Chapter - 5

Deep Learning

Units for Discussion

Unit - 1	Unit 1
Unit - 2	Unit 2
Unit - 3	Unit 3
Unit - 4	Unit 4

Unit 5

Deep Learning

DISCLAIMER

The content is curated from online/offline resources and used for educational purpose only.

Learning Objectives

- Identify and describe the roles of neurons as the basic units of a neural network.
- Differentiate between input, hidden, and output layers and explain their purposes in a neural network.
- Define activation functions and explain their importance in introducing non-linearity to neural networks.
- Explain the concept of a loss function and its role in assessing the performance of a neural network.
- Define optimization algorithms and their importance in updating neural network weights to minimize loss.
- Evaluate the impact of various choices (e.g., activation functions, loss functions, optimization methods) on network performance.

Source : www.freepik.com/

Introduction to Deep Learning

- Deep learning, a subfield of AI, uses neural networks inspired by the human brain to process large data.
- It excels in tasks like image recognition, natural language processing, and reinforcement learning, achieving state-of-the-art results.

Source : https://www.ileafsolutions.com/demystifying-deep-learning-a-beginners-guide

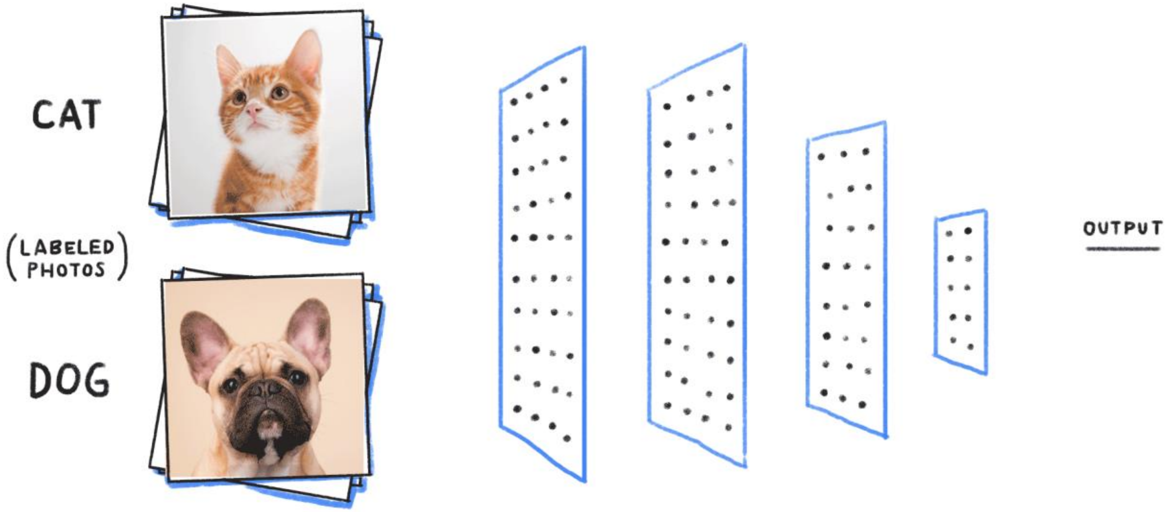
What is Deep Learning?

Artificial Neural Networks: Deep learning algorithms are inspired by the structure and functioning of the human brain.

Pattern Discovery: Deep learning uses networks with multiple hidden layers to automatically discover patterns in complex data sets.

Important Characteristics:

Layered Structure	Automated Feature Extraction
• Neural networks are composed of an input layer, multiple hidden layers, and an output layer.	• Deep learning automatically learns representations directly from raw data, minimizing human intervention.



Source : https://becominghuman.ai/building-an-image-classifier-using-deep-learning-in-python-totally-from-a-beginners-perspective-be8dbaf22dd8

Historical Background

Early Foundations

The first mathematical model of a neuron was proposed in 1943.

