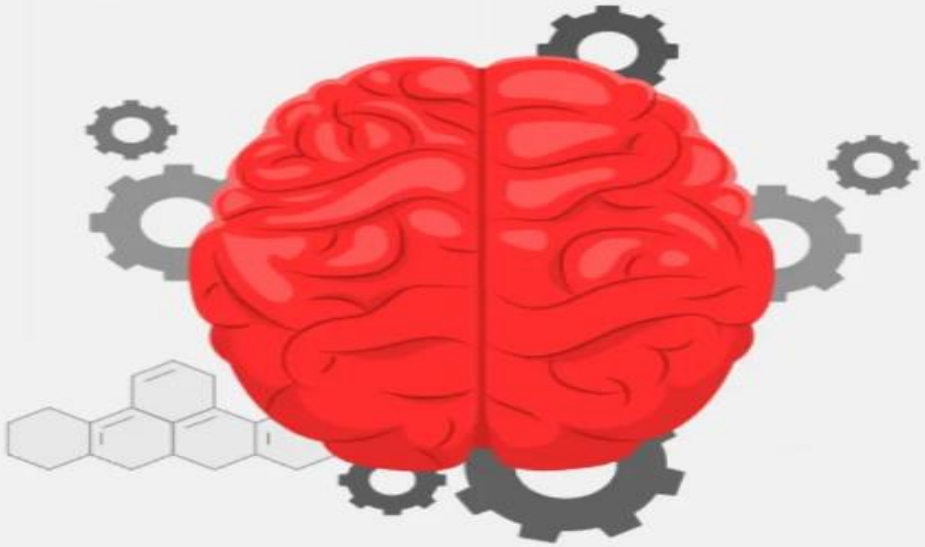
100 Terms In **Deep Learning**

Explained in one line



- **Artificial Neural Network (ANN)**: A computational model inspired by the human brain, comprising interconnected nodes organized in layers for pattern recognition tasks.
- **Backpropagation**: An algorithm used to train neural networks by adjusting weights through minimizing the error between predicted and actual outcomes.
- **Convolutional Neural Network (CNN)**: Specialized deep learning models designed for processing structured grid data, commonly used in image and video recognition.
- **Recurrent Neural Network (RNN)**: Neural networks with loops allowing information persistence, suitable for sequential data analysis like text and time series.
- **Long Short-Term Memory (LSTM)**: A type of RNN capable of learning long-term dependencies, crucial for tasks involving sequences and context.
- **Gradient Descent**: An optimization algorithm used to minimize the error by iteratively adjusting model parameters in the direction of steepest descent.
- **Activation Function**: A mathematical operation applied to neuron outputs in neural networks, introducing non-linearity and enabling complex mapping of inputs to outputs.
- **Dropout**: A regularization technique in neural networks where randomly selected neurons are ignored during training to prevent overfitting.
- **Batch Normalization**: A technique to normalize the inputs of each layer in a neural network, reducing internal covariate shift and accelerating convergence.
- **Generative Adversarial Network (GAN)**: A framework involving two neural networks, the generator and discriminator, competing to generate increasingly realistic data samples.
- **Autoencoder**: A type of neural network designed to learn efficient representations of input data by reconstructing it from a compressed representation.
- **Transfer Learning**: A technique where knowledge gained from training one model is applied to a different but related task, often leveraging pre-trained models.
- **Word Embedding**: A technique to represent words as dense vectors in a continuous vector space, capturing semantic similarities between words.
- **Deep Reinforcement Learning**: A subfield of machine learning where agents learn to take actions in an environment to maximize cumulative reward, often using deep neural networks.
- **Activation Map**: Visual representations of the regions in input space that are important for producing a specific output in a neural network.

- **DropConnect**: A regularization technique similar to dropout, where connections between neurons are randomly dropped during training.
- **Vanishing Gradient Problem**: A challenge in training deep neural networks where gradients become extremely small during backpropagation, hindering learning in early layers.
- **ReLU (Rectified Linear Unit)**: An activation function that introduces non-linearity by outputting the input directly if it is positive, otherwise zero.
- **Softmax**: An activation function used in the output layer of a neural network for multi-class classification, producing probability distributions over classes.
- **Learning Rate**: A hyperparameter determining the step size in gradient descent, influencing the rate at which a model learns during training.
- **Overfitting**: A phenomenon where a model learns to perform well on the training data but fails to generalize to unseen data due to capturing noise or irrelevant patterns.
- **Underfitting**: A scenario where a model is too simple to capture the underlying structure of the data, resulting in poor performance on both training and unseen data.
- **Batch Size**: The number of samples used in one iteration of training a neural network, affecting the stability and speed of convergence.
- **Hyperparameter**: Parameters of a machine learning model that are set prior to training and are not learned from data, such as learning rate and batch size.
- **Kernel**: A small matrix used for convolutional operations in neural networks, extracting features from input data.
- **Pooling**: A downsampling operation in convolutional neural networks, reducing dimensionality and retaining important information.
- **Gradient Checking**: A technique to verify the correctness of gradient computations in backpropagation by comparing them with numerical approximations.
- **Neural Architecture Search (NAS)**: The process of automatically designing optimal neural network architectures, often using reinforcement learning or evolutionary algorithms.
- **Self-organizing Maps (SOM)**: Unsupervised neural networks that create low-dimensional representations of input data while preserving topological properties.
- **Reinforcement Learning**: A type of machine learning where an agent learns to make decisions by interacting with an environment and receiving feedback in the form of rewards.

