MAIN TYPES

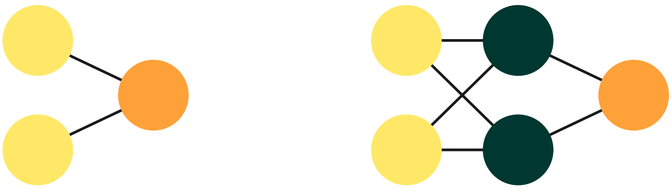
OF NEURAL NETWORKS

Everyone who has been in the world of artificial intelligence for some time knows the story: neural networks were invented back in the 1950s, but were left out for decades because of the limited computing capacity.

Once neural networks began to be explored for their ability to solve problems, it also became evident that their most basic form, the classic deep neural network, did not meet all needs. Researchers in the field then began to develop new ideas, based on both new types of neurons and more complex structures. Currently, among the most widely used **neural networks are deep, convolutional, recurrent, autoencoders, and generative networks**. But the universe of possibilities is wide, and several other architectures specializing in specific tasks have been well established.

In this article, **we present the main 5 types of neural networks** available to the developer in artificial intelligence.

1. PERCEPTRON (P), FEED FORWARD NETWORK (FFN), RADIAL BASIS NETWORK (RBF)



They are the most basic type, where the input information flows in a linear sequence to the output. In each neuron there is a linear mathematical operation of the type

$$Wx + b$$

W — is the weight of the neuron

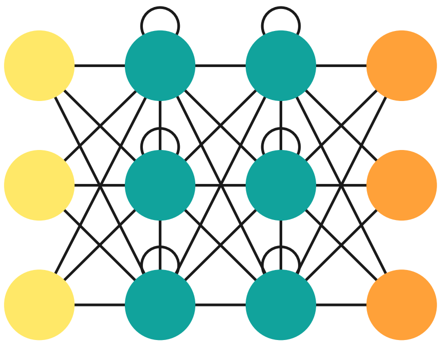
X — is the value of the input data

b — is the bias parameters of the neuron

The result of these operations can go through an activation function before moving on to the layer ahead; in the specific case of radial based activation functions (those that determine the distance from the result to a point of origin), for purely historical reasons, the corresponding network is called the radial based network. The connections between neurons represent the passage of information from one neuron to the next. A neuron that receives more than one input connection adds these values before applying the linear equation for which it is responsible. This neural network can have several layers hidden between the input and the output layer, so it is also often called the deep neural network.

FFNs are able to model various problems where the input data has a timeless impact on the output data. An example is using information from a blood test to determine the presence of a disease.

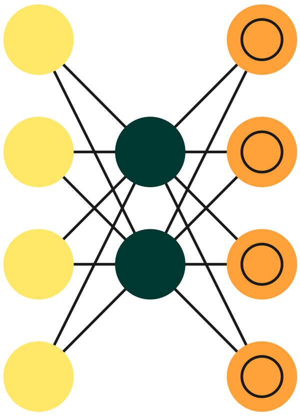
2. RECURRENT NEURAL NETWORK (RNN)



In addition to the data from the anterior layer, the hidden neurons from the recurrent neural network also receive the result of the mathematical operation that they themselves performed in the previous time period. Thus, the RNNs consider a temporal dependency between the input data.

Because these networks have this characteristic, they can model problems with temporal characteristics, such as the weather forecast given the climate history in a window of the past.

