Writing reports Functionality:

step:

The step function is a basic activation function used in machine learning that outputs 1 if the input is greater than or equal to 0 and 0 otherwise. Easy to install, it has pros and cons.

Advantages:

- Simplicity: The neural network step function is straightforward and easy to grasp.
- Binary Output: In binary classification issues, the step function's binary output might be beneficial.
- The step function is computationally efficient.

Disadvantages:

- The neural network's convergence may be slowed by the vanishing gradient issue.
- Output instability: Noisy or tiny changes in the step function's input might cause output instability. In certain applications, this might cause the network's output to bounce between 0 and 1.
- The step function is a basic activation function, but its drawbacks make it inappropriate for many current deep learning systems. ReLU, ELU, and SELU are chosen because of their differentiability and ability to avoid vanishing gradients.

Sigmoid Function:

The sigmoid function, a prominent activation function in machine learning, transfers every input value to a number between 0 and 1. ,

Advantages:

- Non-linearity: Since the sigmoid function is non-linear, it may simulate non-linear data relationships.
- Smoothness: The sigmoid function is differentiable everywhere and smooth.
- Probabilistic interpretation: The sigmoid function may be used in probabilistic models and binary classification issues as a probability distribution.

Disadvantages:

- Vanishing Gradient: The sigmoid function's gradient at 0 or 1 is tiny. This causes the vanishing gradient issue during backpropagation, which makes deep neural networks hard to train.

- Output saturation: When the input is extremely big or very little, the sigmoid function may saturate and produce gradients near to zero. Weights aren't changed, therefore the network stops learning.

The sigmoid function may be beneficial in certain cases, but its drawbacks may restrict its utility in deep learning frameworks. ReLU, ELU, and SELU activation functions have been created to alleviate some of the sigmoid function's shortcomings.

Tanh function:

Machine learning's tanh function converts any input value to a value between -1 and 1. It has pros and downsides.

Advantages:

- Non-linearity: Since the tanh function is non-linear, it may simulate non-linear data relationships.
- Smoothness: All points of the tanh function are differentiable. Derivative-calculating optimization methods like gradient descent benefit from this.
- Zero-centered: The tanh function's zero-centeredness prevents a layer's mean activations from drifting away from zero, which aids neural network training.

Disadvantages:

- Vanishing Gradient: In the region where its output is close to -1 or 1, the tanh function has a modest gradient. This might produce the vanishing gradient issue during backpropagation, when the gradient becomes smaller as it propagates through a neural network's layers, making deep neural networks hard to train.
- Output saturation: The tanh function may saturate when the input is extremely big or very little, resulting in gradients near to zero. The network stops learning since the weights aren't updated.
- Computationally expensive: Unlike the ReLU function, the tanh function takes more computation.

The tanh function is a helpful activation function in certain cases, but its drawbacks may restrict its utility in deep learning frameworks. ReLU, ELU, and SELU activation functions alleviate some of the tanh function's drawbacks.

Relu Function:

The Rectified Linear Unit (ReLU) activation function transfers any input value to a value between 0 and infinity in machine learning. ,

Advantages:

- Non-linearity: The ReLU function is helpful for modeling data's non-linear relationships.
- Sparsity: The ReLU function creates sparse representations with multiple zero activations. This may increase generalization and decrease overfitting.
- ReLU's basic thresholding processes make it computationally efficient.
- Vanishing Gradient: As the gradient is always 1 for inputs higher than 0, the ReLU function doesn't have the vanishing gradient issue. This simplifies training deep neural networks.

Disadvantages:

- Dead neurons: The ReLU function may generate dead neurons with 0 output. .
- Non-zero mean: The ReLU function is not zero-centered, making it challenging to train a neural network since the mean of a layer's activations may move away from zero. To overcome ReLU's drawbacks, several activation functions such Leaky ReLU, ELU, and SELU have been created.

Selu Function:

Machine learning's Scaled Exponential Linear Unit (SELU) function is meant to overcome some of the limitations of previous activation functions. It's both good and bad.

Advantages:

- Self-normalization: The SELU function's mean and variance of each layer's output remain constant regardless of network depth. This may assist prevent disappearing or inflating gradients, which can make deep neural network training problematic.
- Non-linearity: The SELU function's non-linearity makes it suitable for simulating non-linear data interactions.
- Continuous differentiability: All points of the SELU function are differentiable. This makes it valuable in optimization methods that calculate derivatives, such as gradient descent.
- Zero-centered: The SELU function prevents the mean of a layer's activations from migrating away from zero, which helps train a neural network.

Disadvantages:

- Initialization sensitivity: The SELU function is sensitive to neural network weights and biases. To maximize SELU function advantages, precise setup is needed.
- Restricted applicability: The SELU function works best with feedforward neural networks with fully-connected layers and identical input and output dimensions. Other neural networks or layer combinations may not function as well.
- The SELU function is a potential activation function that performs well in various deep learning applications. Its startup sensitivity and restricted application may limit its usability in certain instances.

Elu Function:

It is similar to relu function . It has both advantages and disadvantages.

Advantages:

- Non-linearity: The ELU function may simulate non-linear data relationships due to its non-linearity.
- Smoothness: All points of the ELU function are differentiable. This may help gradient descent and other optimization methods calculate derivatives.
- Negative values: The ELU function enables negative values, which may prevent neurons from dying and boost the network's ability to simulate complicated data connections.
- Zero-centered: The ELU function prevents the mean of a layer's activations from migrating away from zero, which helps train a neural network.

Disadvantages:

- Computational complexity: Since it calculates exponential functions, the ELU function is more costly than the ReLU function.
- Responsive to hyperparameters: The ELU function is sensitive to hyperparameters like the alpha parameter, which defines the function's negative saturation value. The neural network may perform poorly if the hyperparameters are wrong.
- The ELU function has certain benefits over the ReLU function, but its computational complexity and hyperparameter sensitivity may restrict its applicability in some scenarios.