ID: 22308010
Name: Shazeda Sultana

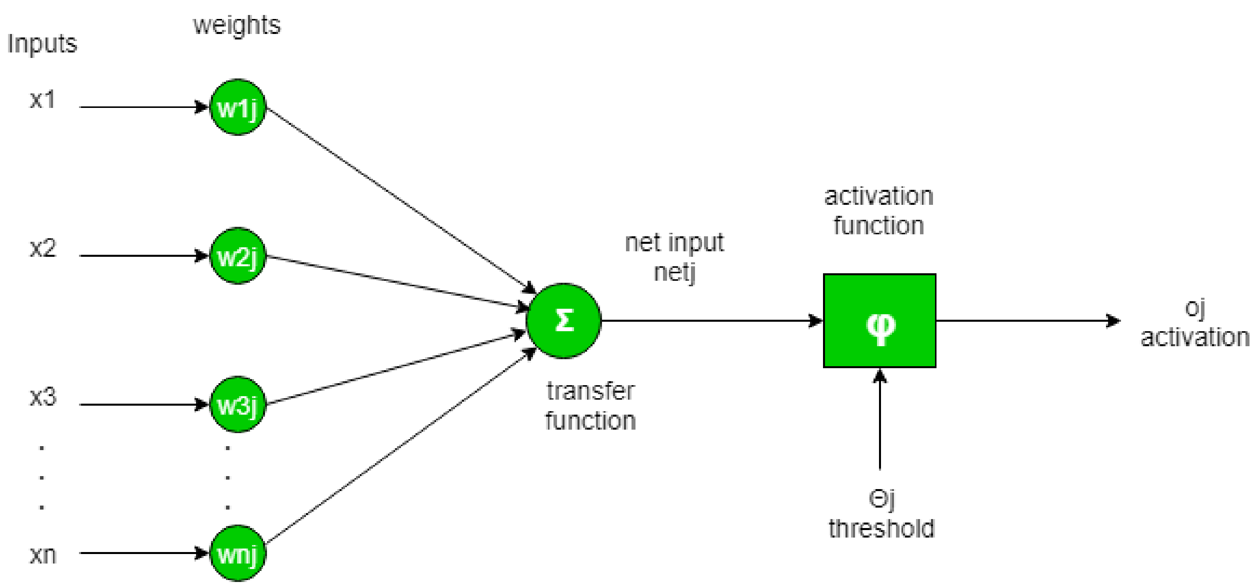
Assignment on Deep Learning

Part-A: Theoretical Concepts:

1. Explain Activation Function:

Simply described, an artificial neuron computes the' weighted aggregate' of its inputs and adds a bias, as seen in the picture below by the net input.

Net input = ∑(weight x input) * bias.



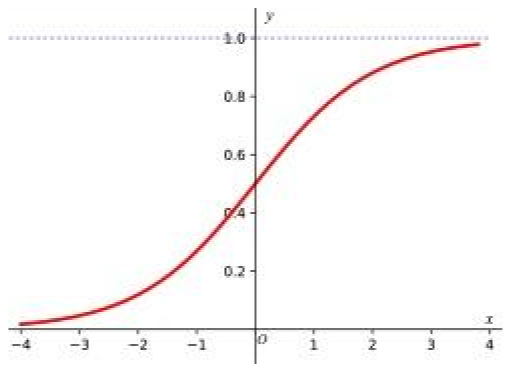
The value of net input might range from- inf to inf. Because the neuron doesn't know how to bind to value, it's ill-equipped to determine the blast pattern. The activation function is an essential element of an artificial neural network. They decide whether or not to actuate a neuron. still, it limits the value of the net input. The activation function is a circular metamorphosis that we perform on the input before moving it to the coming generation of neurons or completing the event.

Define and compare the activation Function given below :

- ❖ **Sigmoid Function:** Sigmoid function is an extensively used activation function. It's defined as

$$\sigma(x) = 1/(1+e^{-x}).$$

The graph of the sigmoid function is shown in Fig. 2.10. When x is small, the value of the sigmoid function is close to 0. When x is large, the value of sigmoid is close to 1. The sigmoid function transforms the nonstop real number into a range of(0,1)
, so that the input value of the coming subcaste is within a fixed range and the weight is more stable.

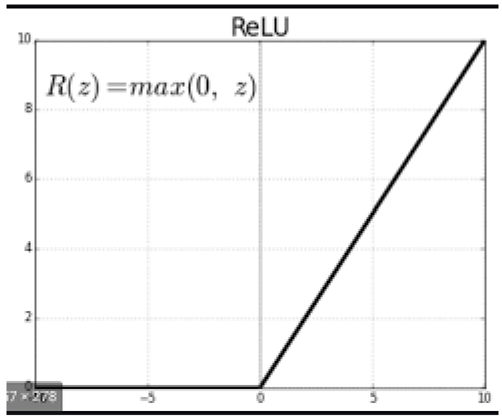


Use Case: Binary classification, Linear regression
Limitations: The network does not learn rapidly close to the borders, which is a drawback of the sigmoid function.

- ❖ **ReLU Activation Function:** The remedied Linear Unit(ReLU) is a popular activation function generally used in deep literacy models. The function labors a value that's directly commensurable to the input, which makes it a simple yet effective mileage for neural networks.

