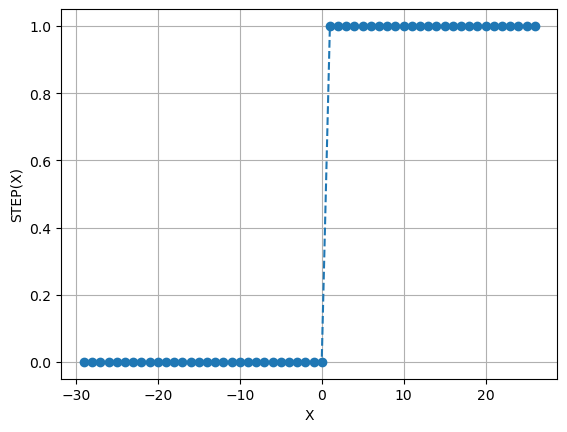
**Step Function:** A step function is a mathematical function that has a constant value within certain intervals and changes abruptly between those intervals. It is also known as a piecewise constant function or a staircase function.

Advantages:

* Requires only a few parameters to define the intervals and the constant values within those intervals.
* Step functions are useful for modeling processes that are discrete, such as counting or threshold-based decisions.
* Can be used for approximations: Step functions can be used to approximate more complex functions by breaking them down into smaller intervals. This can simplify calculations and make them more manageable.

Disadvantage:

* Step functions are discontinuous and not smooth, which can be a problem for some applications.
* Step functions are not suitable for modeling continuous processes or events that change gradually over time.
* Not precise: Step functions can be imprecise for certain applications because the intervals between the constant values are finite.



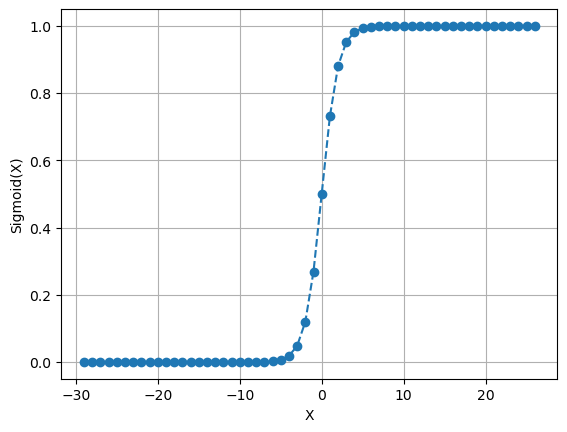
**Sigmoid Function:** This is a function that maps any input value to a value between 0 and 1. It is commonly used for (machine learning & Neural Networks) models where we have to predict the probability as an output. The function represents the shape of “S”.

Advantages:

* Well suited for binary classification problems cause of output can be 0 or 1.
* Nice interpretation as a saturating “firing rate” of a neuron
* Output of the sigmoid function can be interpreted as a probability, which makes it useful for classification tasks

Disadvantage:

* The sigmoid function can suffer from the problem of vanishing/ kill the gradients
* Output are not being zero centered
* Function might not perform well if its input values are very large or very small



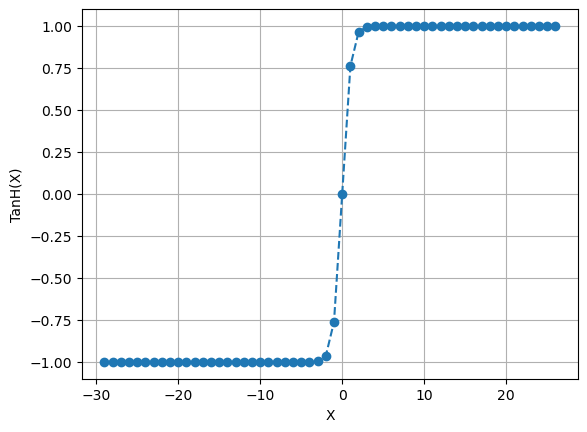
**Tanh Function:** This is a function that maps any input value to a value between -1 and 1. Unlike Sigmoid function this function output shows zero-centered.

Advantages:

* As centered around 0, which means that it can model positive and negative inputs more effectively
* The tanh neuron is simply a scaled sigmoid neuron. The function that is differentiable at all points, making it easy to use in optimization algorithms

Disadvantage:

* Like the sigmoid function, the tanh function can suffer from the problem of vanishing gradients, where the gradient of the function becomes very small as the input approaches -1 or 1.



