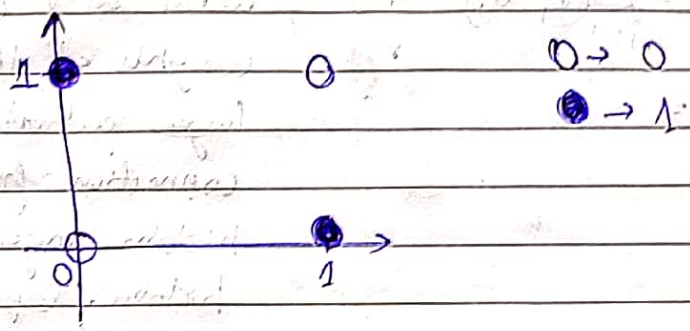
Reshabh Kumar Rajan 2017086

CLASSTEAM® Page No.
Date / /

ML Assignment - 3

3. No. We cannot model the XOR truth table using any number of linear activation functions. This is because, XOR data samples are not linearly-separable :-



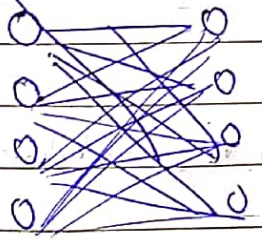
Let us take multiple layers with only linear activation function as the activation function. For the first layer :-

Output = $W^T X + b$, X ~~are~~ is the input vector.

~~For the next layer (first hidden layer) :-~~

~~output = $\sum_{j=1}^{n} w_j (w^T x + b)$~~

Input



Like this, we will have multiple outputs from the first hidden layer like (PTO) :-

CLASSTEAM | Page No.
Date / /

$O_i = W_i^T X + b_i$

In the second hidden layer, the output will be

$O = \sum_{i=1}^{n} (V_i O_i) + b$, where V are the weights of the connection from the hidden layer 1 to hidden layer 2.

$O = V_1 (W_1^T X + b_1) + V_2 (W_2^T X + b_2) + \ldots V_n (W_n^T X + b_n) + b$

$= (V_1 W_1^T + V_2 W_2^T + \ldots V_n W_n^T) X + (V_1 b_1 + V_2 b_2 + \ldots V_n b_n + b)$

Here, we can express weights as:-

$(V_1 W_1 + V_2 W_2 \ldots V_n W_n)$

And the new bias as : $V_1 b_1 + V_2 b_2 + \ldots V_n b_n + b$.

Hence, we still got a linear classifier for a data that is not linearly separable.

4. These are the components of the deep CNN :-

i) Convolution layer : This layer works on the input data to extract various features by using Kernels. This Kernel is chosen in such a way that it extracts the relevant feature (such as a diagonal line) from the input image. Many such Kernels can be used to generate various types of feature matrix, the size of which depends on the Kernel.

ii) Activation layer : Generally ReLu activation function is used because it preserves the value (magnitude) of the features.

iii) Pooling layer : Generally max (average) pooling are used. This uses a kernel with a defined size & stride. Here, max-pooling selects the maximum value from the output, hence selecting only relevant outputs & thereby reducing the dimensions of the input. We use padding for stride.

iv) Fully connected layer : It is like MLP. (Multi layer perceptron). In this layer, every node on the left is connected to every node in the right.

CLASSTEAM | Page No.
Date / /

Deep CNN is better than traditional neural network techniques because unlike in MLP, the neurons in 2 layers are not fully connected, but sparse. Also, the outputs are affected only by a small neighbourhood of the inputs, by the introduction of kernels.

In MLP, fully connected layers may introduce some redundancy.