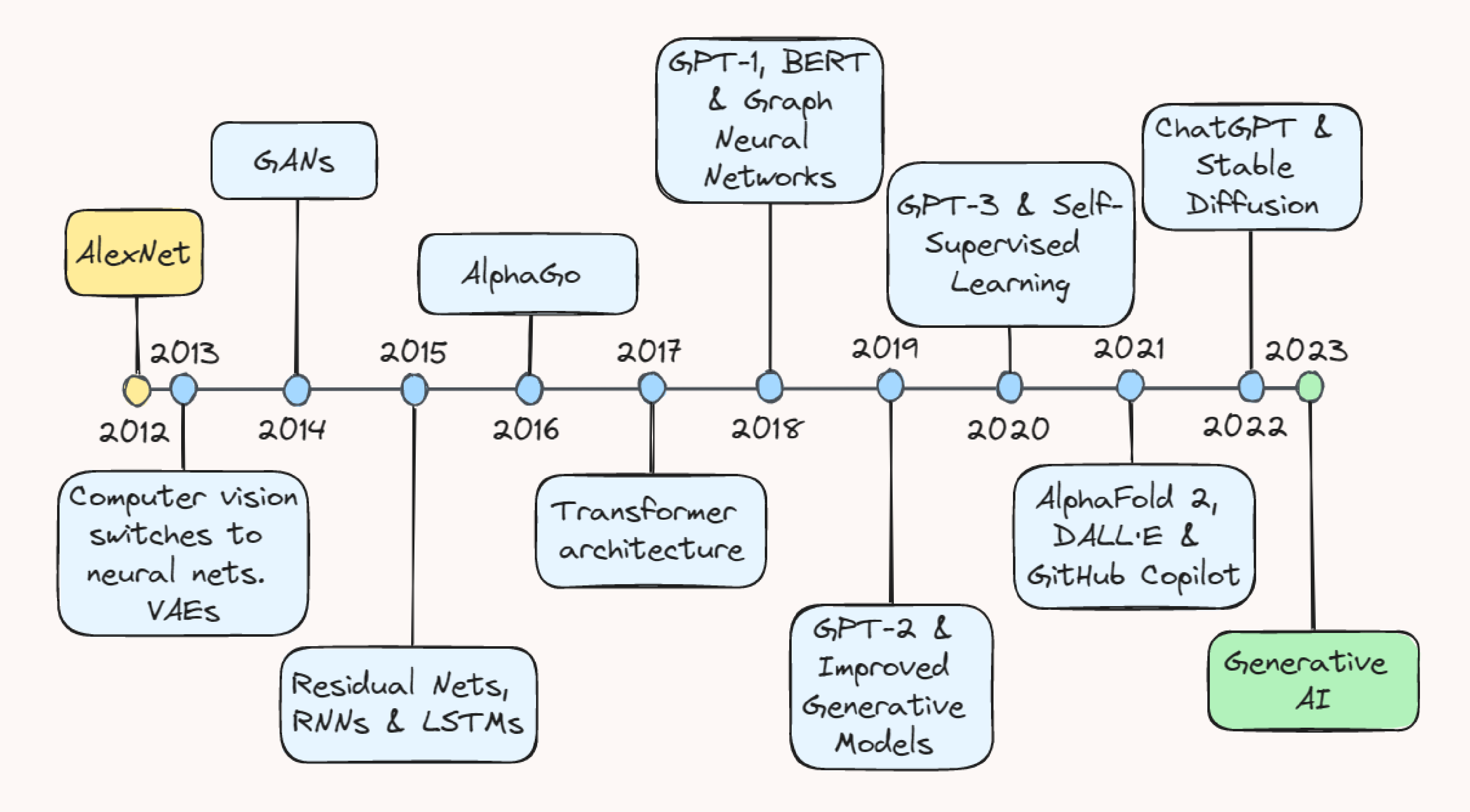
The AI Winter

Limited computing power and datasets led to a decline in interest in neural networks in the late 20th century.

Resurgence of Deep Learning

Advances in GPUs, massive datasets, and new ideas sparked renewed interest in deep learning in the 2000s.

Timeline of last decade



Source : https://towardsdatascience.com/ten-years-of-ai-in-review-85decdb2a540

Why Deep Learning?

