Deep networks and shallow network (hint : xoring technique)

Single Perceptron:

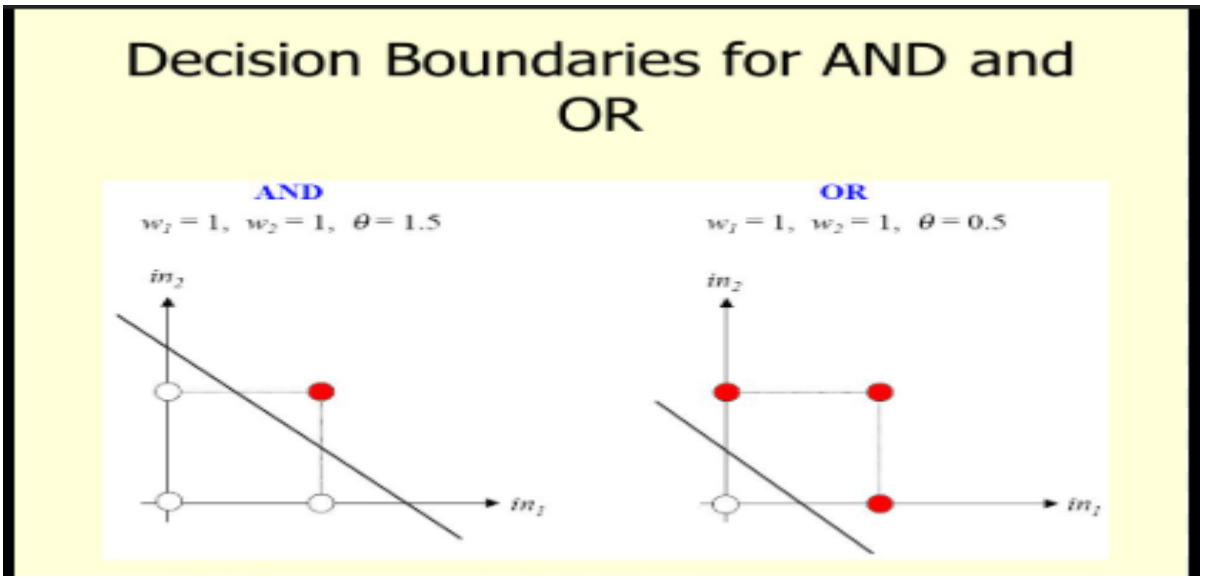
- For a specific input, if the linear sum of weighted inputs is greater than zero, then the neuron fires. It means that a particular input belongs to Class 1.
- For a specific input, if the linear sum of weighted inputs is less than zero, then neurons will not fire. It means that particular input belongs to Class 0.

So, the decision line equation is obtained when the linear sum of weighted inputs is equal to zero (It is a line equation, w1x1 + w2x2 + w0 = 0)

Data Separability:

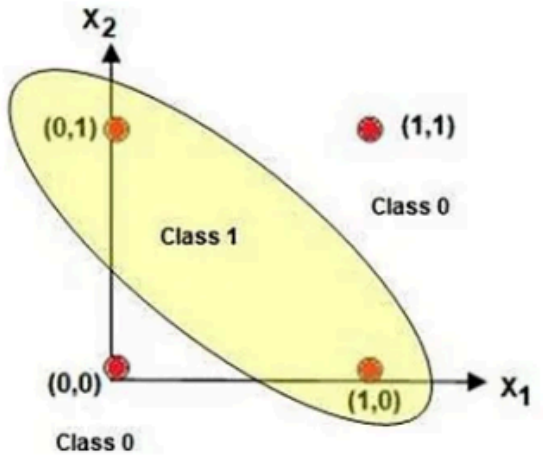
Firstly, let us understand that there are 2 kinds of data in data classification
— Linearly separable and Non-linearly separable data.

- One is where data can be separated into two classes by drawing a line (linear decision boundary) in the feature space
- Second type is where the data cannot be separated using a linear decision boundary.

