47.   NEURAL NETWORKS FUNDAMENTALS

Shmargun A.E.

Belarusian State University of Informatics and Radioelectronics
Minsk, Republic of Belarus

Subbotkina I.G. – Associate Professor

An overview of machine learning technology with a focus on supervised and unsupervised learning is provided in this paper. Neural networks and their various categories, such as Feedforward, Recurrent, Convolutional and Long Short-Term Memory Networks are introduced as well. The importance of using machine learning responsibly due to its potential risks is emphasized.

At present machine learning(ML) technology enables artificial intelligence to acquire knowledge and assimilate patterns from external data without being explicitly programmed. ML algorithms use statistical analysis to find patterns in data and make predictions based on these patterns.

It is important to mention two main kinds of ML: supervised and unsupervised learning. During supervised learning, the algorithm is undergoing a certain training on a labeled dataset, assuming that each data point is labeled with the correct answer. Afterwards, the algorithm, having considered the training, attempts to make predictions on a new set of data, being unlabeled this time. In unsupervised learning, the algorithm does not give any labeled data. It is required to structure the data based on its own capabilities. A number of known machine learning algorithms include linear regression, decision trees, and k-nearest neighbours. These algorithms find their use in a variety of applications that are aimed at fulfilling a number of tasks such as predicting stock prices, identifying fraudulent credit card transactions, and recognizing speech and images.

ML is presented as a tool set of unlimited capabilities, having already revolutionised in many fields like financial sphere, healthcare and logistics. ML must be used with respective responsibility. As any other technological breakthrough, it involves a fair degree of risk and requires a thorough analysis to avoid undesirable consequences. The schematic representation of ML is illustrated in Figure 1.
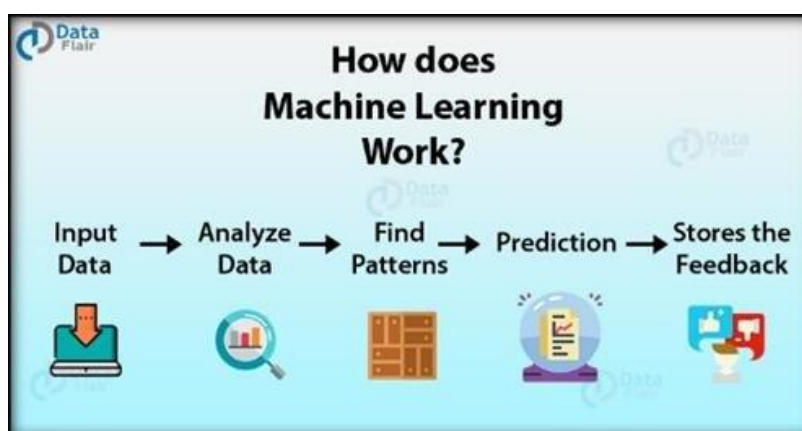

Figure 1 – Schematic representation of ML

Neural networks are a standalone class of ML algorithms that were created according to the structure and functionalities of a human. Such algorithms appear to be capable of handling given complex patterns in data and making predictions based on given patterns. Neural networks have become famous among users in recent years due to their capability of solving a broad selection of tasks, including image recognition, natural language processing, and speech recognition.

The primary operation principle in neural networks consists of multiple layers of interconnected nodes (neurons). The input layer perceives information and transfers it to the following layer where data will undergo a series of mathematical operations. Each subsequent layer performs more complex operations on data until it reaches the output layer for the final prediction.

In the process of training the neural network adjusts the weights of the connections between neurons to minimize the difference between the predicted output and the actual output. This process is called backpropagation and is applied to optimize the network's performance.

Neural networks are divided into various categories. Feedforward Neural Networks are the most trivial types of neural networks, which structure implies data flows in just one direction. Such neural networks are commonly applied for tasks such as classification and regression tasks.