High Accuracy

Deep learning achieves state-of-the-art performance in complex tasks, such as image recognition.

Versatility

Deep learning is applicable across various domains, including natural language processing, healthcare, and self-driving cars.

Automation

Deep learning automates feature learning, minimizing human intervention.

Neurons and Layers

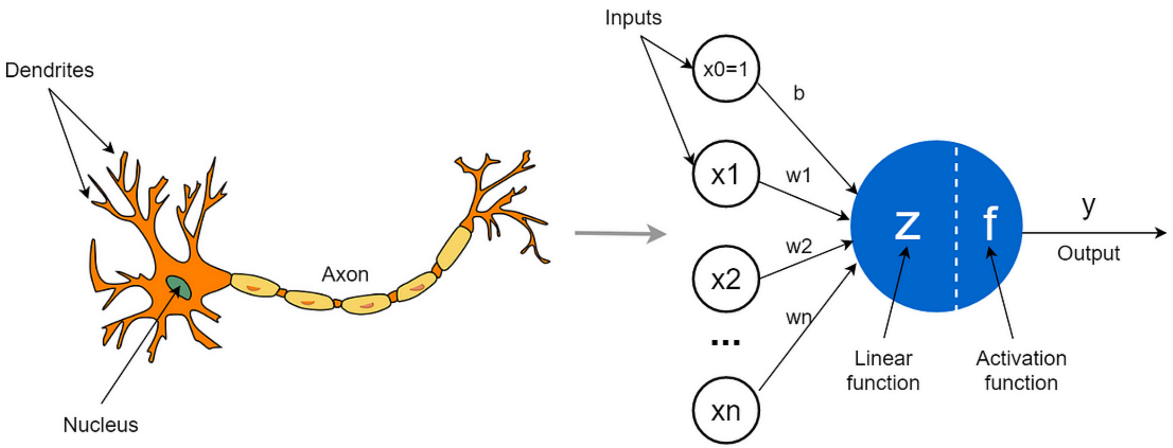
Neurons
The basic building blocks of neural networks, responsible for processing and transforming data.

Layers
Groups of neurons that work together to process data and extract patterns.

Activation Functions
Introduce non-linearity to the model, enabling it to learn complex relationships.

What's a Neuron?

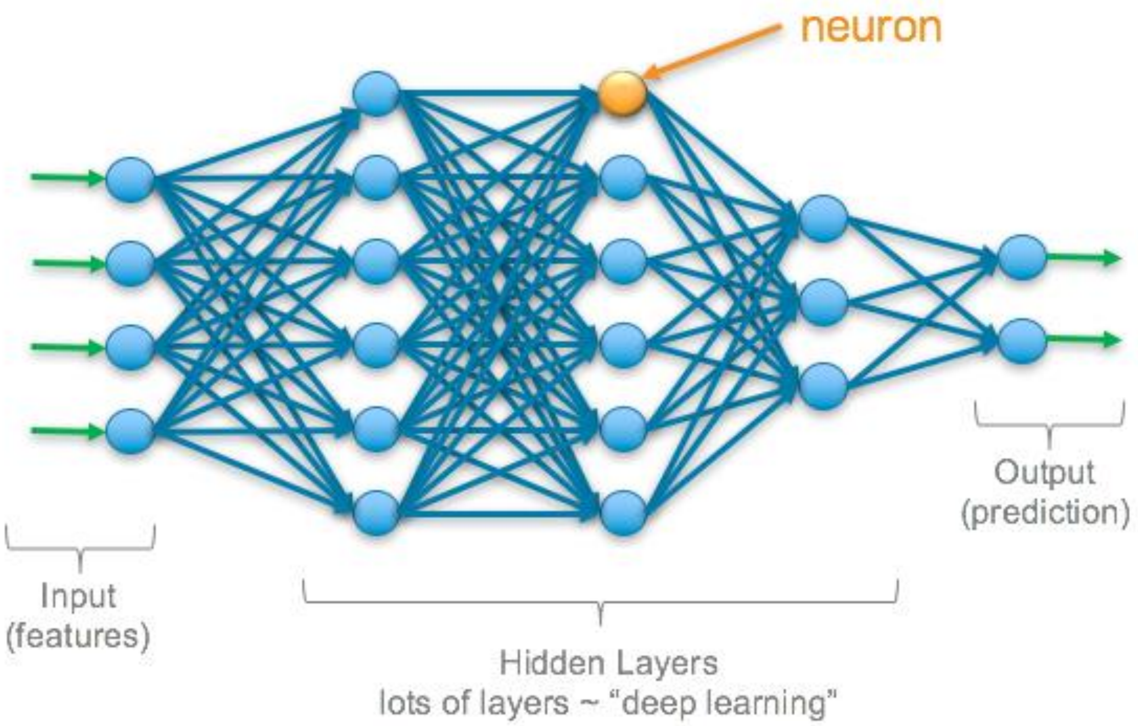
Inputs	Data points or features fed into the neuron.
Weights	Determine the importance of each input.
Bias	Fine-tunes the neuron's output.
Activation Function	Introduces non-linearity to the model



