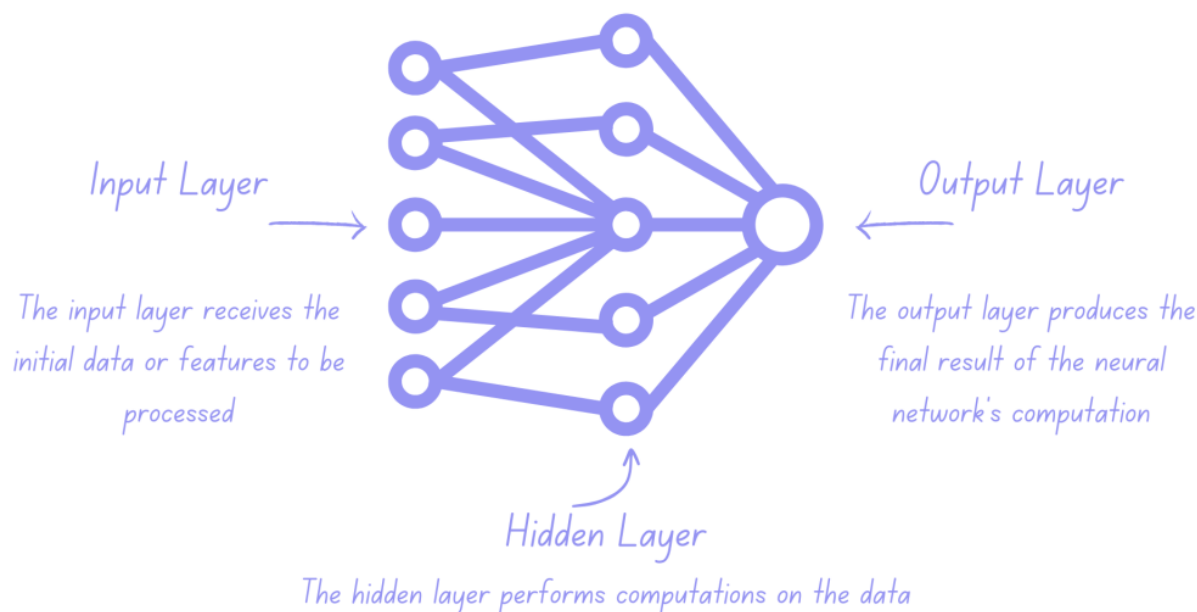


Neural Networks are a type of computational model inspired by the structure and functioning of a human brain. They are composed of interconnected nodes called neurons, organized into layers. Each neuron receives inputs, processes them, and produces an output that is passed on to other neurons. Understanding how neural networks work is one of the first steps to learning Deep Learning. So, if you want to know how a Neural Network works, this article is for you. In this article, I'll explain how Neural Networks work.

## The Structure of Neural Networks



A neural network is a computational model inspired by the functioning of the human brain. It consists of interconnected nodes, called neurons, organized in multiple layers. The structure of a neural network typically includes:

- an input layer;
- one or more hidden layers;
- an output layer;

The input layer receives the initial data or features to be processed. Each neuron in the input layer corresponds to a specific input feature and serves as the starting point for information flow.

The hidden layers, located between the input and output layers, perform computations on the data. Each neuron in the hidden layers receives inputs from the previous layer, applies mathematical operations, and passes the results to the next layer. The number of hidden layers and neurons within each layer can vary depending on the complexity of the task.

The output layer produces the final result of the neural network's computation. Each neuron in the output layer represents a specific output or prediction. The information flows from the input layer through the hidden layers to the output layer, with computations and transformations occurring at each stage.

Connections between neurons are known as weights. Weights determine the strength of influence one neuron has on another. These weights are adjusted during the learning process of the neural network to optimize its performance. Activation functions are applied to the weighted sums of inputs in each neuron, introducing non-linearities and enabling the network to learn complex patterns and make predictions.

## **Now Here's How Neural Networks Work!**

At its core, a neural network contains interconnected nodes called neurons. These neurons are organized into layers: an input layer, one or more hidden layers, and an output layer. Each neuron is associated with a weight, representing the strength of the connection between neurons.

The process of training a neural network involves two main steps:

1. forward propagation
2. and backpropagation

During forward propagation, input data is fed into the network and flows through the layers. Each neuron in the input layer receives an input value and passes it to the neurons in the first hidden layer. Each connection between neurons is assigned a weight, which is multiplied by the input value. These weighted inputs are summed up, and a bias term (a constant) is added to introduce flexibility.

The weighted sum is then passed through an activation function, which introduces non-linearity into the network. The activation function determines whether the neuron should be activated or not, based on the input it receives.

The output of the activation function becomes the input of that particular neuron and is then it's passed to the next layer. This process continues until the output layer is reached, where the final result or prediction of the neural network is received.

Backpropagation adjusts the weights based on the difference between the predicted and desired output. It calculates the error at the output layer and propagates it backwards through the network. The weights are updated using gradient descent, taking the derivative of the error with respect to each weight. The weights are adjusted in the opposite direction of the gradient, aiming to minimize the error. This process is repeated iteratively, refining the network's weights to improve predictions.

This iterative process of forward propagation and backpropagation continues for a set number of epochs or until the network reaches a satisfactory level of accuracy.

## **But Why Do We Use Neural Networks?**

The structure of a neural network allows it to learn and adapt to different tasks by adjusting the weights and biases associated with the connections. They can iteratively improve their performance by comparing their predictions with the desired outputs and adjusting the weights accordingly. That is why Neural networks are widely used despite the existence of other machine learning algorithms due to their unique capabilities.

Neural networks can learn complex patterns and relationships in data, even when the relationships are non-linear or abstract. They can automatically extract relevant features from raw data, eliminating the need for manual feature engineering. Additionally, neural networks are highly flexible and can adapt to different problem domains and data types.

Their ability to model and approximate any arbitrary function makes them powerful tools for a wide range of tasks, such as image and speech recognition, natural language processing, and predictive modelling. Therefore, the use of neural networks is justified by their ability to handle complex and diverse data flexibly and powerfully, which sets them apart from other machine learning algorithms.

## **Summary**

So this is how Neural Networks work. Neural networks work by organizing interconnected nodes, called neurons, into layers: an input layer, one or more hidden layers, and an output layer. Each neuron receives inputs from the previous layer, performs computations using assigned weights, and passes the results to the next layer through an activation function. During training, forward propagation involves feeding input data through the network, calculating weighted sums, applying activation functions, and obtaining predictions. The backpropagation step adjusts the weights based on the prediction error, propagating it

backwards through the network. This adjustment is achieved through gradient descent, where the weights are updated in the opposite direction of the gradient to minimize the error.

The iterative process of forward propagation and backpropagation continues, refining the network's weights to improve predictions until a satisfactory level of accuracy is achieved. Overall, neural networks learn by adjusting weights to capture complex patterns in the data and make predictions or classifications. I hope you liked this article on how Neural Networks work. Feel free to ask valuable questions in the comments section below.