1. Reticular Theory (1871↪️1873)

↪️ Proposed by: Joseph von Gerlach

↪️ Idea: The nervous system is a single continuous network, not made up of separate cells.

2. Staining Technique (1871↪️1873)

↪️ Discovered by: Camillo Golgi

↪️ Contribution: He developed a chemical method to better visualize nervous tissue, supporting the Reticular Theory.

3. Neuron Doctrine (1888↪️1891)

↪️ Proposed by: Santiago Ramón y Cajal

↪️ Idea: The nervous system is made up of individual cells (neurons) that connect to form a network, contradicting the Reticular Theory.

4. The Term Neuron (1891)

↪️ Coined by: Heinrich Wilhelm Gottfried von Waldeyer↪️Hartz

↪️ Contribution: He named the individual nerve cells "neurons" and supported the Neuron Doctrine.

5. Nobel Prize (1906)

↪️ Awarded to: Golgi and Cajal

↪️ Significance: They received the prize for their work on the structure of the nervous system, highlighting the conflict between their theories.

6. Confirmation of Neuron Doctrine (1950s)

↪ Method: Electron microscopy

↪ Outcome: This technology confirmed that nerve cells are indeed individual cells connected by synapses.

7. McCulloch↪Pitts Neuron (1943)

↪ Developed by: Warren McCulloch and Walter Pitts

↪ Contribution: They created a simplified model of how neurons work, laying the groundwork for neural networks.

8. Perceptron (1957↪1958)

↪ Invented by: Frank Rosenblatt

↪ Idea: A type of artificial neuron that could learn and make decisions, seen as a step towards intelligent machines.

9. Limitations of Perceptrons (1969)

↪ Outlined by: Marvin Minsky and Seymour Papert

↪ Contribution: They discussed the limitations of perceptrons, which led to a decline in interest in neural networks.

10. AI Winter (1969↪1986)

↪ Context: A period of reduced funding and interest in artificial intelligence due to unmet expectations.

11. Backpropagation (1986)

↪ Popularized by: Rumelhart et al.

↪️ Significance: A method for training neural networks that improved learning efficiency and effectiveness.

12. Gradient Descent (1847)

↪️ Discovered by: Augustin↪️Louis Cauchy

↪️ Idea: A mathematical method used to minimize errors in models, crucial for training neural networks.

13. Universal Approximation Theorem (1989)

↪️ Concept: A neural network with one hidden layer can approximate any continuous function, showing the power of neural networks.

14. Unsupervised Pre↪️Training (2006)

↪️ Developed by: Geoffrey Hinton and Ruslan Salakhutdinov

↪️ Idea: A technique to initialize weights in deep networks, helping them learn better representations of data.

15. Success in Handwriting Recognition (2009)

↪️ Achieved by: Graves et al.

↪️ Outcome: They outperformed all competitors in an international handwriting recognition competition.

16. Success in Speech Recognition (2010)

↪️ Achieved by: Dahl et al.

↪️ Outcome: They showed significant error reduction in speech recognition systems compared to existing methods.

17. New Record on MNIST (2010)

↪️ Achieved by: Ciresan et al.

↪️ Outcome: They set a new record in digit recognition using deep learning techniques on the MNIST dataset.

18. First Superhuman Visual Pattern Recognition (2011)

↪️ Achieved by: D. C. Ciresan et al.

↪️ Outcome: They achieved a very low error rate in a traffic sign recognition competition.

19. Success on ImageNet (2012↪️2016)

↪️ Key Models: AlexNet, ZFNet, VGGNet, GoogLeNet, MS ResNet

↪️ Outcome: These models won various visual recognition challenges, showcasing the power of deep learning in image classification tasks.

20. Convolutional Neural Networks (1989)

↪️ Developed by: Yann LeCun et al.

↪️ Contribution: They applied CNNs to handwriting digit recognition, which became a foundational technique in image processing.

21. Long Short Term Memory (1997)

↪️ Developed by: Hochreiter and Schmidhuber

↪️ Contribution: LSTMs are a type of RNN that can learn long↪️term dependencies, solving issues with traditional RNNs.

22. Deep Reinforcement Learning (2015)

↪ Achievement: Human↪level control in playing Atari games using deep Q↪networks (DQNs).

23. AlphaGo Zero (2015)

↪ Achievement: Surpassed human players in Go, demonstrating advanced strategic thinking without prior knowledge.

24. DeepStack (2016)

↪ Achievement: Defeated professional poker players, showcasing the application of AI in complex decision↪making games.

25. Dota 2 (2017)

↪ Achievement: An AI bot defeated top professional players in the game, highlighting advancements in AI strategies.

26. Language Modeling (2010↪2015)

↪ Key Contributions: Various models improved language understanding and generation, leading to advancements in natural language processing.

27. Generative Models (2013↪2017)

↪ Key Techniques: Variational Autoencoders and Generative Adversarial Networks (GANs) were developed for generating realistic data.

28. Explainability in AI (2017)

↪ Focus: Increasing emphasis on understanding how deep learning models make decisions, addressing concerns about their reliability and transparency.

Success in Handwriting Recognition (2009)

↪️ Achieved by: Graves et al.

↪️ Outcome: They outperformed all entries in an international Arabic handwriting recognition competition, showcasing the effectiveness of deep learning techniques in recognizing handwritten text.

Success in Speech Recognition (2010)

↪️ Achieved by: Dahl et al.

↪️ Outcome: They demonstrated a relative error reduction of 16.0% and 23.2% over a state↪️of↪️the↪️art system, indicating significant advancements in speech recognition using deep learning methods.

