## PAPER • OPEN ACCESS

# Logical aspects of machine learning

To cite this article: A N Shevlyakov 2018 J. Phys.: Conf. Ser. 1050 012078

View the article online for updates and enhancements.

## You may also like

- <u>The impact of herbicides on the weed</u> infestation and yield of false flax (Camelina Sativa) in the southern wooded steppe of Omskaya oblast Yu V Frizen, E V Nekrasova, T V Nozhenko et al.
- <u>The influence of the underlying terrain on</u> the directional characteristics of the HFrange antennas S V Krivaltsevich, A S Yashchenko, K A Maynenger et al.
- <u>The species composition and the</u> <u>taxonomic structure of phytoplankton in</u> <u>the lakes of Omsk region</u> O P Bazhenova, N N Barsukova, L V Korzhova et al.





DISCOVER how sustainability intersects with electrochemistry & solid state science research



This content was downloaded from IP address 3.22.248.208 on 05/05/2024 at 08:49

**IOP** Publishing

## Logical aspects of machine learning

## A N Shevlyakov<sup>1 2</sup>

<sup>1</sup> Sobolev Institute of Mathematics, Pevtsova 13, Omsk, 644099, Russia Omsk State Technical University, pr. Mira 11, Omsk, 644050, Russia

E-mail: a\_shevl@mail.ru

**Abstract.** Using the basic ideas of universal algebraic geometry, we define a distance function on the set of quantifier-free formulas. We explain the general definition by two examples (partial orders and feature selections) and prove basic facts about the distance in such examples.

#### 1. Introduction

Many machine learning algorithms of prediction and clustering work in metric spaces generated by features of learning objects. Recall that learning objects may have arbitrary types, and there is a need to define a metric between such objects.

One can explain the importance of such metrics by the following example. Suppose we have N forecast statements asserted by N experts. If N is large, we can not effectively process this data. Thus, we have to decrease the number of forecasts removing the "almost identical" statements. If we define a distance function over the set of statements, then a pairs of statements with a small distance can be treated as a pair of "almost identical" statements.

In the current paper we define a distance function on the set of first-order  $\mathcal{L}$ -formulas in variables X, where the values of X belong to a fixed  $\mathcal{L}$ -structure  $\mathcal{A}$  (Section 3). The idea of our approach is based on the results of universal algebraic geometry [1]. In Sections 4, 5 we consider two applications of the defined metric. Namely, in Section 4 we study statements with partial order, whereas in Section 5 we show the application of the defined metric in feature selection.

Our approach was inspired by the following problem in theoretical machine learning. There are popular algorithms of clustering (e.g. FOREL, MeanShift...) which claim the information about the distance distribution of clustered objects. If we give a wrong information to such algorithm, the results of clustering will be either trivial or not interpretable. Our definition of the metric allows to estimate a priori the average distance between objects before clustering (see Sections 4, 5).

#### 2. Main definitions

Let  $\mathcal{L}$  be a language. Below we do not assume that the symbol = belongs to  $\mathcal{L}$  on default.

An  $\mathcal{L}$ -equation is an atomic formula of  $\mathcal{L}$ . An arbitrary set of  $\mathcal{L}$ -equations is called an  $\mathcal{L}$ -system (system, for shortness). A disjunction of a finite number of  $\mathcal{L}$ -systems is a  $\lor$ -system.

Let  $\mathcal{A}$  be an  $\mathcal{L}$ -structure. A point  $P \in \mathcal{A}^n$  is a solution of a system **S** if P satisfies all equations in **S**. A point  $P \in \mathcal{A}^n$  is a solution of a  $\vee$ -system  $\mathbf{S}_1 \vee \mathbf{S}_2 \vee \ldots \vee \mathbf{S}_r$  if P is a solution of at least one system  $\mathbf{S}_i$ . The set of all solutions of an  $(\vee$ -system **S** is denoted by  $V_{\mathcal{A}}(\mathbf{S})$ .