- **Deep Belief Network (DBN)**: A type of neural network consisting of multiple layers of latent variables and trained using unsupervised learning.
- **Weight Initialization**: The process of setting initial values for the weights of neural networks, crucial for effective learning and avoiding convergence issues.
- **Cost Function**: A function that measures the difference between predicted and actual outputs, used as a guide to adjust model parameters during training.
- **One-shot Learning**: Learning a task from only a single example, often achieved through techniques like siamese networks or metric learning.
- **Regularization**: Techniques used to prevent overfitting by adding additional constraints or penalties to the learning process, such as L1 and L2 regularization.
- **Data Augmentation**: Techniques used to increase the size of training datasets by applying transformations like rotation, scaling, and flipping to input data.
- **Attention Mechanism**: A mechanism in neural networks that allows models to focus on different parts of input data, often used in sequence-to-sequence tasks.
- **Word2Vec**: A technique for learning word embeddings from large text corpora, capturing semantic relationships between words.
- **Deep Learning**: A subset of machine learning where models composed of multiple layers of interconnected units learn to represent data with multiple levels of abstraction.
- **Neural Turing Machine (NTM)**: A neural network architecture with an external memory component, capable of learning algorithmic tasks and maintaining state.
- **Transformer**: A neural network architecture based entirely on self-attention mechanisms, revolutionizing tasks like natural language processing.
- **Fine-tuning**: The process of further training a pre-trained model on a specific task or dataset to adapt it to new domains or requirements.
- **Learning Rate Decay**: A technique to gradually reduce the learning rate during training, allowing finer adjustments in later stages and improving convergence.
- **Word Dropout**: A regularization technique where randomly selected words in input sequences are masked or dropped during training, preventing co-adaptation of words.
- **Gated Recurrent Unit (GRU)**: A variation of the RNN architecture designed to address some limitations of the LSTM while maintaining its effectiveness in sequence modeling.
- **Weight Sharing**: A technique where the same set of weights is used across different parts of a neural network, reducing the number of parameters and improving generalization.

- **Early Stopping**: A regularization technique where model training is halted when performance on a validation dataset starts to degrade, preventing overfitting.
- **Neural Style Transfer**: A technique to generate images combining the content of one image with the style of another, using neural networks.
- **Stochastic Gradient Descent (SGD)**: A variant of gradient descent where the parameters are updated using a randomly selected subset of the training data.
- **DropBlock**: A regularization technique similar to dropout but applied to entire blocks of neurons, preventing co-adaptation of neurons within the same block.
- **Unsupervised Learning**: Learning from unlabeled data, where the model must find patterns and structure in the data without explicit supervision.
- **Supervised Learning**: Learning from labeled data, where the model learns to predict labels from input features based on examples provided during training.
- **Semi-supervised Learning**: Learning from a combination of labeled and unlabeled data, leveraging the benefits of both supervised and unsupervised learning.
- **Maximum Likelihood Estimation (MLE)**: A method for estimating the parameters of a statistical model by maximizing the likelihood function with respect to the parameters.
- **Activation Dropout**: A variant of dropout applied directly to activation values in neural networks, promoting the robustness of the model.
- **Bagging (Bootstrap Aggregating)**: An ensemble learning technique where multiple models are trained on different subsets of the training data, reducing variance and improving generalization.
- **Adaptive Learning Rate**: Techniques where the learning rate is adjusted during training based on observed performance, such as AdaGrad, RMSProp, and Adam.
- **Batch Normalization (BN)**: A technique used to stabilize and accelerate the training of neural networks by normalizing the inputs of each layer.
- **Cross-Entropy Loss**: A loss function commonly used in classification tasks, measuring the difference between predicted and actual class probabilities.
- **Data Leakage**: A situation where information from outside the training dataset unintentionally influences model training, leading to overestimation of performance.
- **Decision Boundary**: The surface separating different classes in the input space, learned by classifiers in supervised learning tasks.
- **Ensemble Learning**: A technique where multiple models are trained to solve the same problem, and their predictions are combined to improve overall performance.

