{ki}.$$

Then, the class of object t equal to

$$I=\arg\max_{1\le i\le C}\gamma(i).$$

The results of applying these formulas were obtained for the well-known Iris and Wine databases [14]. For all variants of calculation formulas, the number of classification errors did not exceed 13%, but in the range $0.01\le\delta<0.2$ for the approximation variant, the solution was error-free.

4. Effect of Feature Ordering on the Data Matrix

In the previous sections, hidden patterns in the data were revealed concerning the relationship between the individual feature values of objects and their class. Now, let us consider the issues of the relationship between the entire set of feature values describing an object and its class.

Obviously, the ordering of any feature vector disrupts the composition of the elements of the matrix rows that describe the objects. Therefore, the ordering of the entire set of the TS data is carried out for each of the features separately, and the data matrix G is mapped onto a set of N matrices of ordered data $G_k=\left\|x_{\left(s^k\right)}\right\|_{M*N}$. All these matrices and the matrix G have one identical row

each and the elements of column k of the matrix It is obvious that all these matrices and the matrix G have one identical row, and the elements of column k of the matrix G_k are ordered and arranged in non-decreasing order of the values feature k.

The resulting changes are illustrated by the example:

$$G=\begin{Vmatrix}3 & 2.1 & 5\\ 4 & 0.7 & 1\\ 2 & 0.9 & 6\end{Vmatrix},\ G_1=\begin{Vmatrix}2 & 0.9 & 6\\ 3 & 2.1 & 5\\ 4 & 0.7 & 1\end{Vmatrix},\ G_2=\begin{Vmatrix}4 & 0.7 & 1\\ 2 & 0.9 & 6\\ 3 & 2.1 & 5\end{Vmatrix},\ G_3=\begin{Vmatrix}4 & 0.7 & 1\\ 3 & 2.1 & 5\\ 2 & 0.9 & 6\end{Vmatrix}.$$

Let class i of the TS object number s be determined by the dependence $i=g(s)$. Let us consider the properties of subsets of objects called clusters i, the features of which are described by row (s) of the matrix G_k when $i=g\left((s)^k\right)$.

Let the class i of the object OB number s, equal to the row number of the data matrix G, be given by the dependence $i=g(s)$. Let us consider the properties of object subsets called clusters i, the features of which are described by row (s) of the matrix G_k when $i=g\left((s)^k\right)$. Note that clusters i and classes i have the same length. To estimate the level of coincidence of objects of class i and cluster i, we find the average number of the TS objects for all classes for which the dependence is satisfied

$$g(s)=g\left((s)^k\right), \text{ where } k=1,\cdots,N.$$

The number of such matches, divided by the set length, is called the match index $\psi_k\in(0,1)$ of the feature k. Index analysis was performed for 10 databases [15]. Calculations showed that for a third of the databases considered, the maximum index value exceeds 0.9, 0.7 or 0.5, respectively, for one of the databases it reaches 0.961, and for another $\psi_k\sim 0$ for all k. From the results obtained, it follows that classes and clusters partition many objects into subsets, which partially (in many cases) or almost completely (in some cases) consist of the same objects. Note that the wide range of the index ψ_k values is partly caused by errors in measurements and the selection of features characterizing properties of the class objects.

According to the definition of the matrix G_k, the feature values k are ordered. Any segment of a sequence of ordered features consists of the nearest neighbors by feature value, and therefore, there is a probability that the corresponding objects belong to the same class and have common properties. It follows that in the case where the value of ψ_k significantly exceeds the average value of the index, it can be approximately assumed that the feature k determines the class of the TS objects, and the influence of the remaining features on the class will be insignificant.

Then, by dividing the set $\left\{(s)^k\right\}$ into C subsets whose length is equal to the number of objects in the corresponding class, we find lists of objects of each class i that apparently have common properties. Considering that the order of the classes along the sample length is arbitrary, we obtain $C!$ variants of partitioning each matrix G_k into classes. However, due to the high level of uncertainty, it is

advisable to consider the found classes as clusters that consist of objects with similar properties.

