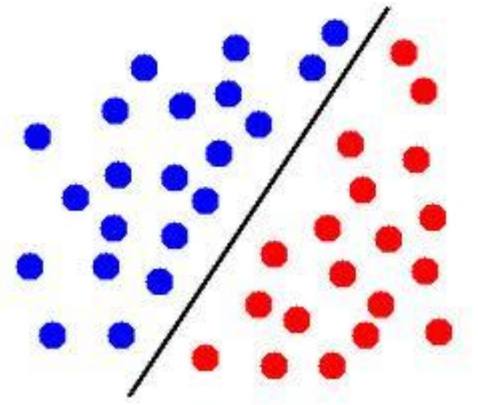
Linear separable means that there is a hyperplane

This means that there is a hyperplane, which splits your input data into two half-spaces such that all points of the first class should be in one half-space and other points of the second class should be in the other half-space.

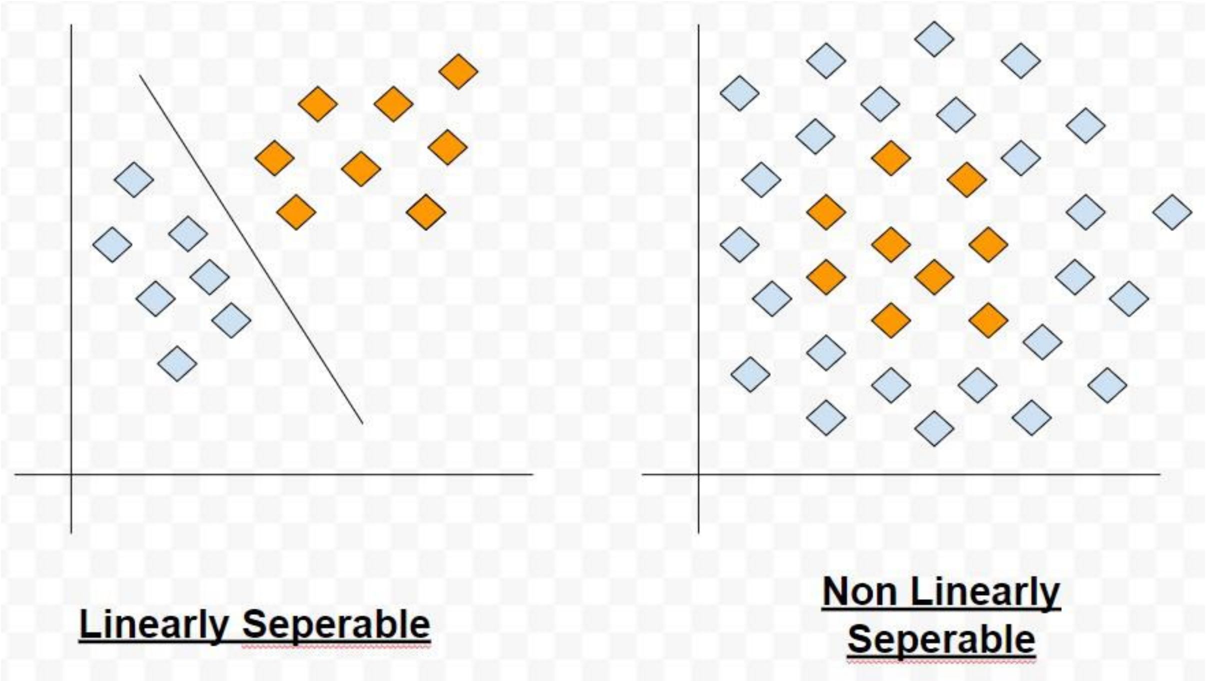
In two dimensional space, it means that there is a line, which separates points of one class from points of the other class.

For example: In the following image, if blue circles represent points from one class and red circles represent points from the other class, then these points are linearly separable.



In three dimensions, it means that there is a plane that separates points of one class from points of the other class.

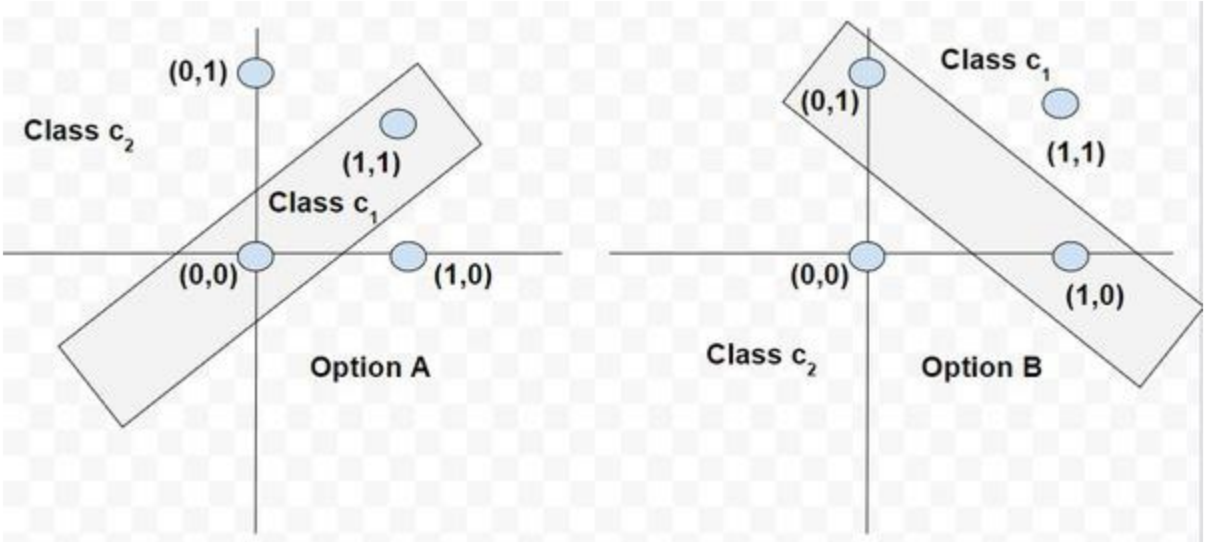
Thus, we see that for a data set with linearly separable classes, perceptrons can always be employed to solve classification problems using decision lines (for 2-dimensional space), decision planes (for 3-dimensional space) or decision hyperplanes (for n-dimensional space). Appropriate values of the synaptic weights can be obtained by training a perceptron. However, one assumption for perceptron to work properly is that the two classes should be linearly separable i.e. the classes should be sufficiently separated from each other. Otherwise, if the classes are non-linearly separable, then the classification problem cannot be solved by perceptron



Multi-layer perceptron: A basic perceptron works very successfully for data sets which possess linearly separable patterns. However, in practical situations, that is an ideal situation to have. This was exactly the point driven by Minsky and Papert in their work in 1969. They showed that a basic perceptron is not able to learn to compute even a simple 2 bit XOR. So, let us understand the reason.
Consider a truth table highlighting output of a 2 bit XOR function:

x_1	x_2	x_1 XOR x_2	Class
1	1	0	c2
1	0	1	c1
0	1	1	c1
0	0	0	c2

The data is not linearly separable. Only a curved decision boundary can separate the classes properly. To address this issue, the other option is to use two decision boundary lines in place of one.



This is the philosophy used to design the multi-layer perceptron model. The major highlights of this model are as follows:

- The neural network contains one or more intermediate layers between the input and output nodes, which are hidden from both input and output nodes
- Each neuron in the network includes a non-linear activation function that is differentiable.
- The neurons in each layer are connected with some or all the neurons in the previous layer.