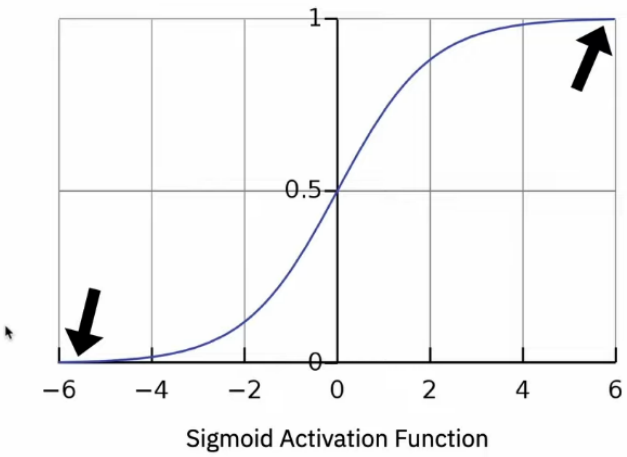
How to choose activation function:

➢ To get rid of unstable gradients we should use Non-saturating activation functions.

➢ Saturating activation functions are those that flatten out and cause vanishing gradients for large input values, hampering training in deep neural networks.

For eg:- sigmoid activation function

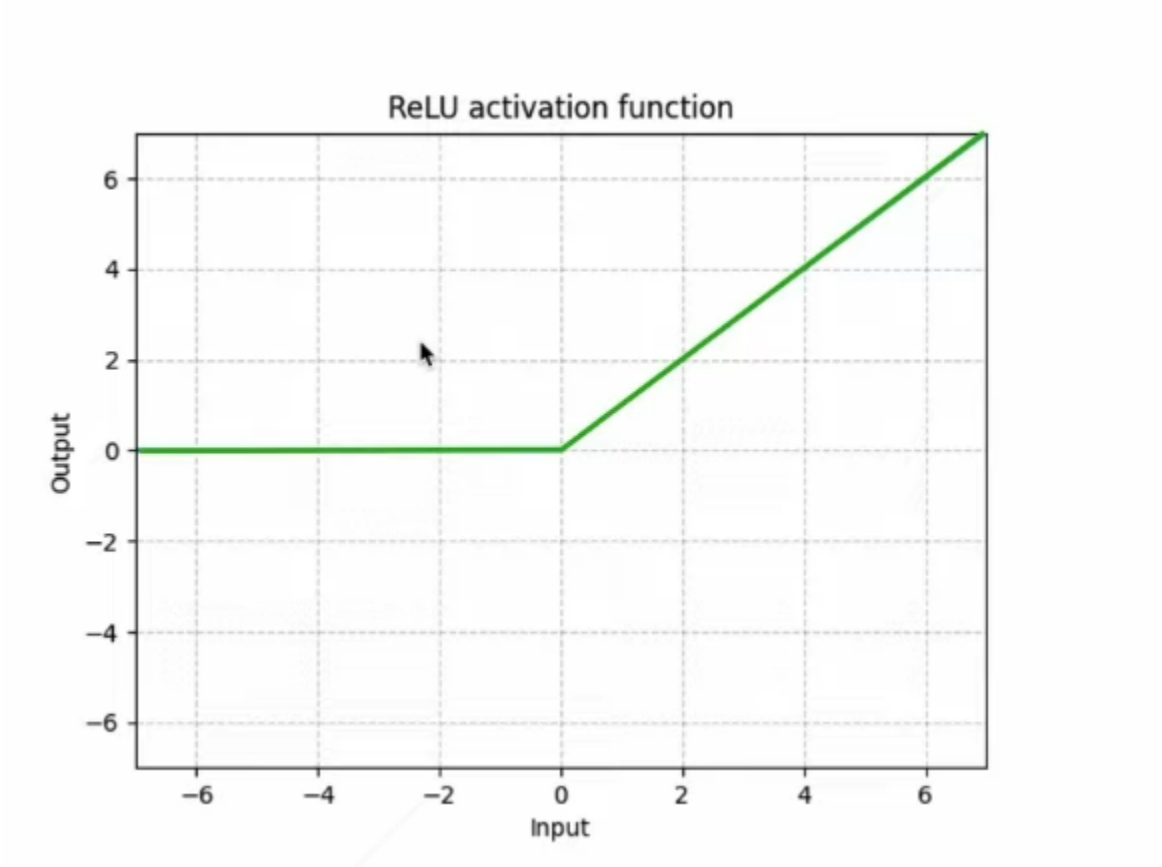
- Saturation on extremes causes the unstable gradients



Sigmoid Activation Function

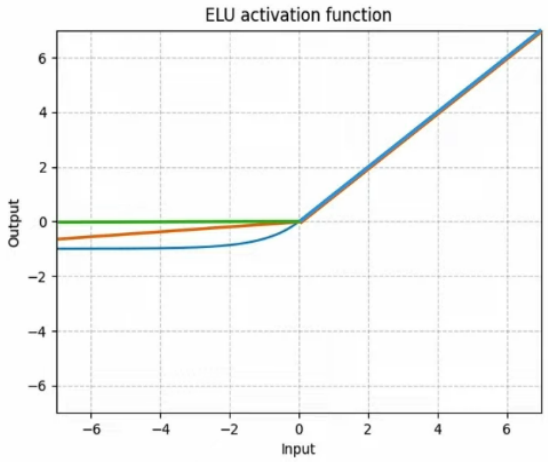
➢(Saturating activation function vannale ,mathi ko sigmoid activation function ko graph herda , yedi hami sanga 6 vanda badhi number input value xa vane tyo automatically 1 (flatten)bhaidinxa)

➢ Eg of non saturating activation function is Rectified Linear Unit (RELU)function,and its graph is as below:



➢ Following are the non saturating activation functions:-

- Rectified Linear Unit (ReLU)
- Leaky ReLU (with parameter α)
- Randomized Leaky ReLU (RReLU)
- Parametric Leaky ReLU (PReLU)
- Exponential Linear Unit (ELU) (with parameter α)
- Scaled ELU
- Gaussian Error Linear Units (GELU)
- ...



Green = ReLU
Orange = Leaky ReLU(with parameter a)
Blue = Exponential Linear Unit(ELU) (with parameter a)