5. Conclusions

The paper proposes an atomistic approach to solving machine learning problems, according to which the elements of classes are not objects, but individual attribute values of objects. It is implemented by a bio-inspired concept for solving classification and clustering problems based on mapping features to ordered features by sorting each feature value of the TS in non-decreasing order, which leads to a decrease in entropy and chaos of the data.

This transformation made it possible to establish that any TS can be partitioned into patterns called macro-features, the elements of which are ordered features of objects of the same class. On macro-features, functions are defined that describe the distribution of feature values, as well as ordered features, over the length of the corresponding class.

Classification of test sample objects comes down to calculating the average frequencies of occurrence of their feature values in the nearest neighborhood of ordered feature values of objects of the same class.

The article develops an approximate method for dividing a data set into clusters of objects that differ in their common properties.

The obtained results indicate the advisability of developing neural networks based on the use of hidden variables. Instead of complex and cumbersome calculations, these networks will use the specified functions. Their monotonicity will ensure widespread use of approximation, as well as minimization of the amount of sampling. Networks of the new type will be distinguished by significantly lower costs of computer time.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Luger, G.F. (2016) Artificial Intelligence: Structures and Strategies for Complex Problem Solving. 6th Edition, Addison-Wesley.

[2] Solso, R. (2006) Cognitive Psychology. 6th Edition, Allyn and Bacon, 589.

[3] Shats, V.N. (2018) The Classification of Objects Based on a Model of Perception. In: Kryzhanovsky, B., *et al.*, Eds., *Advances in Neural Computation, Machine Learning, and Cognitive Research*, Springer International Publishing, 125-131. https://doi.org/10.1007/978-3-319-66604-4_19

[4] Smith, C.U.M. (2004) Biology of Sensory Systems. John Wiley and Sons Limited, 565.

[5] Shats, V.N. (2022) Properties of the Ordered Feature Values as a Classifier Basis. *Cybernetics and Physics*, **11**, 25-29. https://doi.org/10.35470/2226-4116-2022-11-1-25-29

[6] Hastie, T., Tibshirani, R. and Friedman, R. (2009) The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd Edition, Springer, 764.

[7] Shats, V.N. (2024) Feature Ordering as a Way to Reduce the Entropy of the Training Sample and the Basis of the Simplest Classification Algorithms. *Proceeding* 26*th International Conference Neuroinformatics*, Moskow, 24-26 October 2024, 164-173.

[8] Grenander, U. (1976) Lectures on Pattern Theory 1: Pattern Synthesis. Springer-Verlag.

[9] Feynman, R. (1965) The Character of Physical Law. A Series of Lectures Recorded by the BBC at Cornell University USA. Cox and Wyman.

[10] Prigogine, I. and Stengers, I. (1984) Order Out of Chaos: Men's New Dialogue with Nature. Flamingo Edition.

[11] Andrievskii, B.R. and Fradkov, A.L. (2003) Control of Chaos: Methods and Applications. I. Methods. *Automation and Remote Control*, **64**, 673-713. https://doi.org/10.1023/a:1023684619933

[12] David, H.A. and Nagaraja, H.N. (2003) Order Statistics. 3rd Edition, Wiley. https://doi.org/10.1002/0471722162

[13] Kolmogorov, A.N. and Fomin, S.V. (1957) Elements of the Theory of Functions and Functional Analysis. Vol. 1 Metric and Normed Spaces. Graylock Press.

[14] Asuncion, A. and Newman, D. (2007) UCI Machine Learning Repository. Irvine University of California.

[15] Shats, V.N. (2023) Principle Splitting of Finite Set in Classification Problem. *Proceeding* 25*th International Conference Neuroinformatics*, Moskow, 23-27 October 2023, 262-270.