

# List of Activation Functions & Optimizers

## Activation Functions

### Definition:

An activation function is a mathematical function applied to the output of a neuron in a neural network. It introduces non-linearity to the model, allowing it to learn and model complex relationships in data. Activation functions determine whether a neuron should be activated (i.e., its output passed forward) and help neural networks approximate any function.

### List with short description

Activation Function	Description
Sigmoid	Outputs values between 0 and 1; used in binary classification.
Tanh	Outputs values between -1 and 1; zero-centered sigmoid alternative.
Softmax	Converts logits into probabilities; used in multi-class classification.
ELU (Exponential Linear Unit)	Reduces vanishing gradients; outputs smooth transitions for negatives.
ReLU (Rectified Linear Unit)	Outputs 0 for negative values and the input for positive values; widely used.
Leaky ReLU	Allows small gradients for negative values; avoids dead neurons.
Swish	Smooth and differentiable; learned from data.

# Optimizers

## Definition:

An optimizer is an algorithm that adjusts the parameters (weights and biases) of a neural network to minimize the loss function during training. It determines how the model learns by iteratively updating parameters based on the gradients of the loss function with respect to those parameters.

## List with short description

Optimizer	Description
SGD (Stochastic Gradient Descent)	Updates weights based on individual samples or mini-batches.
Momentum	Speeds up SGD by adding a fraction of the previous update vector.
Adam (Adaptive Moment Estimation)	Combines the advantages of Momentum and RMSProp; widely used.
RMSProp	Adapts learning rates for each parameter based on recent gradients.
Adagrad	Adapts learning rate for each parameter; works well for sparse data.
Nadam	Adam with Nesterov momentum for better convergence.
Adadelta	An adaptive learning rate method; improves over Adagrad.