Step Function:

A binary function, the step activation function can either return 0 or 1, depending on the input. It is frequently applied to binary classification issues. The step activation function's key benefit is that it is computationally effective and yields findings that are easy to understand. It cannot be utilized for more sophisticated jobs and can only be applied to binary classification problems. The step function's zero gradient is the primary issue. Because of this, the step function is less useful during back-propagation. The threshold value has the same impact on a value that is 10000000 times that value. also disregards input values that are just below the threshold. Moreover, multi-class classification cannot be done with it.

Sigmoid Function:

A non-linear function, the sigmoid function produces values between 0 and 1. Given that it offers a smooth transition from 0 to 1, it is appropriate for both classification and regression applications. It returns the probability output for each class in multiclass categorization. Another benefit of this function is that it returns a value between (-infinity, +infinity) when utilized as in the linear function (0, 1). The vanishing gradient issue is one of the flaws of this function. Incase This function is not zero-centered; a zero-centered function has zero in the middle of the function range. With zero centers, the majority of machine learning algorithms perform better. Also, the cost of computation is high (exponential in nature).

Tanh Function:

The Tanh activation functions are likewise a sort of continuous activation function similar to Sigmoid but are symmetric over the origin; also, the output values vary between -1 and 1. The most frequent application of this kind of activation function is in categorization issues. Tanh will provide outputs with zero centers. The issue of the values having the same sign is also resolved. Due to the non-linear structure of this function, errors can be easily propagated backwards. The sigmoid function still has the vanishing gradient issue, which is a comparable downside. That costs a lot to compute (exponential in nature).

ReLU Function:

ReLU activation functions are a form of continuous activation function that, depending on whether the input is larger than or less than 0, returns either 0 or the input value. ReLU's non-linear nature makes it simple to backpropagate errors and utilize it to activate neurons in a variety of layers. In addition, when compared to the sigmoid and tanh activations, this function speeds up the convergence of stochastic gradient descent. Because of its computational effectiveness, the network can converge relatively quickly. It has the ability to provide a real zero value. ReLU output is not zero centered, which reduces the neural network's effectiveness. The gradients of the weights will either be uniformly positive or uniformly negative during backpropagation. When the neuron becomes stuck on the negative side and consistently outputs zero, there is a dyi ng ReLU problem. This occurs either when there is a substantial negative bias or when learning r ate is high.

ELU Function:

Similar to the ReLU function, the ELU function produces a negative value for negative inputs. because it guarantees that neurons are active despite negative inputs. The issue of disappearing gradients and expanding gradients is not a concern at ELU. ELU does not experience the issues associated with decaying neurons like ReLU does. In comparison to ReLU and its derivatives, using ELU results in a shorter training time and greater accuracy in neural networks. At any location, the ELU activation function is continuous and differentiable. The possibility of negative ELU values causes the mean of the activations to be pushed closer to zero. Moreover, learning and convergence occur more quickly when mean activations are closer to zero. Due to its non-linearity with negative numbers, ELU takes longer to calculate.

SELU Function:

The activation functions known as SELUs, or Scaled Exponential Linear Units, cause self-normalization. Neuronal activations in the SELU network automatically converge to zero mean and unit variance. The output of the entire layer will remain constant since this function, while comparable to the ELU function, is scaled. As a result, the output doesn't get too small, which helps to lessen the issue of vanishing gradients. SELUs cannot expire as ReLUs can. Compared to other activation functions, SELUs can learn more quickly and more effectively without additional processing. Since this activation feature is still quite new, it is not yet commonly utilized in daily life.