

Module 3: Training Neural Networks

Backpropagation

- Chain rule
- Gradient flow
- Computational graphs

Optimization

- Gradient descent
- Learning rate
- Momentum
- Adam

Challenges

- Vanishing gradients
- Exploding gradients
- Local minima

Regularization

- L1/L2 penalties
- Dropout
- Early stopping

Diagnostics

- Train/Val curves
- Overfitting detection
- Hyperparameter tuning

Best Practices

- Data splitting
- Batch normalization
- Weight init

Key Update Rule:

$$w \leftarrow w - \eta \cdot \nabla_w L$$