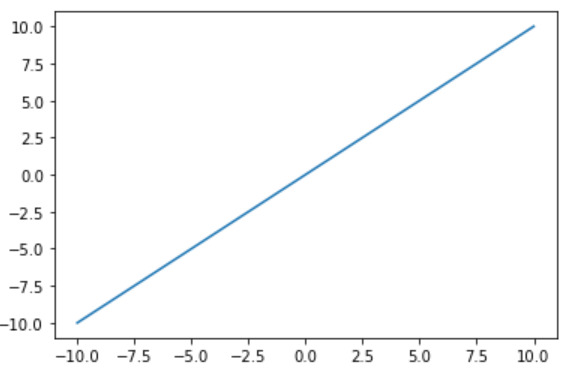
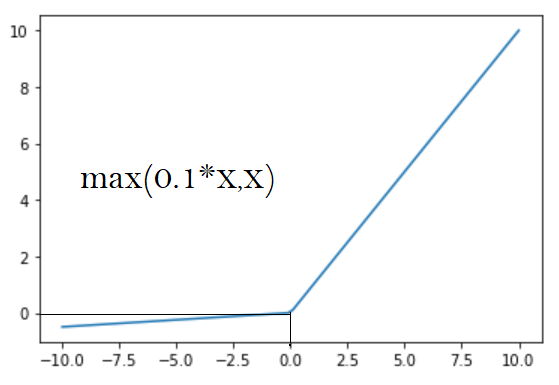
Summarizing Types of Activation Function - Part 2

1. **Linear activation function**: The linear activation function is one in which the activation is proportional to the input. This function does nothing to the weighted sum of the input and returns the value it was given.



The activation function of the last layer is merely a linear function of the input from the first layer, regardless of how many layers there are, assuming they are all linear. The linear activation function has a range of -inf to +inf. The neural network's last layer will operate as a linear function of the first layer.

2. **Leaky ReLU**: The Leaky ReLU function is an improved version of the ReLU activation function. The gradient of the ReLU activation function is 0 for all input values less than zero, which deactivates the neurons in that region and may cause the dying ReLU problem.

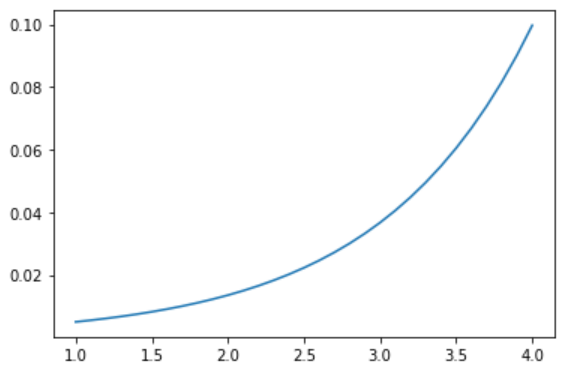


Leaky ReLU is defined to address this problem. Instead of defining the ReLU activation function as 0 for negative values of inputs (x), we define it as an extremely small linear component of x. Here is the formula for this activation function

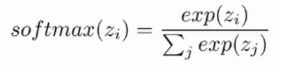
f(x)=max(0.01\*x , x)

If the above function receives any positive input, it returns x; otherwise, it returns a small value equal to 0.01 times x. As a result, it produces an output for any negative values. By making this small change, the gradient on the left side of the above graph becomes non-zero. As a result, it would no longer encounter dead neurons in that area.

3. **Softmax Activation**: The softmax function is often described as a combination of multiple sigmoids. The sigmoid activation function returns values between 0 and 1, which are the probabilities of each of the data points belonging to a particular class. Thus, sigmoid is widely used for binary classification problems.

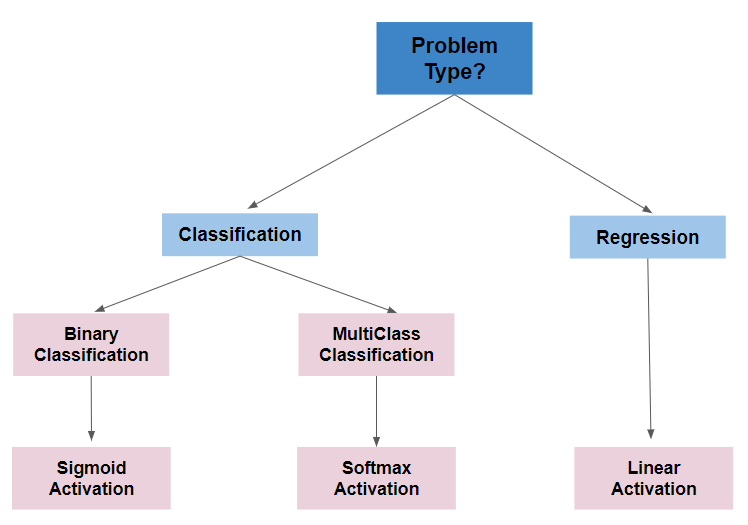


But the softmax activation function is used in the output layer of multi-class classification problems, where this function returns the probability for a datapoint belonging to each individual class. The range of softmax is -inf to inf. Below is the mathematical expression for softmax.



The function normalizes the outputs for each class between 0 and 1 and adds the normalized sum for each class. The output for a particular class is its normalized output divided by the total normalized sum.

**Choosing the right activation function in the output layer**:



Happy Learning!!