Source : https://towardsdatascience.com/the-concept-of-artificial-neurons-perceptrons-in-neural-networks-fab22249cbfc

Layers in a Neural Network

Input Layer	Receives raw data and passes it to subsequent layers.
Hidden Layers	Process data and extract patterns, relationships, or hierarchies.
Output Layer	Produces the network's prediction based on the processed data



Source : https://srnghn.medium.com/deep-learning-overview-of-neurons-and-activation-functions-1d98286cf1e4

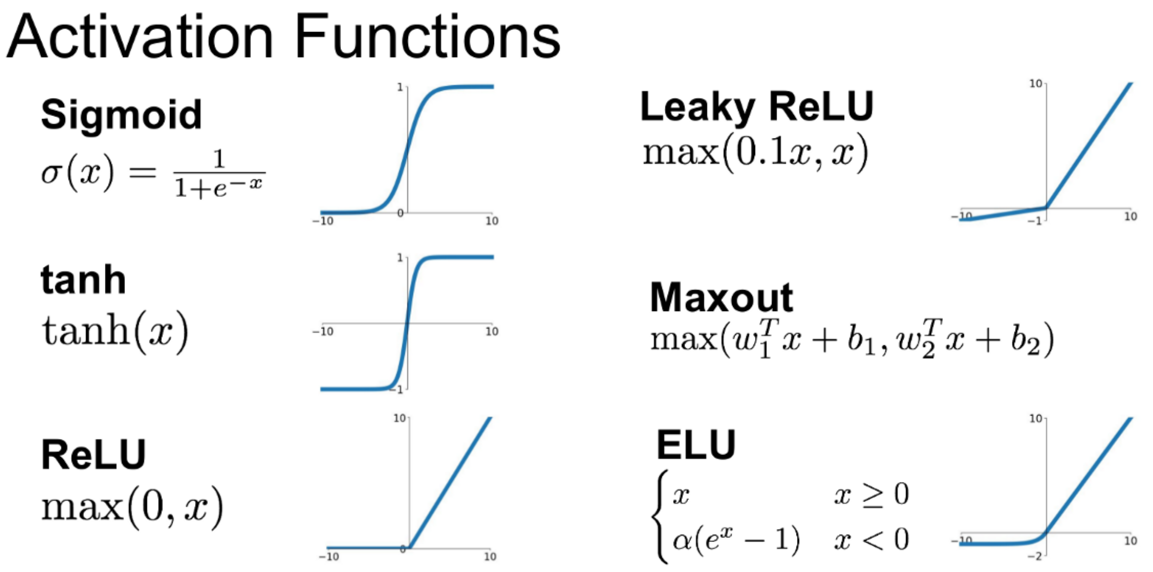
Activation function

- The activation function introduces non-linearity into the network, allowing it to learn and model complex relationships in the data.

