weight initialization techniques Assignment

1. What is the vanishing gradient problem in deep neural networks? How does it affect training?
Answer: The vanishing gradient problem occurs when gradients become very small during backpropagation, slowing down or stopping the learning process, especially in deep networks.

2. Explain how Xavier initialization addresses the vanishing gradient problem.
Answer: Xavier initialization sets weights with a variance that scales with the number of input and output units, helping prevent vanishing gradients by ensuring stable activations and gradients.

3. What are some common activation functions that are prone to causing vanishing gradients?
Answer: Sigmoid and tanh functions are prone to vanishing gradients because their derivatives can be very small for extreme input values.

4. Define the exploding gradient problem in deep neural networks. How does it impact training?
Answer: The exploding gradient problem happens when gradients become excessively large during backpropagation, leading to unstable weights and making training difficult.

5. What is the role of proper weight initialization in training deep neural networks?
Answer: Proper weight initialization ensures that gradients neither vanish nor explode, enabling stable training and faster convergence.

6. Explain the concept of batch normalization and its impact on weight initialization techniques.
Answer: Batch normalization normalizes activations, reducing internal covariate shift and allowing more flexible weight initialization, improving model stability and convergence speed.