<sup>2</sup> Present address: Sobolev Institute of Mathematics, Pevtsova 13, Omsk, 644099, Russia

**IOP** Publishing

A set  $Y \subseteq \mathcal{A}^n$  is called  $(\lor)$ -algebraic over an  $\mathcal{L}$ -structure  $\mathcal{A}$  if there exists an  $(\lor)$ -system **S** in *n* variables with  $V_{\mathcal{A}}(\mathbf{S}) = Y$ . An  $(\lor)$ -algebraic set *Y* is *irreducible* if *Y* is not a proper finite union of other  $\lor$ -algebraic sets. The *radical*  $\operatorname{Rad}_{\mathcal{A}}(Y)$  of an algebraic set *Y* is the set of all  $\mathcal{L}$ -equations satisfying all points  $P \in Y$ .

An algebraic  $\mathcal{L}$ -structure  $\mathcal{A}$  is equationally Noetherian if for any infinite system  $\mathbf{S}$  in variables  $X = \{x_1, x_2, \ldots, x_n\}$  there exists a finite subsystem  $\mathbf{S}' \subseteq \mathbf{S}$  with the same solution set in  $\mathcal{A}$ .

**Proposition 2.1.** [1] Any algebraic set Y over an equationally Noetherian algebraic  $\mathcal{L}$ -structure  $\mathcal{A}$  is a finite union of irreducible sets

$$Y = Y_1 \cup Y_2 \cup \ldots \cup Y_m, \quad Y_i \nsubseteq Y_j \text{ for all } i \neq j,$$

and this decomposition is unique up to a permutation of components.

The subsets  $Y_i$  from Proposition 2.1 are called the *irreducible components* of Y.

The following result gives the decomposition of a  $\lor$ -algebraic set into the union of its irreducible components.

**Proposition 2.2.** Let Y be the solution set of a  $\lor$ -system  $\mathbf{S}_1 \lor \mathbf{S}_2 \lor \ldots \lor \mathbf{S}_r$  over an equationally Noetherian  $\mathcal{L}$ -structure  $\mathcal{A}$ , and

$$Y_i = \mathcal{V}_{\mathcal{A}}(\mathbf{S}_i) = \bigcup_j Y_{ij}$$

be the decomposition of  $Y_i$  into the union of irreducible components.

Then the sets

$$\{Y_{kl} \mid Y_{kl} \nsubseteq Y_{ij} \text{ for all } (i,j) \neq (k,l)\}$$

are irreducible components of Y.

#### 3. The distance

In the current section we define the distance function on the class of first-order formulas according to the following principles:

- (i) we define a distance  $d(\psi_1, \psi_2)$  just for a pair  $(\psi_1, \psi_2)$ , where each  $\psi_i$  is a  $\lor$ -system;
- (ii) we reduce the computation of  $d(\psi_1, \psi_2)$  to distances between irreducible components of  $V_{\mathcal{A}}(\psi_i)$  (where  $\mathcal{A}$  is an appropriate algebraic  $\mathcal{L}$ -structure).

Let us explain the assumptions above.

- (i) Let us explain why we do not consider firs-order formulas with with quantifiers or negations. Actually, this assumption is not strict. Indeed, for any formula  $\psi(X)$  with quantifier one can introduce a new predicate symbol  $R_{\psi}(X)$  with the same domain (Morleyization). Similarly, one can eliminate negations in any formula.
- (ii) We state that the differences in irreducible components of  $\vee$ -algebraic sets  $Y = V_{\mathcal{A}}(D), Z = V_{\mathcal{A}}(E)$  naturally express the distance between the formulas D, E. Let us consider a simple example.

Let  $\mathcal{L} = \{\leq\}$ ,  $\mathcal{A}$  be a linearly ordered  $\mathcal{L}$ -structure. How can we estimate the distance (difference) between:

$$\psi_1 \colon (x_1 \le x_2) \land (x_1 \le x_3), \tag{1}$$

$$\psi_2 \colon (x_1 \le x_3) \land (x_2 \le x_3), \tag{2}$$

The first (second) formula does not assert about the relation between  $x_2$  and  $x_3$  ( $x_1$  and  $x_2$ ). Thus,  $\psi_1$  admits either  $x_2 \leq x_3$  or  $x_3 \leq x_2$ , and  $\psi_2$  admits either  $x_1 \leq x_2$  or  $x_2 \leq x_1$ . Therefore,  $\psi_1, \psi_2$  are respectively equivalent to

