

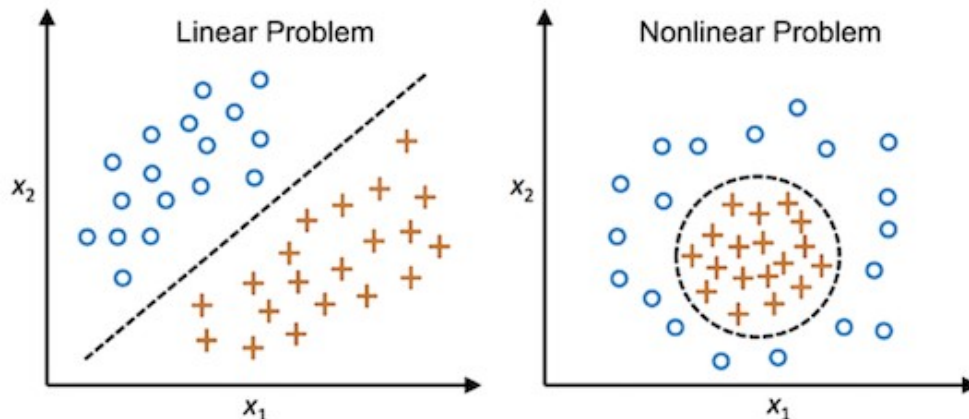
## Limitation of Perceptron Model – perceptron does not work on non-linear data

The output of a perceptron can only be a binary number (0 or 1) due to the hard-edge transfer function. It can only be used to classify the linearly separable sets of input vectors. If the input vectors are non-linear, it is not easy to classify them correctly.

## Linear and Non-Linear Separable Data

Linear Separability refers to the data points in binary classification problems which can be separated using linear decision boundary. If the data points can be separated using a line, linear function, or flat hyperplane are considered linearly separable. Linear separability is an important concept in neural networks

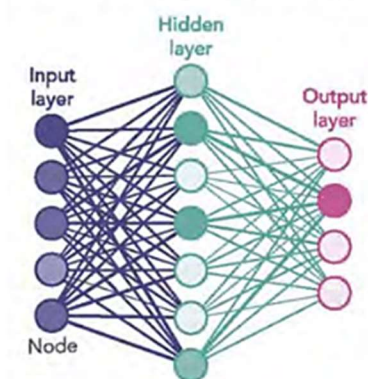
When we cannot separate data with a straight line we use Non – Linear SVM. In this, we have Kernel functions. They transform non-linear spaces into linear spaces. It transforms data into another dimension so that the data can be classified.



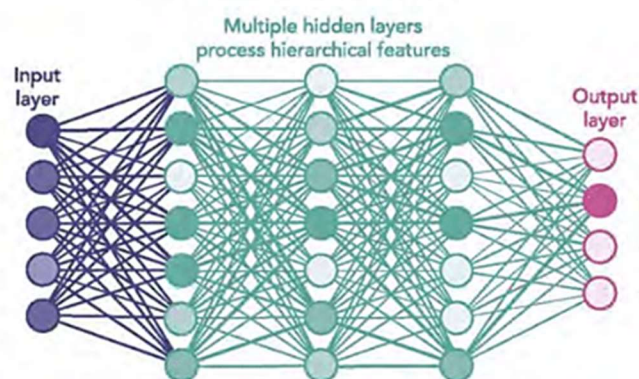
## Shallow Neural Network & Deep Neural Network

Shallow Neural Network has single hidden layer while Deep Neural Network has multiple hidden layer

SHALLOW NEURAL NETWORK



DEEP NEURAL NETWORK



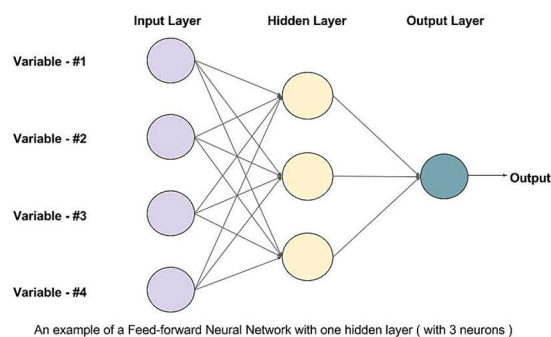
The main difference between deep and shallow neural networks lies in the number of layers they contain. Shallow neural networks, also known as single-layer neural networks, consist of only one hidden layer between the input and output layers.

On the other hand, deep neural networks have multiple hidden layers stacked between the input and output layers. These networks can be quite deep, with tens or even hundreds of layers. Each layer performs computations and passes the transformed data to the next layer.

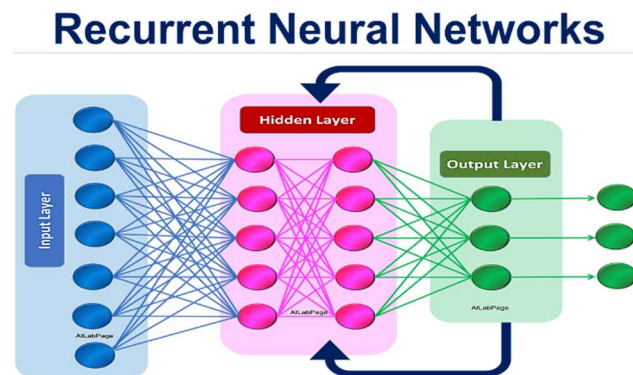
### Note:

- Binary classification – Sigmoid MLPs at output layer
- Multi classification that means no of classes is more than two – Softmax MLPs at output layer
- For salary/ banking purpose (for continuous data i.e. regression) – Linear activation function (by default)
- ReLu MLPs mostly used for input layers
- All types of functions graphs are between 0 to 1, except ReLu

### Network Topologies



### Feed Forward Neural Network



### Recurrent Neural Network

### Multilayer Perceptron – Ipython file

- Import libraries and data
- 80 percent train
- 20 percent test
- Random state 42 – consider good
- Standardized data – by using standard scale
- Then build data
- Neuron should not be less than features, it can be more

- Batch size greater => learning less while Batch size less => means computational less
- Verbose values – 0,1,2 (mainly used 1)
- Epoch greater means => increase in time, mean square error also reduce, accuracy increase
- Normal batch size => 32, 64 (because if it is greater that more data will pass through it while if it is less that greater number of data will pass through it and model will try to learn as many things as possible and both possibilities cannot give good results)
- For greater data batch size could be 100