**Vanishing**

As the backpropagation algorithm advances downwards(or backward) from the output layer towards the input layer,

the gradients often get smaller and smaller and approach zero which eventually leaves the weights of the initial or lower layers nearly unchanged.

As a result, the gradient descent never converges to the optimum. This is known as the **vanishing gradients** problem.

#### Exploding

The gradients keep on getting larger and larger as the backpropagation algorithm progresses.

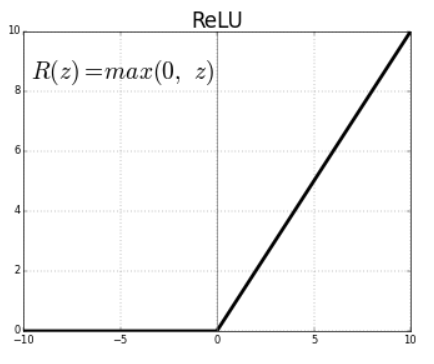
This, in turn, causes very large weight updates and causes the gradient descent to diverge. This is known as the **exploding gradients** problem.

##### **How Can The Exploding Gradient Problem Be Prevented?**

### 1. Proper Weight Initialization

### 2. Using Non-saturating Activation Functions

## Using ReLU ( Rectified Linear Unit )



* *Relu(z) = max(0,z)*
* Outputs 0 for any negative input.
* Range: [0, infinity]

### 3. Batch Normalization

Normalization is a data pre-processing tool used to bring the numerical data to a common scale without distorting its shape.

it is a process to make neural networks faster and more stable through adding extra layers in a deep neural network.

The new layer performs the standardizing and normalizing operations on the input of a layer coming from a previous layer.

**Batch-** collected set of input data