**1. Multi-Layer Perceptrons**

The field of artificial neural networks is often just called neural networks or multi-layer perceptrons after perhaps the most useful type of neural network. A perceptron is a single neuron model that was a precursor to larger neural networks.

The power of neural networks comes from their ability to learn the representation in your training data and how to best relate it to the output variable that you want to predict.

The predictive capability of neural networks comes from the hierarchical or multi-layered structure of the networks.

## **2. Neurons**

The building block for neural networks are artificial neurons. Each neuron also has a bias which can be thought of as an input that always has the value 1.0 and it too must be weighted.

The weighted inputs are summed and passed through an activation function, sometimes called a transfer function.

Traditionally non-linear activation functions are used. This allows the network to combine the inputs in more complex ways and in turn provide a richer capability in the functions they can model. Non-linear functions like the logistic function also called the sigmoid function were used that output a value between 0 and 1 with an s-shaped distribution, and the hyperbolic tangent function also called tanh that outputs the same distribution over the range -1 to +1.

## **3. Networks of Neurons**

Neurons are arranged into networks of neurons.

Different Layers.

### Input or Visible Layers

### Hidden Layers

### Output Layer

## **4. Training Networks**

Once configured, the neural network needs to be trained on your dataset. You must first prepare your data for training on a neural network. Data must be numerical

Neural networks require the input to be scaled in a consistent way. You can rescale it to the range between 0 and 1 called normalization. Another popular technique is to standardize it so that the distribution of each column has the mean of zero and the standard deviation of 1.

### **Stochastic Gradient Descent**

The classical and still preferred training algorithm for neural networks is called stochastic gradient descent. The weights in the network can be updated from the errors calculated for each training example and this is called online learning. It can result in fast but also chaotic changes to the network.

**Once a neural network has been trained it can be used to make predictions.**