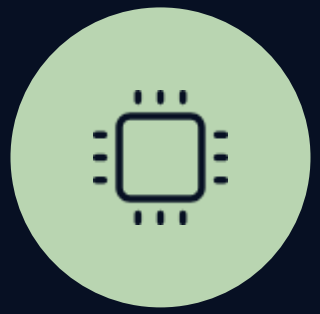
Foundations of Neural Networks

Core concepts and architectural components



Artificial Neurons

The fundamental unit of computation, processing input signals to produce an output.

Perceptrons

Single-layer networks capable of binary classification based on linear decision boundaries.

Layered Architecture

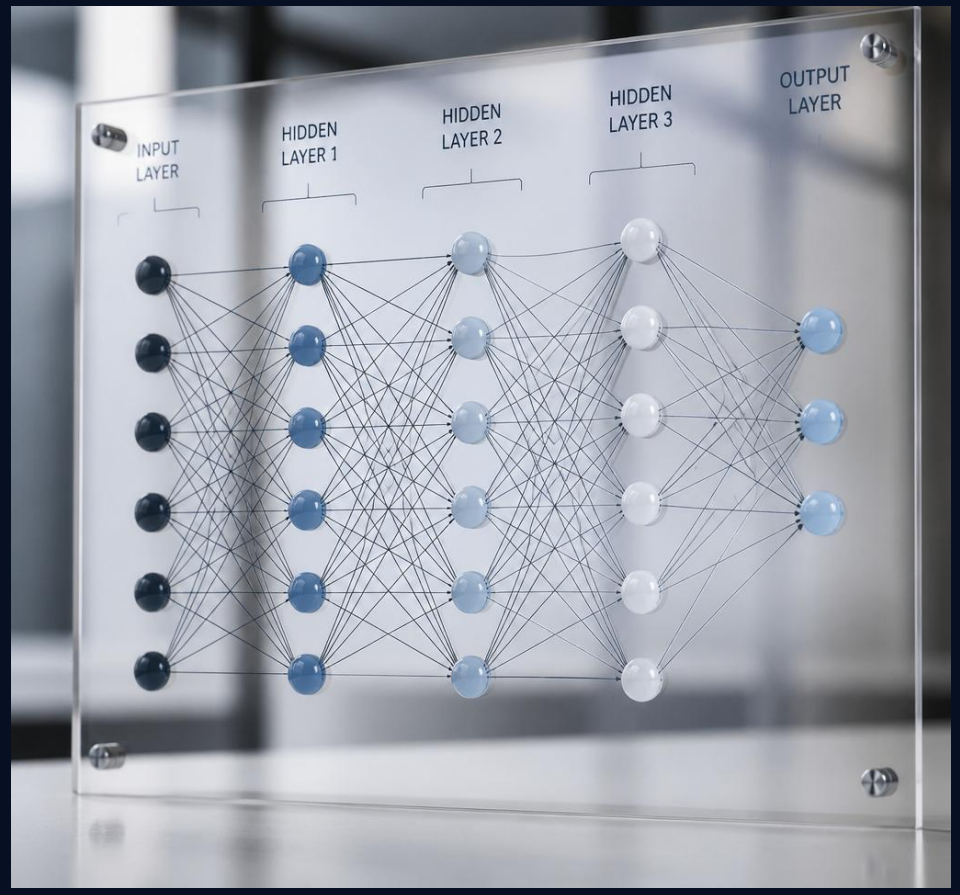
A structured arrangement comprising an input layer, hidden layers for feature extraction, and an output layer.

Key Parameters

Weights and biases that determine the influence of signals and adjust activation thresholds for model learning.

Mathematical Activations and Forward Propagation

The Mechanics of Data Flow



- ## Forward Propagation

  The systematic process of computing network outputs by passing input data through a sequence of weighted layers.

- ## Mathematical Transformation

  Each layer performs a linear combination of inputs ($z = Wx + b$) followed by an activation function ($a = f(z)$).

- ## Non-Linearity

  Activation functions enable the model to learn complex, non-linear relationships within the data.

- ## Common Activation Functions

  ReLU (efficient, mitigates vanishing gradients), Sigmoid (binary classification), and Tanh (better convergence).

Training the Network: The Learning Process

How neural networks learn through error optimization

Loss Functions

Quantifies the difference between the model's prediction and the actual ground truth, providing a numerical error value to minimize.

