Neural Networks

Cracow University of Technology

Name: Madiyar Kaliyev

Homework 1

November 11, 2023

Problem 1

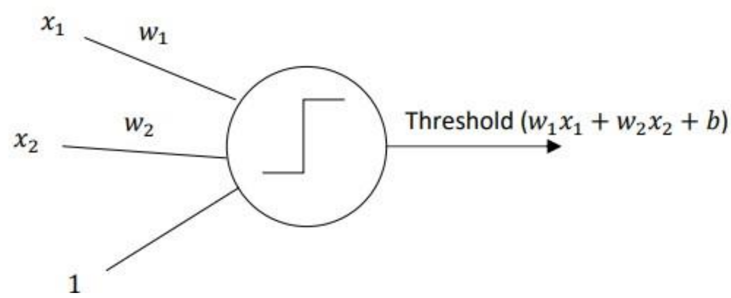
For the neuron activation function f, we can consider many possible choices. What happens if f is the identity function? Explain your answer mathematically.

Answer:

When the activation function is the identity function **(f(x) = x)**, the output of the neuron matches with its input, causing the neural network to behave like a linear model. This linearity limits the network's ability to learn complex patterns in the data. Nonlinear activation functions, such as sigmoid or ReLU, are important because they introduce nonlinearity. This nonlinearity allows the neural network to capture and represent more complex relationships and patterns in the input data, providing better learning and performance.

Problem 2

For the following neural network, find appropriate weights in order to represent the OR function:



Answer:

These weights and bias are used to represent OR function:

w1=1

w2=1

b=-1

The output will be 1 if one of the inputs X1 or X2 is 1, or if both inputs are 1. Otherwise, the output will be 0.

Here is the demonstration:

X1	X2	w1*x1 + w2*x2 + b	Output
0	0	-1	0
0	1	0	1
1	0	0	1
1	1	1	1

The results confirm the correctness of the given parameters

Problem 3

We have a simple neural network with an input layer, a hidden layer, and an output layer. For this neural network, we represent the weight matrices by $w^{(1)}$and $w^{(2)}$ their corresponding biases by $b^{(1)}$ and $b^{(2)}$. We also use w_{ij} to represent the weight between the ith node and the jth node in the previous layer and the current layer repsectively. For both layers, the activation function is the sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$. The outputs of the two layers before applying the activation function are $z^{(1)}$ and $z^{(2)}$, and after applying the activation function re $a^{(1)} = \sigma(z^{(1)})$ and $a^{(2)} = \sigma(z^{(2)})$. Assuming:

$$w^{(1)} = \begin{bmatrix} 0.4 & 0.6 & 0.2 \\ 0.3 & 0.9 & 0.5 \end{bmatrix}, \ b^{(1)} = [1, 1, 1]; \ \text{ and}$$

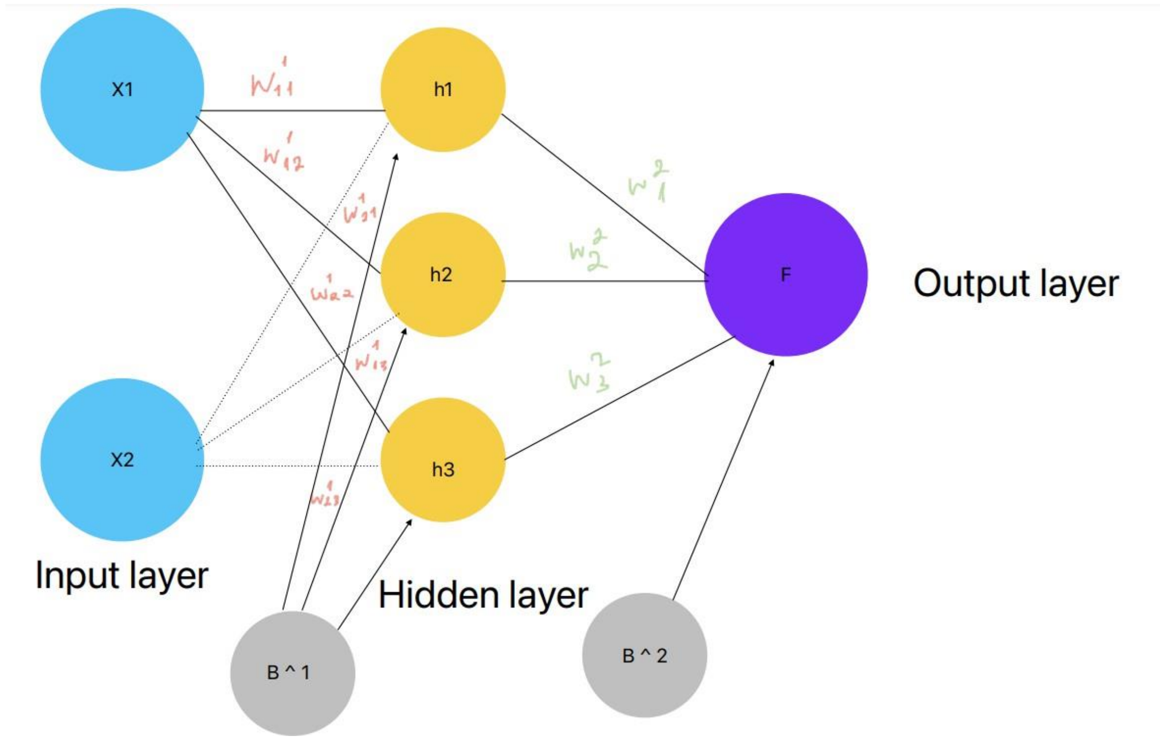
$$w^{(2)} = \begin{bmatrix} 0.2 \\ 0.2 \\ 0.8 \end{bmatrix}, \ b^{(2)} = [0.5].$$

I. Draw the diagram of the network.

II. If the input is $a^{(0)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, what will be the network output and what is the output of the loss? Show your calculations step by step.

Answer:

I. Diagram:



II. If the input is a (0) = [1 1], what will be the network output and what is the output of the loss? Show your calculations step by step.

The calculation of h1, h2 and h3 is needed to get a^2. An input $a^0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, the following calculation is done:

$h_1 = f\left([1 \;\; 1] \cdot \begin{bmatrix} 0.4 \\ 0.3 \end{bmatrix} + 1 \right) = f(1 + 0.4 + 0.3) = f(1.7) = 0.85$

$h_2 = f\left([1 \;\; 1] \cdot \begin{bmatrix} 0.6 \\ 0.9 \end{bmatrix} + 1 \right) = f(1 + 0.6 + 0.9) = f(2.5) = 0.93$

$h_3 = f\left([1 \;\; 1] \cdot \begin{bmatrix} 0.2 \\ 0.5 \end{bmatrix} + 1 \right) = f(1 + 0.2 + 0.5) = f(1.7) = 0.85$

The result is: $a^1 = [0{,}85,\ 0.93,\ 0.85]$

Next, it's possible to calculate a2 that equals:

$a^2 = f\left([0.85, 0.93, 0.85] \cdot \begin{bmatrix} 0.2 \\ 0.2 \\ 0.3 \end{bmatrix} + 0.5 \right) =$

$= f(0.5 + 0.85 \cdot 0.2 + 0.92 \cdot 0.2 + 0.85 \cdot 0.8) = f(1.534) = 0.823$

Network's output is 0.823.