REGULARIZATION in NN

-MUKESH KUMAR

AGENDA

- Drop out
- Batch Normalization

Dropout

- **Purpose**: Dropout is a regularization technique to prevent overfitting by reducing the reliance on any single neuron.
- **How It Works**: During each training step, a random subset of neurons is "dropped out" (set to zero), typically with a probability (e.g., 0.2-0.5).
- **Effect on Training**: Forces the model to learn more distributed representations, as neurons can't rely solely on specific neighboring neurons.

Dropout

- **Effect on Testing**: Dropout is disabled during inference; all neurons are active, and activations are scaled to match training conditions.
- **Best Placement**: Usually placed after fully connected (dense) layers, as these tend to overfit more easily than convolutional layers.
- **Dropout Rate**: A dropout rate of 20-50% is commonly used, but the exact rate may be tuned based on the dataset and model complexity.

Batch Normalization

- **Purpose**: Batch Normalization (BatchNorm) accelerates and stabilizes training by normalizing activations across each mini-batch.
- **Normalization Process**: For each mini-batch, BatchNorm centers (mean = 0) and scales (variance = 1) each feature, standardizing the activations.

Batch Normalization

- **Effect on Gradient Flow**: Reduces issues like vanishing/exploding gradients, leading to faster and more stable convergence, especially in deep networks.

- **Best Placement**: Commonly placed after dense or convolutional layers and before the activation function to normalize activations effectively.

- **Improves Generalization**: By stabilizing training, BatchNorm indirectly improves generalization and can reduce the need for strong regularization techniques.