Q1. An activation function in the context of artificial neural networks is a mathematical function applied to the output of each neuron in a neural network. It introduces non-linearity into the network, allowing it to learn complex patterns in the data.

Q2. Some common types of activation functions used in neural networks include:

- Sigmoid
- Hyperbolic tangent (tanh)
- Rectified Linear Unit (ReLU)
- Leaky ReLU
- Softmax

Q3. Activation functions affect the training process and performance of a neural network in several ways:

- They introduce non-linearity, allowing the network to approximate complex functions.
- They control the output range of neurons, which affects the stability and convergence of the training process.
- They influence the network's ability to learn and adapt to different types of data distributions.

Q4. The sigmoid activation function works by squashing the input value into the range [0, 1]. Its formula is $\sigma(x) = \frac{1}{1+e^{-x}}$. Advantages of the sigmoid function include smoothness and suitability for binary classification problems. However, it suffers from vanishing gradient problems, where gradients become extremely small for large input values or in deep networks.

Q5. The Rectified Linear Unit (ReLU) activation function is defined as $f(x) = \max(0, x)$. It differs from the sigmoid function by being linear for positive input values and zero for negative input values. Unlike the sigmoid function, ReLU does not saturate for large positive values, which helps mitigate the vanishing gradient problem.

Q6. The benefits of using the ReLU activation function over the sigmoid function include:

- Faster convergence during training, especially in deep neural networks.

- Reduced likelihood of the vanishing gradient problem due to its non-saturating nature.
- Simplicity in computation, as it involves only simple thresholding operations.

Q7. The "leaky ReLU" is a variant of the ReLU activation function that addresses the vanishing gradient problem by allowing a small, non-zero gradient for negative input values. It is defined as $f(x)=\max(\alpha x,x)$, where α is a small constant. By introducing this small slope for negative inputs, leaky ReLU prevents neurons from dying out during training.

Q8. The softmax activation function is commonly used in the output layer of neural networks for multi-class classification tasks. It converts raw scores or logits into probabilities by exponentiating each score and normalizing the results to sum up to one. This makes it suitable for outputting probability distributions over multiple classes.

Q9. The hyperbolic tangent (tanh) activation function is similar to the sigmoid function but squashes the input values into the range [-1, 1]. Compared to the sigmoid function, tanh has stronger gradients, making it more suitable for training deeper networks. However, it still suffers from the vanishing gradient problem, especially for extremely large or small input values.