- **Exploding Gradient Problem**: A challenge in training neural networks where gradients become excessively large during backpropagation, leading to numerical instability.
- **Feedforward Neural Network**: A type of neural network where connections between neurons do not form cycles, and information flows in one direction, from input to output.
- **Gated Linear Unit (GLU)**: An activation function used in certain types of neural networks, introducing gating mechanisms to control information flow.
- **Hessian Matrix**: A square matrix of second-order partial derivatives of a scalar-valued function, used to characterize the curvature of loss surfaces in optimization.
- **Jacobian Matrix**: The matrix of all first-order partial derivatives of a vector-valued function, representing the local linearization of the function.
- **KL Divergence (Kullback-Leibler Divergence)**: A measure of how one probability distribution diverges from a second, often used in variational autoencoders.
- **Learning Rate Scheduler**: A technique used to adjust the learning rate during training based on predefined schedules or conditions.
- **Mean Squared Error (MSE)**: A common loss function used in regression tasks, measuring the average squared difference between predicted and actual values.
- **Neuroevolution**: A method of training neural networks using evolutionary algorithms, evolving architectures and weights over successive generations.
- **Orthogonalization**: A principle in machine learning where different concerns, such as optimization and generalization, are addressed independently during model development.
- **Principal Component Analysis (PCA)**: A dimensionality reduction technique used to transform high-dimensional data into a lower-dimensional space while preserving variance.
- **Quadratic Weighted Kappa**: A metric used to measure agreement between two raters, often used in tasks involving ordinal classification, like medical diagnoses.
- **Residual Network (ResNet)**: A type of neural network architecture designed to address the vanishing gradient problem by introducing skip connections.
- **Self-Supervised Learning**: Learning from unlabeled data using supervision from the data itself, often achieved through pretext tasks like predicting context in text.
- **T-distributed Stochastic Neighbor Embedding (t-SNE)**: A dimensionality reduction technique used for visualization, emphasizing the local structure of data points..

- **Universal Approximation Theorem**: A theorem stating that feedforward neural networks with a single hidden layer containing a finite number of neurons can approximate any continuous function.
- **Variational Autoencoder (VAE)**: A type of autoencoder with a probabilistic formulation, capable of generating new data samples similar to those in the training data.
- **Weight Decay**: A regularization technique where a penalty is applied to the magnitude of weights during optimization to prevent overfitting.
- **Xavier Initialization**: A method for initializing weights in neural networks to maintain activation variances and gradients across layers during training.
- **You Only Look Once (YOLO)**: A real-time object detection system that processes images in a single pass through a neural network, enabling fast inference.
- **Zero-shot Learning**: Learning to recognize classes without seeing any examples during training, often achieved through attribute-based classification or knowledge transfer.
- **Attention Score**: A scalar value representing the relevance of a specific element in a sequence to a given query, often computed using attention mechanisms.
- **Batch Reinforcement Learning**: A reinforcement learning setting where multiple agents learn simultaneously, interacting with the environment in batches.
- **Capsule Network**: A neural network architecture designed to better capture hierarchical relationships and spatial hierarchies in data compared to traditional CNNs.
- **Domain Adaptation**: Adapting a model trained on a source domain to perform well on a different target domain, often with limited labeled data in the target domain.
- **Echo State Network (ESN)**: A type of recurrent neural network with a fixed, randomly initialized reservoir of neurons, often used in time-series prediction tasks.
- **Federated Learning**: A distributed machine learning approach where model training is performed across multiple decentralized devices, preserving data privacy.
- **Gaussian Mixture Model (GMM)**: A probabilistic model representing data as a mixture of multiple Gaussian distributions, often used in clustering and density estimation.
- **Hierarchical Clustering**: A method of cluster analysis where data points are grouped based on their similarity into nested clusters, forming a hierarchy.
- **Instance-based Learning**: A learning paradigm where models make predictions based on instances or examples from the training data, without explicitly building a generalizable model.

- **Kernel Trick**: A method used in kernel machines, such as support vector machines, to implicitly map input data into high-dimensional feature spaces.
- **Locally Connected Layer**: A layer in neural networks where each neuron is only connected to a local region of the input, reducing the number of parameters.
- **Meta-learning**: Learning to learn, where models are trained on a variety of tasks and adapt quickly to new tasks with minimal additional training.
- **Neighborhood Components Analysis (NCA)**: A method for learning a distance metric from labeled data, emphasizing local neighborhood structure in feature space.
- **Online Learning**: A learning paradigm where models are updated continuously as new data becomes available, allowing adaptation to changing environments.
- **Pooling Layer**: A layer in convolutional neural networks used to progressively reduce the spatial dimensions of the input volume, typically using max or average pooling.
- **Quantum Neural Network (QNN)**: A type of neural network that leverages principles from quantum computing to perform certain computations more efficiently, still an emerging field with ongoing research.

-----THANNKYOU-----