U. Yefremov, PhD in CS., As. Prof.
S. Kukharonok, Senior lecturer
A. Yeremenko, Student
Zhytomyr Polytechnic State University

CORE CONCEPTS OF MACHINE LEARNING

Learning is a process familiar to everyone. People learn every day and demonstrate excellent results in this process. They observe patterns in the changing environment around them, construct a certain model of the changes in this environment, and make certain decisions. The environment reacts in a certain way to the decisions made, and people again adjust their model of the world.

Machine learning is a simplified (machine-adapted) version of the learning process that occurs with humans. Typically, machine learning has a certain set of examples, observations, and responses to these observations. The task is to construct models that will effectively describe the available data and make reliable predictions.

Machine learning differs from what we are used to because we are trying to teach the computer to learn.

There are two types of learning:

1. *Inductive learning*
2. *Deductive learning*

Inductive learning is familiar to everyone because it involves observing the world and constructing models that explain the causes of various phenomena. These models are then repeatedly tested, some of them "surviving" to be used and improved, while others are completely discarded. *Deductive learning* is similar to school mathematics, where students are given ready-made formulas and taught how to apply them in practice.

Machine learning is inductive learning because we teach the machine mainly to learn from examples, observing a large number of real examples, building models on them, testing them, and applying them to more examples.

The goal of machine learning is to predict the outcome based on input data. The more diverse the input data, the easier it is for the machine to find patterns and the more accurate the result.

So, if we want to teach a machine, we need three things: *data, features*, and *algorithms.*

Data. If we want to detect spam, we need examples of spam emails; to predict stock prices, we need a history of prices; to learn about users' interests, we need their likes or posts. We need as much data as possible. Tens of thousands of examples is a minimum [1].

Data is collected in different ways. Some people do it manually – the process is slower, there is less data, but it is error-free. Other people do it automatically – they just feed the machine everything they find and hope for something better. The smartest ones, like Google, use their own users for free labeling. Think of ReCaptcha, which sometimes asks to "find all the road signs in the picture" – and that's it.

Signs. We call them features. Features, properties, characteristics, signs – they can be the mileage of a car, user gender, stock price, or even the frequency counter of a word appearing in a text.

The machine needs to know what exactly to look for. It is good when the data is simply laid out in tables – the names of their columns are the features. But what if we have a hundred gigabytes of cat pictures? When there are a lot of features, the model works slowly and inefficiently. Often, selecting the right features takes more time than anything else in training. However, there are also reverse situations when the user decides to select only the "correct" features from his point of view and introduces subjectivity into the model, in result it starts to lie wildly.

Algorithm. The same task can almost always be solved in different ways. The choice of method determines the accuracy, speed of operation, and size of the resulting model. But there is one nuance: if the data is "garbage", then even the best algorithm will not help. Don't get hung up on percentages; it's better to gather more data.

For further understanding, it is necessary to recognize the differences between *artificial intelligence*, *machine learning*, and *neural networks*.

Artificial intelligence is the name of the entire field, like biology or chemistry. *Machine learning* is one branch of artificial intelligence. Important, but not the only one that exists. *Neural networks* can be considered a type of machine learning. They are popular, but there are others just as good. Deep learning is the architecture of neural networks, one of the approaches to their construction and training. In practice, few distinguish where deep neural networks are used and where they are not very relevant. People usually mention the name of a specific network and that's it.

You can only compare things on the same level. Otherwise, it's complete nonsense, like "Which is better: a car or a wheel?". Don't equate terms without reason to avoid misunderstandings.

In practice, machine learning encompasses various techniques and methodologies, including *supervised* and *unsupervised* learning.

Supervised learning is similar to having a teacher who tells it how to do things right. It explains that in this picture, there is a cat, and in this one, there is a dog. So, the teacher has already divided all the data into cats and dogs, and the machine learns from specific examples [2].

In *unsupervised* learning, the machine is simply given a lot of animal photos on the table and told, "Figure out who looks like whom here." The data is not labeled, there is no teacher for the machine, and it tries to find any patterns on its own without explicit guidance.

Obviously, with a teacher, the machine learns faster and more accurately, so it is much more often used in practical tasks. These tasks can be divided into two types: *classification*, which predicts the category of an object, and *regression*, which predicts a position on a numerical line.

Today, a computer only executes precise instructions provided by a human. When writing any application, a programmer uses a high-level language, and then a translator program translates this application into machine language directives that the computer processor understands. Thus, it becomes clear that the computer itself is fundamentally incapable of thinking, but high-level programs make it relatively intelligent.

Despite the distant prospect of ubiquitous artificial intelligence, its components are successfully used for solving practical tasks. A vivid example of this is machine learning.

REFERENCES

1. Eli N. Weinstein, Jeffrey W. Miller Bayesian Data Selection – 2023. – p. 2.
2. Jonathan K. Su On Truthing Issues in Supervised Classification – 2024. – p. 10 - 11.