**Non-programming Assignment**

1. **To which values initialize parameters (W, b) in a neural network and why?**

Parameters WWW (weights) are typically initialized using methods like Xavier initialization or He initialization to avoid exploding or vanishing gradients. Xavier initialization ensures that the variance of activations remains constant across layers. Biases bbb are often initialized to zero.

1. **Describe the problem of exploding and vanishing gradients.**

**Exploding gradients** occur when gradients grow exponentially during backpropagation, leading to unstable learning. **Vanishing gradients** happen when gradients become too small, slowing down or halting the training process because weight updates are minimal.

1. **What is Xavier initialization?**

Xavier initialization is a technique where weights are drawn from a normal distribution with zero mean and a variance of 1/nin1/n\_{in}1/nin​, where ninn\_{in}nin​ is the number of input neurons. It helps maintain variance across layers and prevents exploding/vanishing gradients.

1. **Describe training, validation, and testing data sets and explain their role.**

**Training set** is used to fit the model, **validation set** helps tune hyperparameters and assess generalization during training, and **test set** is for final evaluation on unseen data. All are necessary to prevent overfitting and ensure the model generalizes well.

1. **What is a training epoch?**

A **training epoch** is one complete pass through the entire training dataset. Typically, training involves multiple epochs to allow the model to learn from data iteratively.

1. **How to distribute training, validation, and testing sets?**

A common split is 70-80% for training, 10-15% for validation, and 10-15% for testing. Larger datasets may use smaller percentages for validation and test sets.

1. **What is data augmentation and why is it needed?**

**Data augmentation** is a technique to artificially increase the size of the training set by applying transformations like rotation, flipping, etc., to existing data. This improves model generalization by reducing overfitting and increasing robustness.