Homework 1

Neural Networks

Pedro Nunes, PK - pedronunes0028@gmail.com

October 2023

1 Problem 1

The activation function is a function used in neural networks to calculate the output of a node, based of its weights and inputs. It is particularly useful to solve problems where the solution is nonlinear.

A neuron usually follows the following formula

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right)$$

When we use the identity function, we're using a linear function.

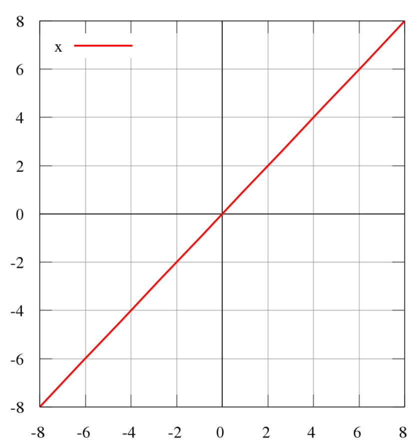

Figure 1: Linear Function

Mathematically, the function is represented by

$$f(x) = x$$

If we replace it in the neuron formula, we get

$$y_{\text{linear}} = \sum_{i=1}^{n} w_i x_i + b$$

This means that the function only scales the result based on the slope. So, if the slope is equal to 1, then our function isn't doing anything, it's just outputting the same value it got. Notably, if it's different then 1, it will scale the output according to the factor.

2 Problem 2

Let us start by defining the OR function for each pair of inputs

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

Our neural network should follow the same behaviour. We're using a threshold function, where our threshold is equal to 0, following the formula:

$$f(x) = \begin{cases} 0 & \text{if } x \leq \theta \\ 1 & \text{if } x > \theta \end{cases}$$

When applied to our neuron formula, we get the following

$$y = \begin{cases} 0 & \text{if } \sum_{i=1}^{n} w_i x_i + b \leq \theta \\ 1 & \text{if } \sum_{i=1}^{n} w_i x_i + b > \theta \end{cases}$$

Since we want to get 1 every time a number is different than 0, we can measure our weights as

$$\begin{aligned}
w_1 &= 1 \\
w_2 &= 1
\end{aligned}$$

If our threshold function requires the number to be greater or equal than 0 to give us a 1, we can add a negative bias, for instance -0.5 , so that we get a zero when both inputs are zero.

As an example, let's calculate the output for $\text{OR}(1,1)$:

$$y = f\left(\sum_{i=1}^{2} w_i x_i + b\right) = f(1+1) = f(2)$$

Since $2 > 0$, the output of $f(2)$ is 1, which means our model is working as intended

3 Problem 3

My purposed diagram for the network is

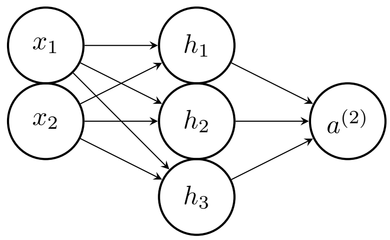

Figure 2: Neural Network Architecture

In the neural network architecture shown in Figure 2, all neurons are represented by the following formula

$$f(W_{i,j} \cdot x + b_{i,j})$$

where:

$$\begin{aligned}&i \text{ is input neuron index,}\\&j \text{ is the hidden neuron index,}\\&x \text{ is the input value,}\\&f \text{ is the sigmoid function}\\&W_{i,j}\text{is the weight associated}\\&b_{i,j}\text{is the bias}\end{aligned}$$

To get $a^{(2)}$, we first need to calculate h_1, h_2, and h_3. Given an input $a^{(0)} = \begin{bmatrix}1\\1\end{bmatrix}$, we perform the following calculations:

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

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

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

So, we have that

$$a^{(1)} = [0.85, 0.93, 0.85]$$

We can then calculate a2, which is equal to

$$a^{(2)} = f\left(\begin{bmatrix}0.85 & 0.93 & 0.85\end{bmatrix} \cdot \begin{bmatrix}0.2\\0.2\\0.8\end{bmatrix} + 0.5\right)$$

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

Our network output is 0.823.