Common activation functions are:

- Rectified Linear Unit (ReLU)
- Sigmoid Function
- Hyperbolic Tangent (tanh) Function
- Leaky Rectified Linear Unit (Leaky ReLU)
- Softmax Function

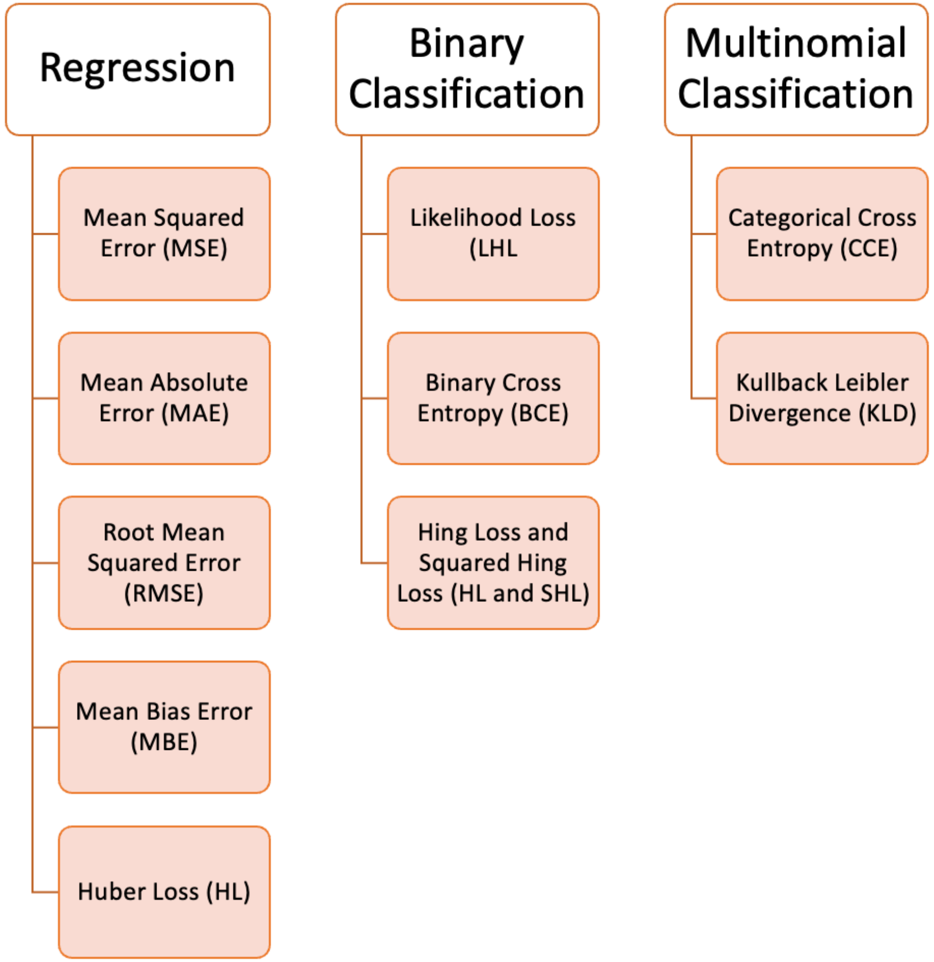
Activation functions play a crucial role in determining the capacity and behavior of neural networks, affecting their convergence speed, stability, and performance on different tasks.



Source : https://www.researchgate.net/figure/Activation-functions_fig3_364953720

Loss functions

- Loss functions, also known as cost functions measures how well a model performs by quantifying the difference between predicted values and actual values.
- The goal is to minimize this difference during model training.
- The loss functions depends on the type of problem in the hand. The loss functions for regression problems and classification problems are different.



Source : https://towardsai.net/p/l/10-commonly-used-loss-functions-explained-with-python-code

Regression Loss Functions

Mean Squared Error (MSE):

- Measures the average squared difference between actual y_i and predicted $\hat{y}_i$ values.
- **Behaviour:**
  - Penalizes large errors more heavily due to squaring, making it sensitive to outliers.
  - Provides smooth gradients, making it easy to optimize.
- **Use Case:** Continuous data, such as predicting housing prices or stock prices.
- **Drawbacks:** Outliers can disproportionately influence the loss, leading to suboptimal solutions.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Regression Loss Functions

Mean Absolute Error (MAE):

- Measures the average absolute difference between actual and predicted values.
- **Behaviour:**
  - Treats all errors equally without squaring.
  - Produces less sensitivity to outliers compared to MSE.
- **Use Case:** Predictive tasks where robust treatment of outliers is critical.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

Drawbacks: Gradients are not smooth (at $y_i = \hat{y}_i$), which can make optimization harder.