$$(x_1 \le x_2 \le x_3) \lor (x_1 \le x_3 \le x_2),$$
  
 $(x_1 \le x_2 \le x_3) \lor (x_2 \le x_1 \le x_3),$ 

and we actually decompose  $\psi_i$  into the union of its elementary cases. The difference between  $\psi_1, \psi_2$  is hidden in the differences between the elementary cases of the statements  $\psi_i$ , However, the elementary cases of  $\psi_i$  are exactly the irreducible components of  $\vee$ -algebraic sets  $\mathcal{VA}(\psi_i)$ .

Let us mention the another advantage of the metric based on irreducible components. For many important (see Sections 4, 5) algebraic structures  $\mathcal{A}$  first-order definable sets and  $\vee$ algebraic sets have a complicated structure, and it is difficult to directly work with them. Whereas the description of irreducible sets over  $\mathcal{A}$  is often simple and there are clear algorithms which decompose any  $\vee$ -algebraic set into the union if its irreducible component.

To compute the distance  $d(\psi_1, \psi_2)$  between  $\lor$ -systems  $\psi_1, \psi_2$  we propose the following:

- (i) fix an  $\mathcal{L}$ -structure  $\mathcal{A}$ ;
- (ii) denote  $Y = V_{\mathcal{A}}(\psi_1), Z = V_{\mathcal{A}}(\psi_2).$
- (iii) let  $Y = \bigcup_{i=1}^{k} Y_i$ ,  $Z = \bigcup_{i=1}^{l} Z_i$  be the irreducible decompositions of Y, Z;
- (iv) put  $d(\psi_1, \psi_2) = d(Y, Z) = \{\max d(Y_i, Z_j) \mid Y_i \notin Z, Z_i \notin Y\}.$

Thus, the distance between Y, Z is reduced to the distance between their irreducible components. The choice of a distance between irreducible sets depends on the universal theory of  $\mathcal{A}$ . One can naturally define d for the most important algebraic structures  $\mathcal{A}$  (see the next two sections).

#### 4. Partial orders

Motivation. There are many models in economics and machine learning, where the order between parameters is significant as well as their exact values. For example, the information about the order between the parameters

 $x = \{ \text{US dollar to ruble rate for tomorrow} \},\$ 

 $y = \{ US \text{ dollar to ruble rate for the day after tomorrow} \}$ 

is sufficient for the optimal tomorrow's currency exchange.

Thus, there appears an important problem in the studying of first-order formulas with the relation of partial order. The approach developed in Section 3 allows us to define a distance between such formulas.

Let  $\mathcal{L} = \{\leq\}$  be the language with the predicate symbol of a partial order, and we consider  $\lor$ -systems of  $\mathcal{L}$ -equations over a linearly ordered  $\mathcal{L}$ -structure  $\mathcal{A}$ . One can directly prove that  $\mathcal{A}$  is equationally Noetherian, and Proposition 2.1 holds.

Let us define an algorithm which decomposes any  $\lor$ -algebraic set into a union of its irreducible components. By Proposition 2.2, it is enough to give such algorithm for algebraic sets.

**Proposition 4.1.** Let **S** be a system in variables  $X = \{x_1, x_2, ..., x_n\}$ . The irreducible components of  $Y = V_{\mathcal{A}}(\mathbf{S})$  are all possible linear orderings of X consistent with **S**. For example, the irreducible components of  $V_{\mathcal{A}}(\{x_1 \le x_2, x_3 \le x_4\})$  are defined by the following systems:

$$\{x_3 \le x_4 \le x_1 \le x_2\}, \{x_3 \le x_1 \le x_4 \le x_2\}, \{x_3 \le x_1 \le x_2 \le x_4\}, \{x_3 \le x_1 \le x_2 \le x_4 \le x_2 \le x_4 \le x_2 \le x_4 \le x_2 \le x_4 \le x_4 \le x_2 \le x_4 \le x_4$$