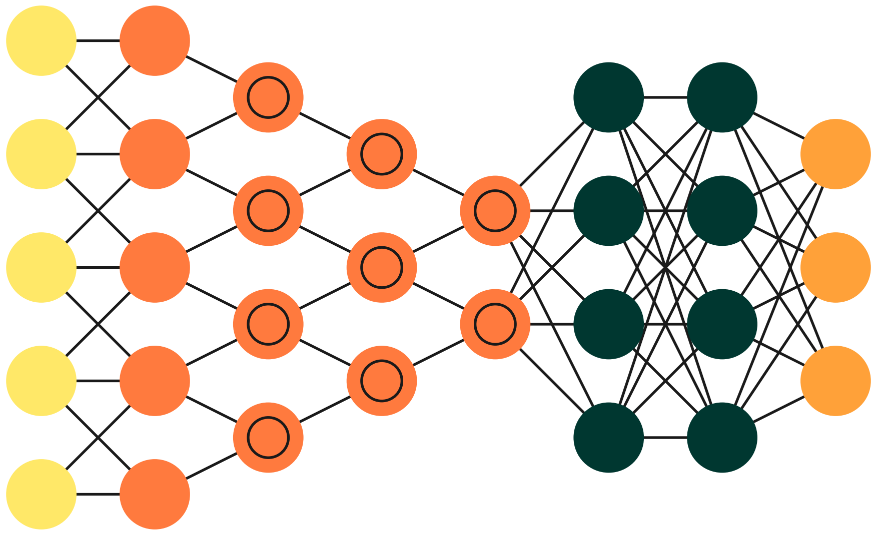
3. AUTO-ENCODER (AE)



Auto-encoders are designed to represent input information in a smaller dimensional space. Therefore, the central hidden layer of this network, which represents this space, must have fewer neurons than the input layer. The output layer is a copy of the input information, so that during training, auto-encoders learn to represent the original information in less space, but with enough information to reconstruct the original data. The first half of the network, which compresses the information, is called an encoder, and the second is a decoder.

Auto-encoders can be used both to compress data for storage and / or transmission, and to represent the data in a reduced form so that, for example, another neural network specialized in a specific task can use them.

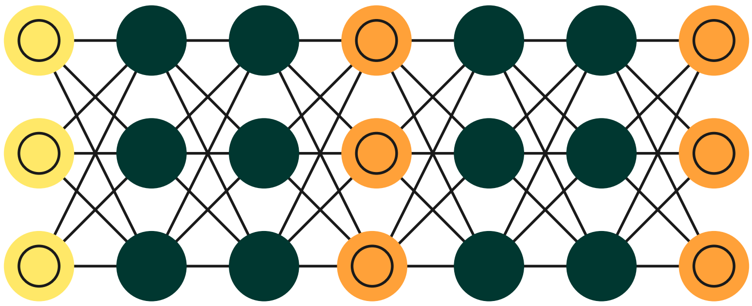
4. CONVOLUTIONAL NEURAL NETWORKS (CNN) OR DEEP CONVOLUTIONAL NETWORK (DCN)



The convolutional neural network has a very different structure from those presented so far. In the convolution layers, the information passes through several filters (which in practice are numerical matrices) with the function of accentuating regular local patterns, while reducing the size of the original data. The results of various filters are summarized by pooling operations. In the deepest part of the convolutions, data in a reduced dimensional space is expected to contain enough information about these local patterns to assign a semantic value to the original data. These data then go through a classic FFN structure for the classification task.

For these characteristics, the most common application of CNNs is in the classification of images; the filters accentuate the attributes of the objects necessary for their correct classification. A CNN specializing in classifying faces, for example, in the first layers recognizes contours, curves and borders; further on, it uses this information to recognize mouth, eyes, ear and nose; and in the end, it recognizes the entire face. In addition to images, any information with local regularity can benefit from the use of CNNs, such as audio for example.

5. GENERATIVE ADVERSARIAL NETWORK (GAN)



The generative adversarial network can be described as containing internally two networks that work together in a competitive way. The first network — the generative part — is responsible for generating content, being trained in the task of recreating the original information. The second network — the adversarial part — is responsible for judging the content, comparing the creation of the generative network with the original information. When the adversarial network believes that the information generated cannot pass for the original, the generative network is forced to improve its performance, until it succeeds in deceiving the other.

This gives the GANs the ability to create unprecedented content, since the exit from the network was produced by the generative part and deemed acceptable by the adversary party. This can be considered a primitive form of creativity, which is why GANs are often referred to as artificial intelligence artists.