Classification Loss Functions

Binary Cross-Entropy Loss:

- Measures the difference between true labels $(y_i \in \{0, 1\})$ and predicted probabilities $\hat{y}_i$.
- **Behaviour:**
  - Penalizes incorrect predictions logarithmically.
  - Ensures higher penalties for confident but wrong predictions.
- **Use Case:** Binary classification tasks like spam detection, disease diagnosis.
- **Drawbacks:** Requires numerical stability adjustments to avoid $log(0)$

$$\text{Loss} = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right]$$

Classification Loss Functions

Categorical Cross-Entropy Loss:

- Extends binary cross-entropy to multi-class problems by summing over all K classes.
- **Behaviour:**
  - Encourages the predicted probability for the true class to be as close to 1 as possible.
  - Penalizes low probabilities for the true class heavily.
- **Use Case:** Multi-class classification tasks like image classification (e.g., identifying handwritten digits).
- **Drawbacks:** Sensitive to imbalanced class distributions.

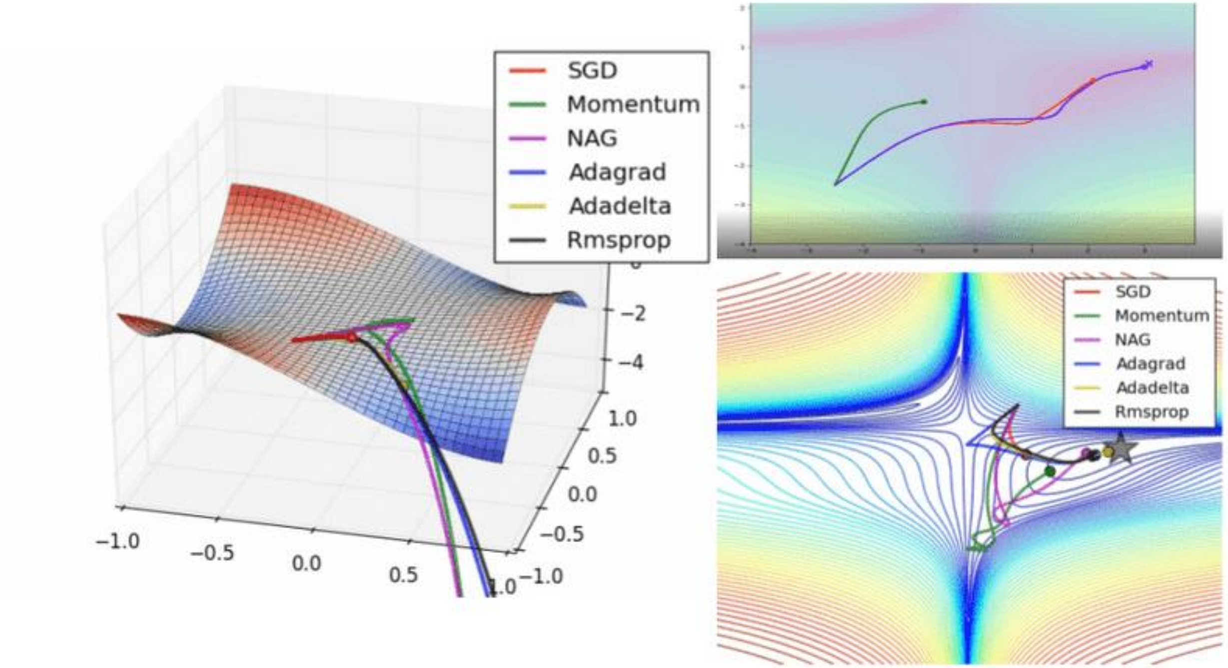
$$\text{Loss} = -\sum_{i=1}^{n} \sum_{k=1}^{K} y_{i,k} \log(\hat{y}_{i,k})$$

Optimization Algorithms

- Optimization algorithms play a crucial role in training machine learning models by iteratively updating model parameters to minimize a chosen loss function.

