Improving the Neural Network

Overfitting

- Dropout Layers
- Early Stopping
- Data Augmentation
- Regularization (L1, L2)
- Cross-Validation
- Ensemble Methods

Normalization

- Normalizing Inputs
- Batch Normalization
- Layer Normalization
- Weight Normalization
- Feature Scaling

Vanishing Gradient

- Activation Functions
- Weight Initialization
- Skip Connections (Residual Networks)
- Gradient Clipping
- Proper Network Depth

Hyperparameter Tuning

- Number of Hidden Layers
- Nodes per Layer
- Batch Size
- Learning Rate
- Dropout Rate
- Activation Functions
- Optimizer Choice

Optimizers

- Momentum
- Adagrad
- RMSProp
- Adam
- Nadam
- AdamW
- SGD with Nesterov Momentum

Learning Rate Scheduling

- Step Decay
- Exponential Decay
- Cosine Annealing
- Cyclical Learning Rate
- OneCycle Policy

Data Handling

- Data Preprocessing
- Balancing Classes
- Data Augmentation
- Feature Selection
- Noise Reduction

Evaluation & Monitoring

- Validation Curves
- Learning Curves
- Model Checkpointing
- Performance Metrics

Architecture Design

- Residual Connections
- Convolutional Layers
- Recurrent Layers
- Attention Mechanisms
- Model Pruning

Regularization Techniques

- Weight Decay
- Noise Injection
- Label Smoothing
- Early Stopping

Gradient Checking & Clipping

- Numerical Gradient Verification
- Gradient Scaling
- Exploding Gradient Prevention

Transfer Learning & Find Tuning

- Pretrained Models
- Layer Freezing
- Domain Adaptation