Question 3

99222 - Frederico Silva, 99326 - Sebastião Carvalho

November 27, 2023

Question 3

In this exercise, you will design a multilayer perceptron to compute a Boolean function of D variables, $f : \{-1, +1\}^D \to \{-1, +1\}$, defined as:

(a) Perceptron's Limitations (5 points)

Show that the function above cannot generally be computed with a single perceptron. *Hint: think of a simple counter-example.*

Answer To demonstrate that the specified Boolean function cannot be computed by a single perceptron, we simply need to show that there exists a counter-example where the data is not linearly separable.

Let's consider a simple case where $D = 2$, $A = -1$, and $B = 1$. The function f is defined as:

$$f(x) = \begin{cases} 1 & \text{if } \sum_{i=1}^{D} x_i \in [-1, 1], \\ -1 & \text{otherwise} \end{cases}$$

Now, let's consider the following inputs: $x_1 = (+1, +1)$, $x_2 = (-1, -1)$, $x_3 = (-1, +1)$, and $x_4 = (+1, -1)$.

In this setup:

- For x_1, the sum $\sum x_i = 2$. Since 2 is not in the range [-1, 1], $f(x) = -1$.
- For x_2, the sum $\sum x_i = -2$. Since -2 is also not in the range [-1, 1], $f(x) = -1$.
- For x_3 and x_4, the sum $\sum x_i = 0$. This falls within the range [-1, 1], so $f(x) = 1$ for these inputs.

The visual representation of the points can be seen in Figure **??**. The red points represent the inputs that should be classified as $+1$ and the blue points represent the inputs that should be classified as -1.

The critical point here is that a single perceptron is fundamentally a linear classifier, which means it can only separate data points using a straight line in the feature space. However, in this example, there is no straight line that can separate these points accordingly in a 2D space to satisfy the function f.

This example thus serves as a counter-example proving that the given function cannot generally be computed with a single perceptron, as it requires a non-linear decision boundary which a single perceptron cannot provide.

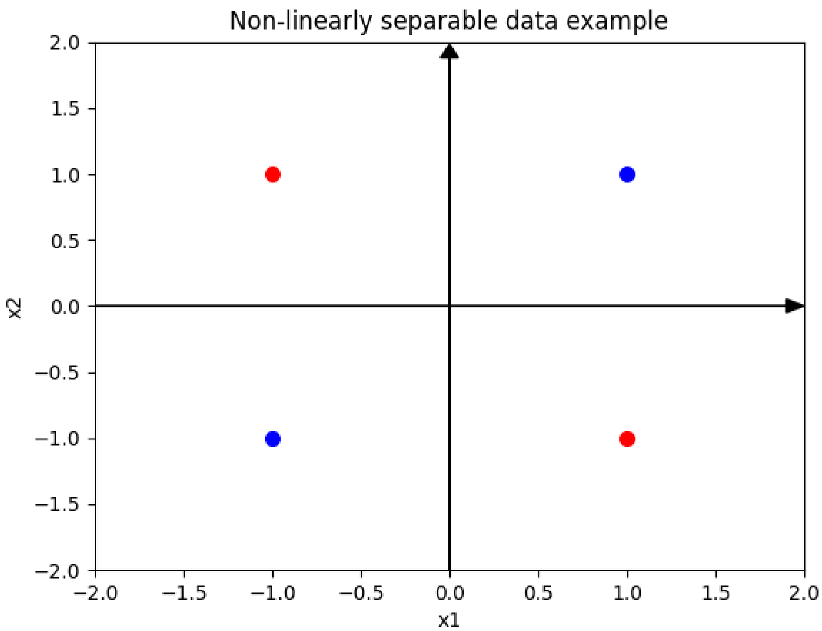

Figure 1: Classification of points using the function f

(b) Neural Network With Hard Threshold Activation (15 points)

Answer First we will start by defining the weights and biases of the network.

$W^{(1)} = \begin{bmatrix} 1 & \dots & 1 \\ -1 & \dots & -1 \end{bmatrix}$, where $W^{(1)}$ is a matrix of size 2 x D, and D is the size of the input vector.

$b^{(1)} = \begin{bmatrix} -A \\ B \end{bmatrix}$, where A is the lower bound of the sum of the input vector, and B is the upper bound of the sum of the input vector.

$W^{(2)} = \begin{bmatrix} 1 & 1 \end{bmatrix}$.
$b^{(2)} = \begin{bmatrix} -1 \end{bmatrix}$.