Common algorithms:

- Gradient Descent
- Momentum
- Adaptive Learning Rate Methods
- Adaptive Learning Rate with Second Order Methods
- Regularization



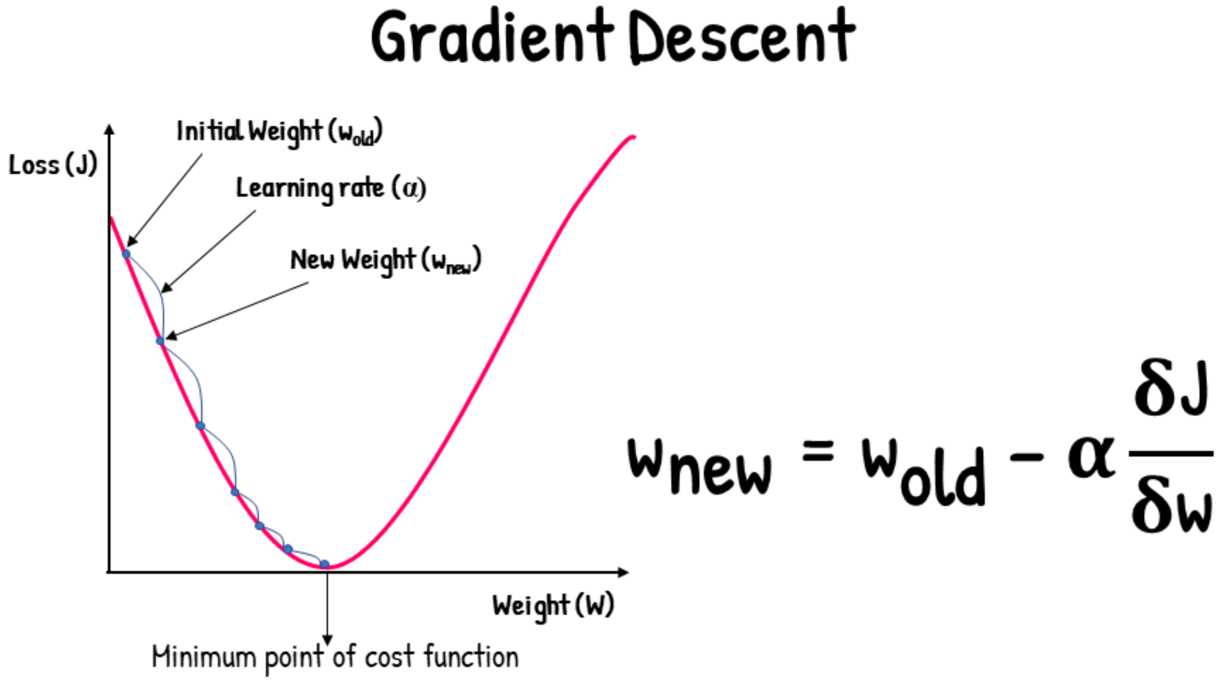
Source : https://www.slideshare.net/slideshow/deep-learning-optimization-basic/86898891

Optimizer: Gradient Descent

- Gradient descent minimizes the loss function of a machine learning model.
- It iteratively updates the parameters of the model in the opposite direction of the gradient of the loss function, aiming to reach the minimum of the loss function.