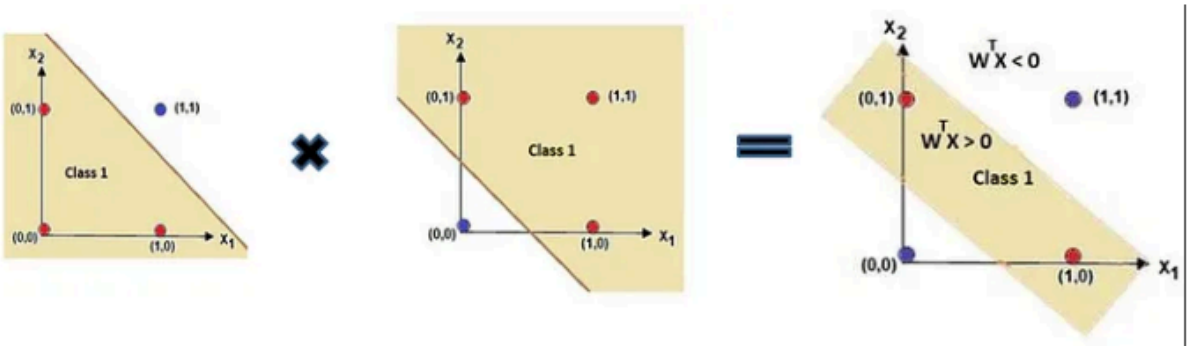


For example here we find AND and OR representation can be done using a single decision boundary because a single perceptron is only capable of drawing a line (or a hyperplane) but this is not the case when trying to represent XOR.

XOR Problem



Truth Table for XOR Gate		
X1	X2	Output Class
0	0	0
0	1	1
1	0	1
1	1	0



Truth Table - NAND		
X1	X2	Output Class
0	0	1
0	1	1
1	0	1
1	1	0

Truth Table - OR		
X1	X2	Output Class
0	0	0
0	1	1
1	0	1
1	1	1

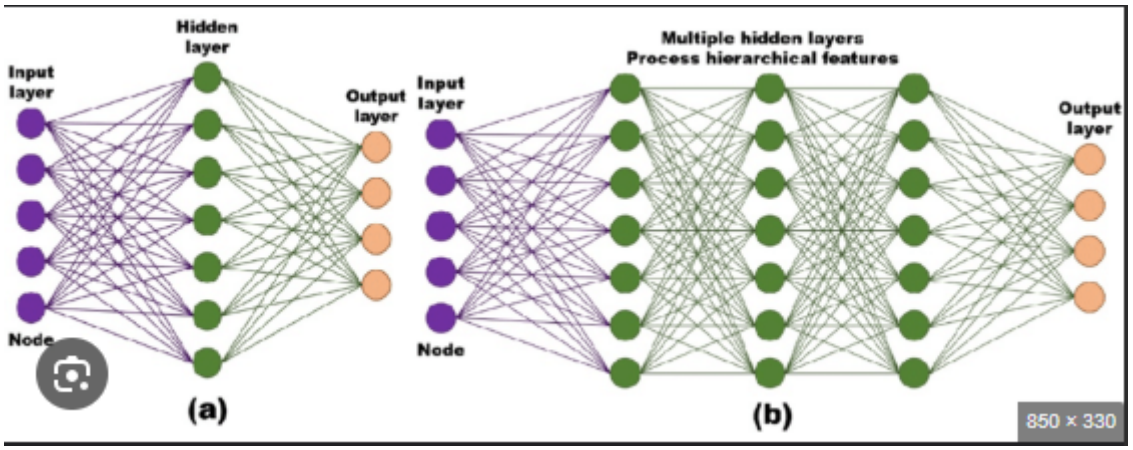
Truth Table - XOR		
X1	X2	Output Class
0	0	0
0	1	1
1	0	1
1	1	0

While a single perceptron cannot do the job, a network of perceptrons can accomplish this task. One Perceptron can be used to implement an OR Gate, while another perceptron can be used to implement a NAND Gate. The respective outputs can then be sent to another Perceptron as inputs to implement an AND Gate thereby giving us the XOR Output.

The XOR problem is a classic example in neural network theory that shows the limitations of shallow networks.

Shallow Networks: consists of one hidden layer (or none) apart from the input and output layers. A simple feedforward neural network with one hidden layer is considered a shallow network.

Deep Networks: has multiple hidden layers, which allows it to learn more complex patterns and hierarchical representations.



In XOR problem:

- A simple single-layer perceptron (which has no hidden layers) cannot solve the XOR problem because it only works for linearly separable problems.
- Even a single hidden layer with a limited number of neurons and linear activation functions will struggle.
- `A deep neural network (or a shallow network with at least one hidden layer using nonlinear activation functions) can learn XOR.
- The hidden layer allows the model to transform the input space into a new representation where XOR becomes linearly separable.
- Activation functions like ReLU, sigmoid, or tanh help introduce non-linearity.