- The ReLU Activation Function offers several advantages, including

- ➢ Simplicity As a direct function, ReLU is easy to apply and computationally effective.
- ➢ working the evaporating grade Problem ReLU helps to alleviate the evaporating grade problem in neural networks, enabling them to learn from the backpropagation process more effectively.
- ➢ Sparsity ReLU activation function results in a meager representation which is beneficial in terms of memory storehouse and calculation.



Formula which is denoted by: f(x)=max(0,x)
Required the-

- f(x)>0; f(x) =0 x;
- f(x)<=0; f(x) ==0;

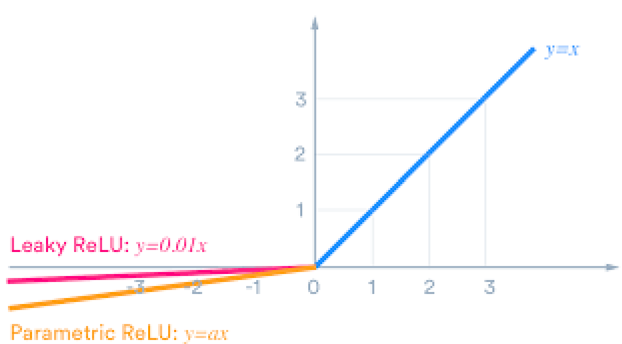
Challenges and Limitations

Despite its advantages, ReLU comes with certain limitations. The most significant bone is the" Dead ReLU" problem, where neurons can occasionally come wedged during the literacy process and cease to modernize, making them unresponsive to variations in error.

Use Case

- Generally used in retired layers of deep literacy models, especially CNNs and MLPs.
- Suitable for tasks where meager activation is salutary.

❖ **Leaky Relu:** To alleviate the dying ReLU problem, Leaky ReLU introduces a small grade for negative inputs, conserving some exertion in the neurons. Still, it struggles with thickness for negative inputs, using a fixed pitch throughout training.



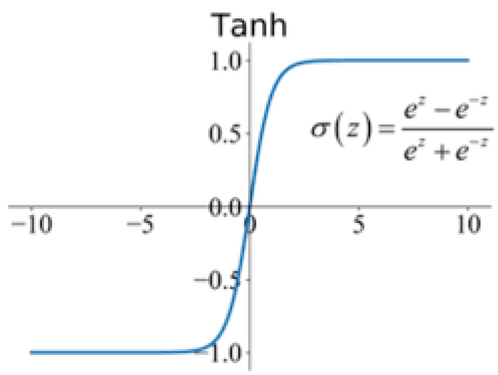
When enforcing Leaky ReLU, it’s important to experiment with conforming the literacy rate, along with regular evaluations. This can help in determining the optimal configuration for Leaky ReLU in a given neural network.

- ☐ **Use Case:** Addresses dying ReLU.
- ☐ **Limitation**: Requires additional parameter tuning.

❖ Tanh:

The tanh activation function, also called the hyperbolic digression activation function, is a fine function generally used in artificial neural networks for their retired layers. It transforms input values to produce affair values between -1 and 1. It's expressed as the rate of the difference between the exponential of the input value and the exponential of its negation to the sum of these exponentials.

Mathematically, the tanh activation function can be represented as
Formula :
tanhf(x) = (e^x - e^-x)/(e^x + e^-x)



Use Case: The tanh function is constantly used in the sheltered layers of a neural network. Because of its zero-centered nature, when the data is also formalized to have to mean zero, it can affect further effective training.

Limitations: All the negative values become zero immediately

2. <u>Discuss Optimization Algorithm</u>

SGD(Stochastic Gradient Descent)

- **Advantage** -
    1. Simple
    2. memory-effective.
- **Limitation** - Slow confluence.

Adam

- **Advantage -**
  1. Combines instigation
  2. Adaptive literacy rate.
- **Limitation** - Sensitive to hyperparameters.

RMSprop

- **Advantage-**
  1. Good for non-stationary objects.
- Limitation- May overfit without proper tuning.
- Learning Rate - Affects confluence speed and stability. Dynamic adaptation through schedulers improves training.

❖ Error Analysis

Errors

- Misrepresentation of comparable classes (for example, cat versus canine).
- Poor performance in less frequent classes.
- Perceptibility to rotation or other supplements.

Solutions:

- Increase training data for deficiently performing classes.
- Use more advanced supplements(e.g., brilliance adaption).
- Fine-tune the model or try to transfer knowledge.

☐ Model Design

A CNN with 3 convolutional layers, ReLU activation, maximum pooling, and 2 completely connected layers was enforced. Regularization ways like powerhouse and batch normalization were used to help overfitting.

❖ Result and Discussion

Cross-Validation Results:

Fold 1: 65.67%

Fold 2: 62.27%

Fold 3: 62.74%

Average Cross-Validation Accuracy: 63.56%

Standard Deviation: 1.50%

☐ Conclusion

The CNN effectively classified CIFAR-10 images, pressing the significance of proper preprocessing and model design. unborn advancements could include transfer literacy and deeper infrastructures for better performance.