Recurrent Neural Networks (RNNs) have loops in their architecture that enable them to process and assimilate sequential data packages such as time series or a natural language. Such neural networks have a potential to be used for tasks such as speech recognition and language translation.

Convolutional Neural Networks (CNNs) are designed with a specific purpose of image recognition. They are widely utilized in computer vision applications such as object detection and image segmentation.

Long Short-Term Memory (LSTM) Networks present a special type of RNN that has ability to store or remember information for extended periods of time. Such networks are used in natural language processing tasks such as speech recognition and language translation.

The schematic representation of neural network processes is illustrated in Figure 2.
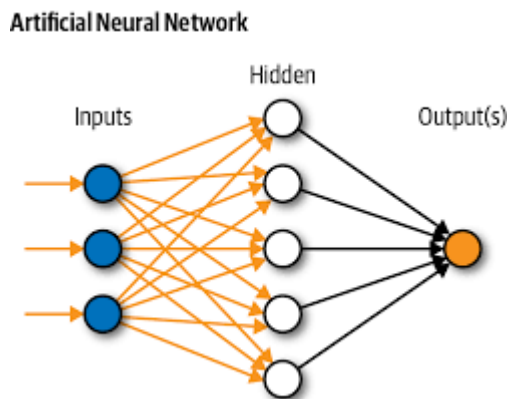
Artificial Neural Network

Figure 2 – Schematic representation of neural network processes

Chat Generative Pre-trained Transformer (GPT) is one of the most popular neural networks. It is a large language model developed by OpenAI, based on the GPT-3.5 architecture. Chat GPT has been trained on a vast amount of textual data, allowing it to generate coherent and contextually appropriate responses to a wide range of natural language inputs.

The architecture of Chat GPT consists of multiple layers of self-attention and feedforward neural networks. Each layer is designed to capture different levels of contextual information, from local word relationships to global document-level coherence.

Chat GPT was trained on a massive dataset of diverse textual data, including books, articles, and online conversations. The training process was unsupervised, meaning that the model learned to generate responses solely based on the patterns and structures present in the input data. The training was done using a technique called language modeling where the model is trained to predict the next word in a sentence given the preceding words. This process allows the model to learn the statistical patterns of language and generate coherent and contextually appropriate responses.

Chat GPT has numerous applications in various industries, including customer service, healthcare, and education. In customer service, it can be used to provide personalised and efficient responses to customer queries, improving customer satisfaction and reducing response time. In healthcare, Chat GPT gives a medical advice to patients based on their symptoms, reducing the burden on healthcare professionals. In education, it recommends personal learning experiences to students, adapting to their individual needs and abilities.

Chat GPT is a powerful language model that has the potential to revolutionize the way we communicate with machines. Its ability to generate coherent and contextually appropriate responses to a wide range of natural language inputs makes it a valuable tool in various industries. However, there are also concerns about the ethical implications of such technology, particularly in terms of privacy and bias. Further research is needed to address these concerns and ensure that Chat GPT is developed and used responsibly.

Neural networks are powerful machine learning algorithms which are capable of processing and seeking solutions for complex problems and performing accurate predictions. Neural networks have a broad range of applications in the areas of image recognition, natural language processing and predictive analytics. There is an assumption, that in future there will be an opportunity to use neural networks in more innovative spheres.

References:
1. Getting Started with Neural Networks in Python [Electronic resource]. – Mode of Access: *https://ai.plainenglish.io/getting-started-with-neural-networks-in-python-b4efc76ef91*. – Date of Access: 07.03.23.
2. Machine learning, explained [Electronic resource]. – Mode of Access: *https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained*. – Date of Access: 12.03.23.
3. Supervised vs. Unsupervised Learning: What's the Difference? [Electronic resource]. – Mode of Access: *https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning.* – Date of Access: 16.03.23.
4. Grokking Algorithms [Electronic resource]. – Mode of Access: *https://www.manning.com/books/grokking-algorithms.* – Date of Access: 20.03.23.