Success in Visual Pattern Recognition (2011)

↪️ Achieved by: D. C. Ciresan et al.

↪️ Outcome: They achieved a 0.56% error rate in the IJCNN Traffic Sign Recognition Competition, marking the first superhuman performance in visual pattern recognition tasks.

Convolutional Neural Networks (CNN)

↪️ Developed by: Yann LeCun et al.

↪️ Contribution: CNNs are designed for processing structured grid data, such as images. They utilize convolutional layers to automatically learn spatial hierarchies of features, making them particularly effective for tasks like image classification and recognition. The architecture typically includes convolutional layers, pooling layers, and fully connected layers, allowing the network to learn increasingly abstract representations of the input data.

Better Optimization Methods

↪ 1983: Nesterov introduced methods that improved convergence speed and accuracy in optimization.

↪ 2011: Adagrad was developed, allowing for adaptive learning rates based on the parameters' historical gradients.

↪ 2012: RMSProp was introduced, addressing the diminishing learning rates problem in Adagrad.

↪ 2015: Adam was developed, combining the benefits of both Adagrad and RMSProp for efficient training of deep learning models.

↪ 2016: Eve was introduced as an optimization method that further improved training efficiency.

↪ 2018: Beyond Adam methods were explored, continuing the trend of enhancing optimization techniques in deep learning.

The Curious Case of Sequences

↪ Context: Sequences are prevalent in various domains such as time series, speech, music, text, and video. Each unit in a sequence interacts with other units, necessitating models that can capture these interactions effectively.

↪ Key Models:

↪ Hopfield Network (1982): A content↪addressable memory system for storing and retrieving patterns.

↪ Jordan Network (1986): A recurrent network where the output state of each time step is fed to the next time step, allowing for interactions between time steps.

↪ Elman Network (1990): Similar to the Jordan network but feeds the hidden state to the next time step.

↪ Long Short Term Memory (LSTMs, 1997): A type of RNN that can learn long↪term dependencies, addressing issues like exploding and vanishing gradients.

↪ Sequence To Sequence Learning (2014): Initial successes using RNNs/LSTMs for large↪scale sequence learning problems, introducing attention mechanisms that inspired further research.

The Paradox of Deep Learning

↪ Context: Despite its success, deep learning faces several challenges:

↪ High Capacity: Models are susceptible to overfitting due to their ability to memorize training data.

↪ Numerical Instability: Issues like vanishing and exploding gradients complicate training.

↪ Sharp Minima: The tendency to converge to sharp minima can lead to overfitting.

↪ Non↪Robustness: Deep learning models can be sensitive to small perturbations in input data.

↪ Current Focus: There is an increasing emphasis on explainability and theoretical justifications to understand why deep learning works so well, aiming to bring more clarity and sanity to the field.

Advancements in Deep Learning from 2012 to 2016

1. Success on ImageNet (2012↪2016):

↪ Deep learning models, particularly Convolutional Neural Networks (CNNs), achieved significant breakthroughs in image classification tasks, winning various visual recognition challenges.

↪ Notable architectures developed during this period include:

↪ AlexNet (2012): Achieved a top↪5 error rate of 16.0%.

↪ ZFNet (2013): Improved upon AlexNet with an error rate of 11.2%.

↪ VGGNet (2014): Further reduced the error rate to 7.3%.

↪ GoogLeNet (2014): Achieved an error rate of 6.7%.

↪ MS ResNet (2015): Set a new record with an error rate of 3.6% using a very deep architecture (152 layers).

2. Advancements in Optimization Methods:

- New optimization techniques were developed, including:
  - Adam (2015): A widely used optimization algorithm that combines the benefits of Adagrad and RMSProp for efficient training.
  - Eve (2016): Introduced as a further enhancement in optimization methods.

3. Generative Models:

- The introduction of Generative Adversarial Networks (GANs) in 2014 by Goodfellow et al. allowed for the generation of realistic data, significantly impacting areas like image synthesis.

4. Natural Language Processing:

- Deep learning techniques began to dominate tasks in natural language processing, including language modeling, machine translation, and conversational agents.

Significance of the Neuron Doctrine

- The Neuron Doctrine, proposed by Santiago Ramón y Cajal, established that the nervous system is composed of discrete individual cells (neurons) rather than a continuous network. This was a pivotal shift in understanding the structure and function of the nervous system.

- **Key Points:**
  - It laid the foundation for modern neuroscience by emphasizing the role of individual neurons in transmitting signals.
  - The doctrine helped clarify how neurons communicate through synapses, leading to a better understanding of neural networks in both biological and artificial contexts.

The acceptance of the Neuron Doctrine marked the beginning of a more systematic study of neural circuits and their functions, influencing the development of artificial neural networks in computer science.

Contribution of Backpropagation to Neural Networks

Backpropagation is a key algorithm used for training artificial neural networks, allowing them to learn from data by adjusting weights based on the error of predictions.

Key Contributions:

- Error Minimization: Backpropagation computes the gradient of the loss function with respect to each weight by applying the chain rule, enabling the network to minimize the error in its predictions.
- Efficient Training: The algorithm allows for efficient computation of gradients, making it feasible to train deep networks with many layers.
- Foundation for Deep Learning: The rediscovery and popularization of backpropagation in the 1980s and its subsequent use in deep learning architectures have been crucial for the success of modern neural networks.
- Facilitated Complex Models: By enabling the training of multilayer networks, backpropagation has allowed for the development of complex models capable of capturing intricate patterns in data, leading to advancements in various fields such as computer vision, speech recognition, and natural language processing.

ChatGPT