XOR is not linearly separable

Proof. If XOR is linearly separable, a perceptron can separate it. We will construct a two input, one output perceptron to compute XOR. We will show that this leads to a contradiction, and so our assumption that XOR is not linearly separable must be false.

Our perceptron's inputs are x_1 and x_2 and have associated weights w_1 and w_2 respectively. Then we have $y = w \cdot x + b$ and the output of the perceptron is

$$z = \begin{cases} 1 & \text{if } y = w \cdot x + b > 0, \\ 0 & \text{otherwise} \end{cases}$$

Our training data is $D = \{(0,0), (0,1), (1,0), (1,1)\}$ and the desired output values are $d = \{0, 1, 1, 0\}$.

Because I am lazy, I'd rather not learn the bias independently, so let $x_3 \equiv 1$ and when we set w_1 and w_2 to random values in $[0, 1]$ to initialise learning, we also set w_3 to a similar random value. (We use random values because numerical analysts tell us that this provides better numerical stability.)

If we now agree that our vectors are now three-dimensional without changing their names, the perceptron's output may be written

$$z = \begin{cases} 1 & \text{if } y = w \cdot x > 0, \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

We must now write $D = \{(0,0,1), (0,1,1), (1,0,1), (1,1,1)\}$.

Suppose that XOR is linearly separable. Then when we learn the final weights for our perceptron, we will have succeeded in separating the elements of D according to the values of d. Then we may write the four training cases of our simple perceptron (one line for each element of D and d respectively) thus:

$$\begin{aligned} z((0,0,1)) &= 0 \\ z((0,1,1)) &= 1 \\ z((1,0,1)) &= 1 \\ z((1,1,1)) &= 0 \end{aligned}$$

In other words, using (1), we can express the same system in terms of $w_1x_1 + w_2x_2 + w_3x_3$, and so we have

$$\begin{aligned} w_1 0 + w_2 0 + w_3 1 &\leqslant 0 \\ w_1 0 + w_2 1 + w_3 1 &> 0 \\ w_1 1 + w_2 0 + w_3 1 &> 0 \\ w_1 1 + w_2 1 + w_3 1 &\leqslant 0 \end{aligned}$$

Dropping the zero terms and remembering the multiplicative identity, this yields

$$w_3 \leqslant 0 \tag{2}$$
$$w_2 + w_3 > 0 \tag{3}$$
$$w_1 + w_3 > 0 \tag{4}$$
$$w_1 + w_2 + w_3 \leqslant 0 \tag{5}$$

Summing (3) and (4), we see that $(w_1 + w_2 + w_3) + w_3 > 0$. Since $w_3 \leqslant 0$ by (2), $-w_3 \geqslant 0$, this is thus equivalent to

$$(w_1 + w_2 + w_3) > -w_3 \geqslant 0$$

We see then that (5) yields a contradiction, and so XOR is not linearly separable.

$\square$