Gradient problem in RNN

Recurrent Neural Networks (RNN) are a powerful type of artificial neural network that can process sequential data such as text, speech, or video. However, they also suffer from some common challenges such as the vanishing and exploding gradient problems.

Gradients are the values that indicates how much a parameter in a neural network should change to reduce the errors. Gradients are essential for updating the weights and biases of the network and improving the performance.

The vanishing and exploding gradient problems occur when the gradients become either too small or too large during backpropagation.

□ Vanishing problem :- When performing a backpropagation process to update weights with the help of a chain rule calculation. The number of layers is increasing the derivative of values become very smaller values and this leads to the new weight and the old weight becoming approximately matching or equal to each other. If the gradients vanish, the network can't learn from the past and loses its ability to capture long-term dependencies. This can lead to poor generalization and underfitting.

Signs for Vanishing problem:-

- Slow training progress.
- Low weight updates.
- Dead neurons: The neurons in some layers have very small or zero outputs, causing them to become inactive.

Solution for gradient Vanishing includes:

1. Activation function such ReLU, tanh
2. batch normalization
3. gradient clipping
4. Use LSTM or GRU

Exploding gradient: Exploding is a problem where a calculated derivative is being large to the level that procedures a new weight with high variety and gap from the old weight which will also lead to never converge to the global minima as well.

If the gradients explode, the network becomes unstable and sensitive to small changes in the input. This can lead to numerical overflow, erratic behavior and overfiting.

Signs for exploding problem:-

- Large Weight updates
- Null Values
- Oscillating performance: The model's accuracy fluctuates over time, indicating instability.