Module 3 Summary: Training Neural Networks

BACKPROPAGATION

- Chain rule
- Gradient flow
- Computational graph

GRADIENT DESCENT

- SGD, Mini-batch
- Momentum
- Adam optimizer

LOSS FUNCTIONS

- MSE (regression)
- Cross-entropy
- Custom losses

TRAINING LOOP

Forward + Backward + Update

INITIALIZATION

- Xavier/Glorot
- He initialization
- Avoid vanishing grad

REGULARIZATION

- L1/L2 penalty
- Dropout
- Early stopping

HYPERPARAMETERS

- Learning rate
- Batch size
- Architecture

Key Takeaways

1. Backprop = Chain rule applied systematically | 2. Learning rate is the most critical hyperparameter | 3. Regularization prevents overfitting