Variants of Gradient descent:

- Batch Gradient Descent
- Stochastic Gradient Descent (SGD)
- Mini-batch Gradient Descent



Source : https://www.linkedin.com/pulse/understanding-gradient-descent-algorithm-its-role-linear-mhango-kjbvf

Optimizer: Adam (Adaptive Moment Estimation)

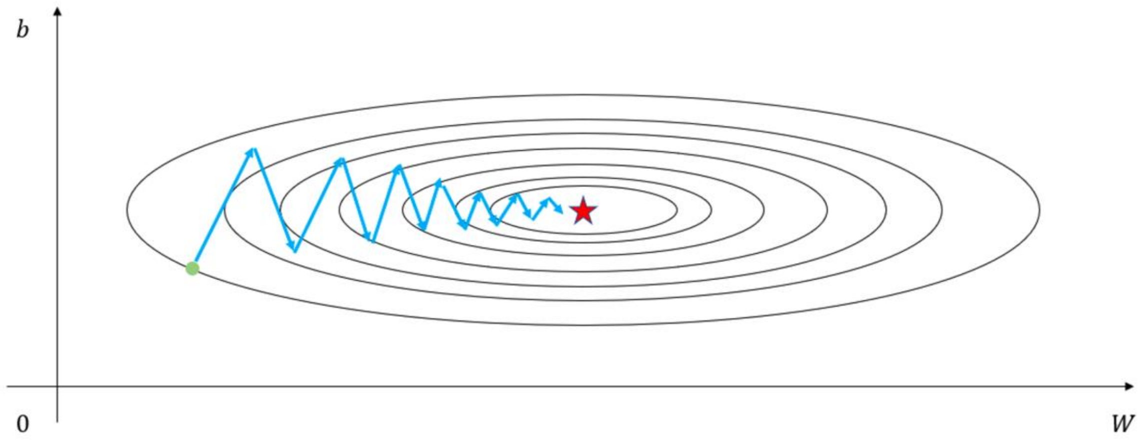
- It's known for its efficiency and effectiveness in a wide range of deep learning tasks.

Advantages:

- Adaptive learning rates for each parameter.
- It performs well with sparse gradients.
- Relatively easy to implement and computationally efficient.

Disadvantages:

- Requires tuning of hyperparameters
- It accumulates momentum and second moment across training, which could lead to bad performance when dealing with noisy gradients.



Source : https://datalya.com/blog/2017/gradient-descent-with-momentum

Lab Activity

Hands On

Lab 1

- Place Your Content Here

Conclusion

- Neural networks rely on neurons as their fundamental units, working across input, hidden, and output layers to process and generate predictions.
- Activation functions introduce non-linearity, enabling complex modeling, while loss functions measure performance, and optimization algorithms adjust weights to minimize errors.
- Key design choices, such as activation functions, loss functions, and optimization methods, critically impact network performance, making them essential to creating effective models.

Source : www.freepik.com/

References

- https://artemoppermann.com/activation-functions-in-deep-learning-sigmoid-tanh-relu/
- https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6
- https://taylorandfrancis.com/knowledge/Medicine_and_healthcare/Neurology/Artificial_neural_networks/
- https://en.wikipedia.org/wiki/Neural_net
- https://www.ibm.com/think/topics/loss-function


Q&A
Let’s Start

Quiz

1.What does a neuron in a neural network represent?

a) A single layer in the network.

b) A mathematical function performing computations like weighted sums and activation.

c) A manual feature engineering component.

d) A unit that only stores input data.

Answer: B

A mathematical function performing computations like weighted sums and activation.

Quiz

2. What is the primary role of weights in a neuron?

a) To determine the importance of each input.
b) To eliminate biases in predictions.
c) To serve as the activation function.
d) To store the final outputs.

Answer: A
To determine the importance of each input.

Quiz

3. Which activation function outputs the maximum of 0 and the input value?

a) Sigmoid.

b) ReLU (Rectified Linear Unit).

c) Tanh.

d) Linear.

Answer: B

ReLU (Rectified Linear Unit).

Quiz

4. What is a key characteristic of deep learning compared to traditional machine learning?

a) Dependence on manual feature engineering.

b) Automated feature extraction from raw data.

c) Reliance on rule-based systems.

d) Focus solely on linear tasks.

Answer: B

Automated feature extraction from raw data.

Quiz

5. Which activation function is often used in binary classification tasks?

a) ReLU.

b) Tanh.

c) Sigmoid.

d) Softmax.

Answer: C

Sigmoid

Thank You