 $\{x_1 \le x_3 \le x_4 \le x_2\}, \{x_1 \le x_3 \le x_2 \le x_4\}, \{x_1 \le x_2 \le x_3 \le x_4\}.$ 

Propositions 2.2,4.1 allows us to decompose any  $\vee$ -algebraic set over  $\mathcal{A}$ . Let us define a distance between irreducible algebraic set over  $\mathcal{A}$  by the following way:

- (i) since any irreducible set Y defines a linear order on the set of variables  $X = \{x_1, x_2, \dots, x_n\}$ , one can denote the ordered vector of X by  $\overrightarrow{Y}$  (for, example, for  $Y = V_{\mathcal{A}}(\{x_3 \le x_1 \le x_4 \le x_2\})$  we have  $\overrightarrow{Y} = (3, 1, 4, 2)$ );
- (ii) the distance between irreducible algebraic sets Y, Z is the Hamming distance between  $\overrightarrow{Y}, \overrightarrow{Z}$ .

**Example 4.2.** Suppose  $\psi_1, \psi_2$  are defined by (1,2) then:

(i) using  $\mathcal{A}$  defined above, we have

$$Y = \mathcal{V}_{\mathcal{A}}(\psi_1) = \mathcal{V}_{\mathcal{A}}((x_1 \le x_2 \le x_3)) \cup \mathcal{V}_{\mathcal{A}}(x_1 \le x_3 \le x_2),$$
$$Z = \mathcal{V}_{\mathcal{A}}(\psi_2) = \mathcal{V}_{\mathcal{A}}(x_1 \le x_2 \le x_3) \cup \mathcal{V}_{\mathcal{A}}(x_2 \le x_1 \le x_3).$$

(ii) by the definition,

$$d(\psi_1, \psi_2) = d(\mathcal{V}_{\mathcal{A}}(x_1 \le x_3 \le x_2), \mathcal{V}_{\mathcal{A}}(x_2 \le x_1 \le x_3)) = d((1, 3, 2), (2, 1, 3)) = 3.$$

One should remark in conclusion that the definition of  $d(\psi_1, \psi_2)$  allows to obtain upper and lower bound for the average value of  $d(\psi_1, \psi_2)$  on the set of  $\lor$ -systems in *n* variables.

#### 5. Unary predicates

**Motivation.** In machine learning and statistics, feature selection is the process of selecting a subset of relevant features (variables, predictors) for further use in model construction. Any algorithm of feature selection divides the set of original features into the groups of relevant and irrelevant features (also, an algorithm may decline to classify some features). Different algorithms of feature selections may give different answers about the relevancy (irrelevancy) of features. Therefore, there arises a question: how to estimate the closeness between the results of feature selection algorithms?

The answer comes from the general scheme defined in Section 3. Let  $\mathcal{L} = \{R, R'\}$  be the language of two unary predicates, and  $\mathcal{A}$  an  $\mathcal{L}$ -structure, where the predicates R(x), R'(x) satisfy  $\mathcal{A} \models \forall x \neg (R(x) \leftrightarrow R'(x))$ . The  $\mathcal{L}$ -structure  $\mathcal{A}$  is the set of features, and the predicate R(x) (R'(x)) has the interpretation: "a feature x is relevant (irrelevant)".

The result of any feature selection algorithm may be expressed by an appropriate  $\lor$ -system. For example, the following  $\lor$ -systems

$$\psi_1: \ (R(x_1) \lor R'(x_2)) \land (R'(x_1) \lor R(x_2)), \tag{3}$$

$$\psi_1: \ (R(x_1) \lor R(x_2)) \land (R'(x_1) \lor R'(x_2)) \tag{4}$$

respectively state that

- (i) "exactly one of the features  $x_1, x_2$  is relevant";
- (ii) "the features  $x_1, x_2$  are relevant or irrelevant simultaneously".

Mechanical Science and Technology Update (MSTU-2018)

IOP Conf. Series: Journal of Physics: Conf. Series **1050** (2018) 012078 doi:10.1088/1742-6596/1050/1/012078

The distance on the set of  $\lor$ -systems induces a metric on the class of feature selection algorithms. The following statement is an analogue of Proposition 4.1.