Gradient Descent

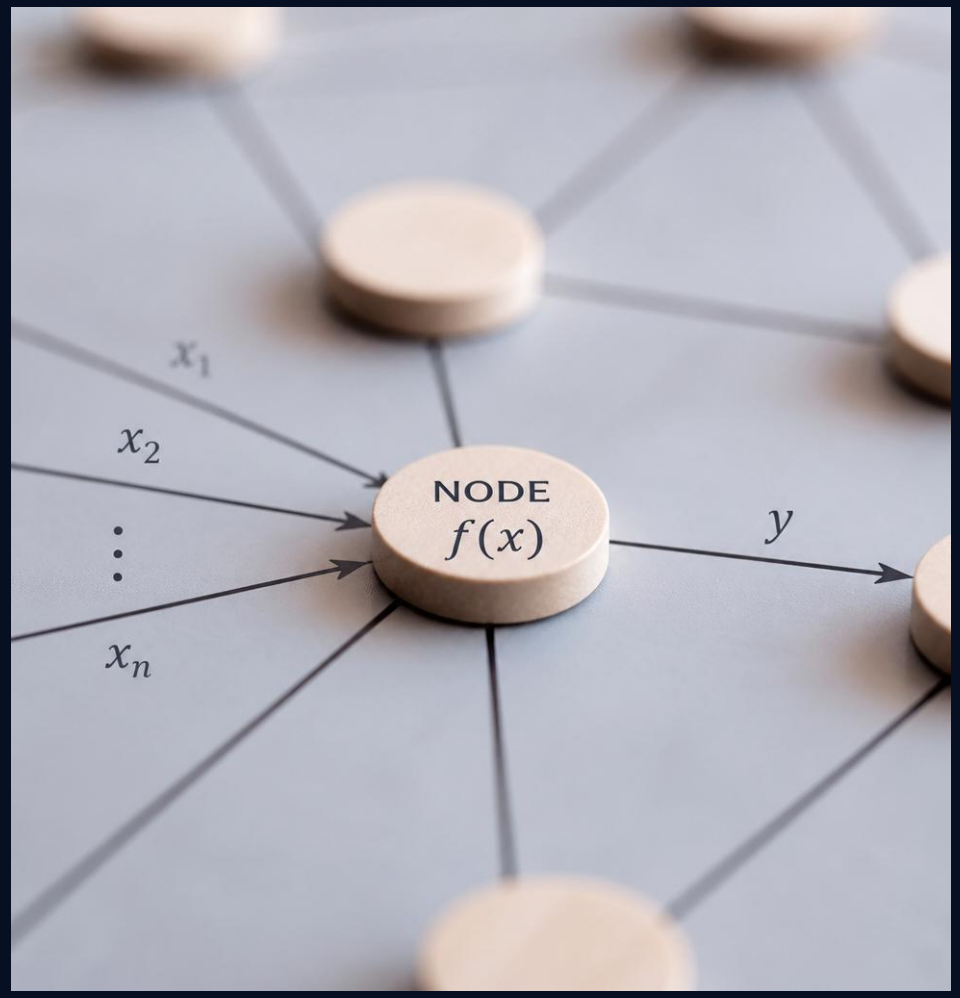
The core optimization algorithm that progressively adjusts weights and biases to reduce error.

Backpropagation

Efficiently computes gradients and propagates error signals backward through the network layers for iterative weight updates.

Advanced Architectures: CNNs, RNNs, and LSTMs

Key deep learning architectures for diverse data types



- CNNs (Convolutional Neural Networks)

  Optimized for spatial data and image processing; uses filters to extract features and recognize patterns in grid-like inputs.

- RNNs (Recurrent Neural Networks)

  Designed for sequential data streams; maintains internal state to process data points that depend on previous inputs.

- LSTMs (Long Short-Term Memory)

  A specialized RNN variant that solves the vanishing gradient problem and retains long-term dependencies over extended sequences.

Deep Learning Applications

Translating Theory to Real-World Tasks

- **Computer Vision**

  Encompasses image classification for content categorization and object detection for locating specific items.

- **Natural Language Processing (NLP)**

  Includes text classification, sentiment analysis, and automated translation for enhanced communication.

Mini Project: Hands-on Implementation

Objective: Build an Image or Text Classifier

Select model structure (CNN/RNN/Transformer) and define input/output layers.

Select loss function (e.g., Cross-Entropy) and choose optimizer with learning rate.

Train on standard datasets like MNIST or IMDb and monitor performance metrics.

Validate on test set and report final accuracy and model performance.