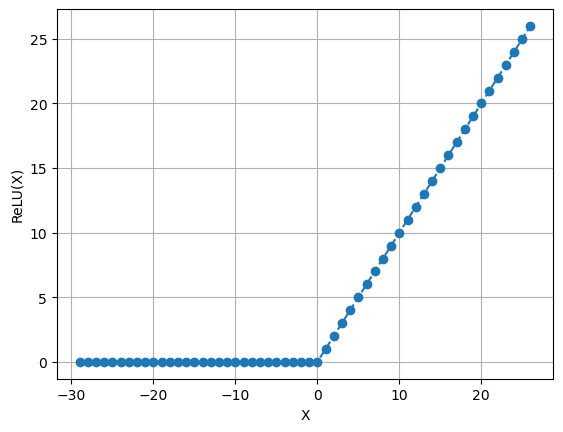
**ReLU:** The Rectified Linear Unit (ReLU) function is a mathematical function that maps any input value to a value between 0 and infinity. It is widely used in the deep learning sector.

Advantages:

* It does not suffer from the vanishing gradient problem that affects the sigmoid and tanh functions. Due to its linear, non-saturating form.
* The tanh neuron is simply a scaled sigmoid neuron. The function that is differentiable at all points, making it easy to use in optimization algorithms
* Computes **f(x) = max(0,x)**
* Converges much faster than sigmoid/tanh in practice

Disadvantage:

* Not being zero-centered output, which can make it difficult to use in some optimization algorithms
* Any negative input given can turn into zero immediately in the graph, which makes the mapping for negative values not a good look.



**ELU:** The Exponential Linear Unit (ELU) function is a mathematical function that maps any input value to a value between negative infinity and infinity. Computes **f(x) = max(0,x)**

Advantages:

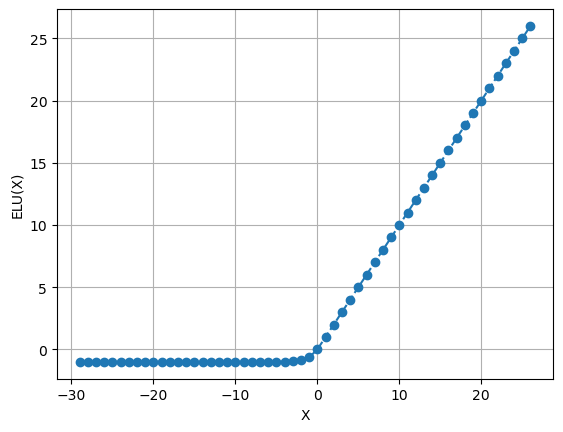
* Has the benefits of ReLU

Produces outputs that are centered around 0

* Has negative saturation regime compared which means that it can help reduce the effects of noises

Disadvantage:

* Computationally expensive than some other activation functions, such as the ReLU function



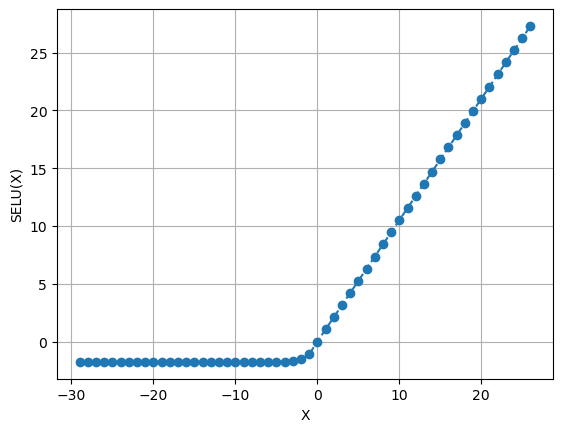
**SeLU:** The Scaled Exponential Linear Unit (SeLU) function is a mathematical function that is similar to the ELU function but includes a scaling parameter that allows it to self-normalize.

Advantages:

* The function has a self-normalizing property, which means that it can help prevent the vanishing gradient problem that affects the sigmoid and tanh functions
* The SeLU function has a negative saturation regime, which means that it can help reduce the effects of noisy or outlier data
* Has negative saturation regime compared which means that it can help reduce the effects of noises

Disadvantage:

* The ELU function is sensitive to the scale of the input data, which means that it may not perform well if the input values are very large or very small.

****