Writing a report About Every Function:

Stepping function:

The step function is a basic activation function used in machine learning that returns 1 if the input is greater than or equal to 0 and 0 otherwise. It is a simple threshold function that has both benefits and downsides.

Advantages:

Simplicity: The step function is a straightforward function that is simple to comprehend and apply in a neural network.

Binary Output: The step function generates a binary output, which may be beneficial in certain applications like binary classification issues.

Computationally efficient: The step function is computationally efficient since it takes relatively little processing.

Disadvantages:

Non-differentiability: Since the step function is not differentiable, it cannot be employed in backpropagation methods that require derivative computation.

Vanishing Gradient: The gradient of the step function is zero everywhere except at the point of discontinuity, which might pose issues when training deep neural networks. The vanishing gradient issue may stymie or perhaps prohibit neural network convergence.

Instability of output: If the function's input is noisy or has minor variations, the step function may create unstable output. This might cause the network's output to hop back and forth between 0 and 1, making it difficult to utilise in certain applications.

Ultimately, the step function is a basic activation function that may be beneficial in certain applications, but its drawbacks make it inappropriate for many contemporary deep learning systems. More complex activation functions, such as ReLU, ELU, and SELU, are chosen because to their differentiability and ability to avoid vanishing gradients.

sigmoid function:

The sigmoid function is a prominent activation function in machine learning that transfers every input value to a value between 0 and 1. It has both pros and cons.

Advantages:

Non-linearity: Since the sigmoid function is non-linear, it may be used to simulate non-linear data relationships.

Smoothness: The sigmoid function is a smooth function that is differentiable at all locations. This makes it handy in optimization procedures that involve the computation of derivatives, such as gradient descent.

The sigmoid function may be viewed as a probability distribution, making it helpful in probabilistic models and binary classification applications.

Disadvantages:

Vanishing Gradient: The sigmoid function has a modest gradient in the region when its output is near to 0 or 1. This might result in the vanishing gradient issue during backpropagation, when the gradient grows less and smaller as it propagates through the layers of a neural network, making deep neural networks difficult to train.

Output saturation: The sigmoid function may saturate when the input is extremely big or very little, resulting in gradients near to zero. Since the weights are no longer updated, the network may cease to learn.

The sigmoid function is not zero-centered, which makes training a neural network problematic since the mean of the activations in a layer may move away from zero.

Overall, the sigmoid function is a helpful activation function in certain contexts, but its drawbacks restrict its use in deep learning networks. Alternative activation functions, such as ReLU, ELU, and SELU, have been created to solve some of the sigmoid function's drawbacks.

Tanh function:

The tanh function is a prominent activation function in machine learning that translates every input value to a value between -1 and 1. It has both pros and cons.

Advantages:

Non-linearity: Since the tanh function is non-linear, it may be used to describe non-linear data interactions.

Smoothness: The tanh function is a smooth function that is differentiable at all locations. This makes it handy in optimization procedures that involve the computation of derivatives, such as gradient descent.

Zero-centered: The tanh function is zero-centered, which may aid in neural network training by preventing the mean of activations in a layer from drifting away from zero.

Disadvantages:

Vanishing Gradient: The tanh function has a modest gradient in the region when its result is near to -1 or 1. This might result in the vanishing gradient issue during backpropagation, when the gradient grows less and smaller as it propagates through the layers of a neural network, making deep neural networks difficult to train.

Output saturation: The tanh function may saturate when the input is extremely big or very little, resulting in gradients near to zero. Since the weights are no longer updated, the network may cease to learn.

Computationally expensive: The tanh function takes more computing than other activation functions, such as the ReLU function.

Overall, the tanh function is a helpful activation function in certain cases, but its drawbacks restrict its use in deep learning frameworks. Alternative activation functions, such as ReLU, ELU, and SELU, have been created to alleviate some of the shortcomings of the tanh function.

Relu function:

The Rectified Linear Unit (ReLU) function is a prominent activation function in machine learning that transfers any input value to a value between 0 and infinity. It has both pros and cons.

Advantages:

Non-linearity: Since the ReLU function is non-linear, it may be used to describe non-linear data relationships.

Sparsity: The ReLU function may provide sparse representations in which many of the activations in a layer are zero. This may assist to avoid overfitting and increase generalisation.

Computational efficiency: The ReLU function is computationally efficient since it just needs basic thresholding operations.

Vanishing Gradient: The ReLU function does not suffer from the vanishing gradient issue since the gradient is always 1 for inputs higher than 0. This simplifies the training of deep neural networks.

Disadvantages:

Dead neurons: The ReLU function may generate dead neurons, where the output is always 0. This may happen when the ReLU function's input is negative, and it can cause a huge percentage of the network to remain inactive, reducing the network's potential to represent complicated connections in data.

Non-zero mean: Since the ReLU function is not zero-centered, it might be challenging to train a neural network because the mean of the activations in a layer may move away from zero.

Overall, the ReLU function is a helpful activation function in many scenarios, although its drawbacks restrict its use in various deep learning systems. Alternative activation functions, such as Leaky ReLU, ELU, and SELU, have been created to alleviate some of the shortcomings of the ReLU function.

Selu function:

The Scaled Exponential Linear Unit (SELU) function is a relatively recent activation function used in machine learning that is intended to solve some of the limitations of older activation functions. It has both pros and cons.

Advantages:

Self-normalization: The SELU function is meant to be self-normalizing, which means that the mean and variance of the output of each layer in a neural network will remain constant regardless of network depth. This may help lessen the possibility of disappearing or exploding gradients, which can make training deep neural networks problematic.

Non-linearity: Since the SELU function is non-linear, it may be used to simulate non-linear data interactions.

Continuous differentiability: The SELU function is a continuous function that is differentiable at all places. This makes it handy in optimization procedures that involve the computation of derivatives, such as gradient descent.

Zero-centered: The SELU function is zero-centered, which may aid in neural network training by preventing the mean of activations in a layer from migrating away from zero.

Disadvantages:

Initialization sensitivity: The SELU function is sensitive to the initialization of weights and biases in a neural network. This implies that careful setup is essential to get the full advantages of the SELU function.

Restricted applicability: The SELU function is intended to perform best with feedforward neural networks that employ fully-connected layers with identical input and output dimensions. It may not operate as effectively in other kinds of neural networks or with various layer configurations.

Overall, the SELU function is a promising activation function that has shown strong performance in various deep learning applications. Nevertheless, its sensitivity to initialization and restricted application may limit its effectiveness in certain scenarios.

Elu function:

The Exponential Linear Unit (ELU) function is a machine learning activation function that is comparable to the ReLU function. It has both pros and cons.

Advantages:

Non-linearity: Since the ELU function is non-linear, it may be used to describe non-linear data relationships.

Smoothness: The ELU function is a smooth function that is differentiable at all locations. This makes it handy in optimization procedures that involve the computation of derivatives, such as gradient descent.

Negative values: The ELU function allows for negative values, which may assist avoid dead neurons and increase the network's ability to understand complicated connections in data.

Zero-centered: The ELU function is nearly zero-centered, which may aid in neural network training by avoiding the mean of activations in a layer from migrating away from zero.

Disadvantages:

Computational complexity: The ELU function is more computationally costly than the ReLU function since it needs the calculation of exponential functions.

Responsive to hyperparameters: The ELU function is sensitive to the selection of hyperparameters, such as the alpha parameter, which sets the function's negative saturation value. Improper hyperparameter selection may result in poor neural network performance.

Overall, the ELU function is a useful activation function that has certain benefits over the ReLU function, but its computational complexity and hyperparameter sensitivity may restrict its applicability in some scenarios.