**Proposition 5.1.** Let **S** be a system in variables  $X = \{x_1, x_2, ..., x_n\}$ . A variable  $x \in X$  is called undefined if neither R(x) nor R'(x) belongs to **S**. Then  $Y = V_{\mathcal{A}}(\mathbf{S})$  is the proper union

$$Y = \mathcal{V}_{\mathcal{A}}(\mathbf{S} \cup \{R(x)\}) \cup \mathcal{V}_{\mathcal{A}}(\mathbf{S} \cup \{R'(x)\}).$$
(5)

Applying the splitting (5) of Y to all undefined variables, we obtain the irreducible decomposition of the original algebraic set.

One can directly check that  $\mathcal{A}$  is equationally Noetherian, and Proposition 2.1 holds. Let us define a distance between irreducible algebraic set over  $\mathcal{A}$  by the following way:

- (i) by Proposition 5.1, for any irreducible set Y over  $\mathcal{A}$  and for any variable  $x \in X$  we have either  $R(x) \in \operatorname{Rad}_{\mathcal{A}}(Y)$  or  $R'(x) \in \operatorname{Rad}_{\mathcal{A}}(Y)$ ;
- (ii) therefore Y is defined by a (0,1)-vector  $\overrightarrow{Y} = (y_1, y_2, \dots, y_n)$  such that

$$y_i = \begin{cases} 0, \text{ if } R'(x_i) \in \operatorname{Rad}_{\mathcal{A}}(Y), \\ 1, \text{ if } R(x_i) \in \operatorname{Rad}_{\mathcal{A}}(Y) \end{cases}$$

(iii) the distance between irreducible algebraic sets Y, Z is the Hamming distance between  $\overrightarrow{Y}, \overrightarrow{Z}$ .

**Example 5.2.** Let  $\lor$ -systems  $\psi_1, \psi_2$  be defined by (3,4). Let  $\mathcal{A}$  be an  $\mathcal{L}$ -structure satisfying the formula

$$\forall x \neg (R(x) \leftrightarrow R'(x)).$$

We have that the sets  $Y_i = V_{\mathcal{A}}(\psi_i)$  have the following irreducible decompositions

$$Y_1 = \mathcal{V}_{\mathcal{A}}(\{R(x_1), R'(x_2)\}) \cup \mathcal{V}_{\mathcal{A}}(\{R'(x_1), R(x_2)\}) = Y_{11} \cup Y_{12},$$

$$Y_2 = V_{\mathcal{A}}(\{R(x_1), R(x_2)\}) \cup V_{\mathcal{A}}(\{R'(x_1), R'(x_2)\}) = Y_{21} \cup Y_{22}.$$

The irreducible sets  $Y_{ij}$  define the following vectors  $\overrightarrow{Y_{ij}}$ :

$$\overrightarrow{Y_{11}} = (1,0), \ \overrightarrow{Y_{12}} = (0,1), \ \overrightarrow{Y_{21}} = (1,1), \ \overrightarrow{Y_{22}} = (0,0).$$

We have

$$d(Y_{11}, Y_{21}) = d(Y_{11}, Y_{22}) = d(Y_{12}, Y_{21}) = d(Y_{12}, Y_{22}) = 1.$$

Thus,  $d(\psi_1, \psi_2) = 1$ .

The definition of the distance allows us to estimate the average distance between two irreducible sets. One can directly check that the number of (0, 1)-vector pairs  $(\overrightarrow{Y}, \overrightarrow{Z})$  with the Hamming distance k is equal  $2^n \binom{n}{k}$ . Therefore, the average distance between two irreducible sets is

$$\frac{1}{2^n \cdot 2^n} \sum_{k=0}^n k 2^n \binom{n}{k} = n/2$$

#### Acknowledgments

The author was supported by Russian Research Fund (project 17-11-01117).

#### References

 [1] Daniyarova E, Myasnikov A and Remeslennikov V 2017 Algebraic geometry over algebraic structures (Novosibirsk: SO RAN)