

General Tricks

Sigmoid-Specific

Normalization

1. Zero-mean
2. Equal-variance
3. Uncorrelated

Weight Initialization

Choosing Target Values

Activation Functions

Trick 7 RBF VS Sigmoid

Shuffling

Trick 1

1 Preprocessing

Trick 4

2 Forward Propagation

Trick 5

Trick 6

Linear 1

Activation 1

...

Linear n

Activation n

Loss Computation

Linear 1 Gradient

Activation 1 Gradient

...

Linear n Gradient

Activation n Gradient

Loss Gradient

3 Backward Propagation

4 Parameter Update

Trick 2

Iteration

Trick 3

Learning Rate Selection

Stochastic
VS
Batch Learning

Training Flow & Optimization Tricks

7 practical tricks from *Efficient BackProp* are mapped to their corresponding stages in the training process.