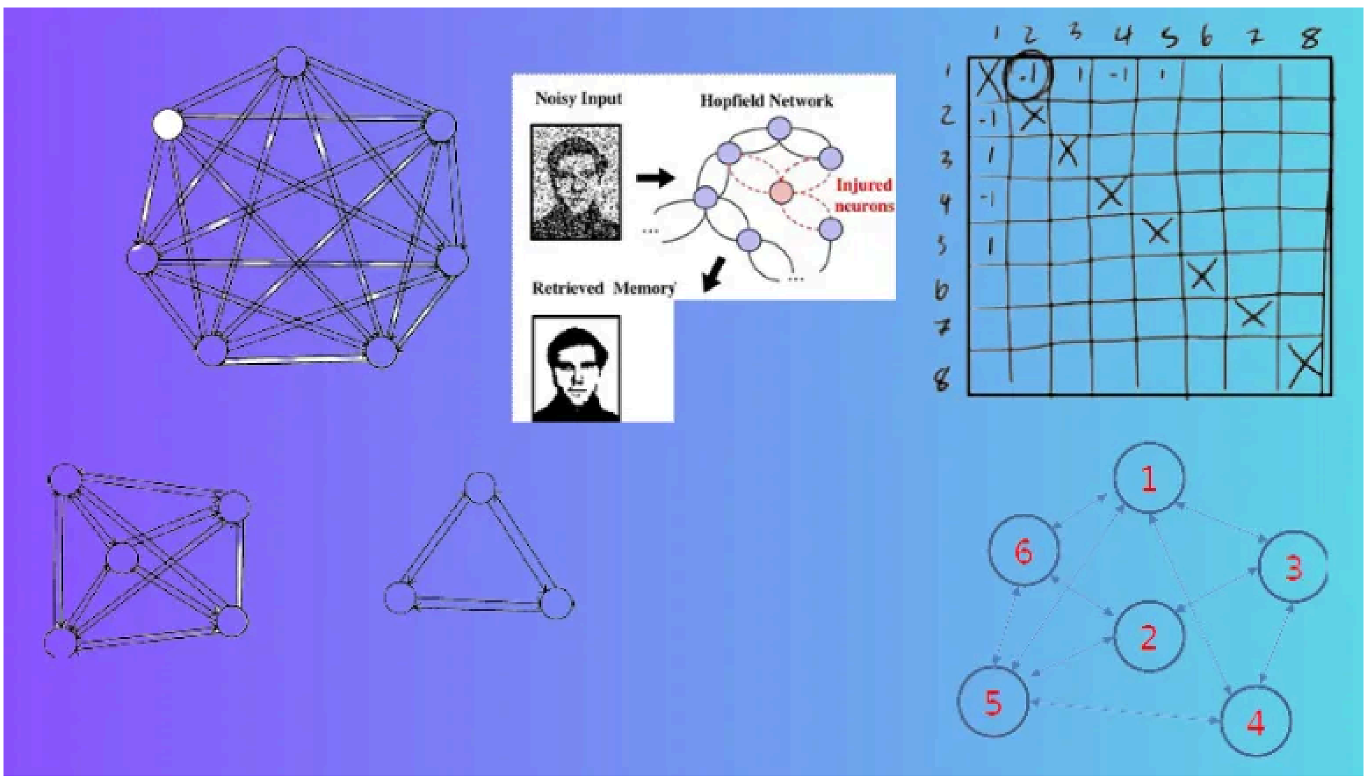
New research in Hopfield Networks: A short intro

Joe El Khoury - GenAI Engineer · Follow

10 min read · Aug 20, 2023

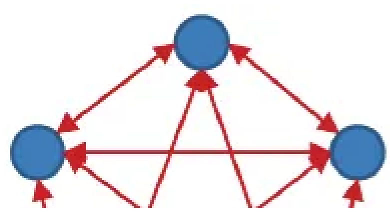


In the era of technological advancement, the ability of a system to recognize and recall patterns, especially when they're imperfect, is of immense value. The Hopfield network, a recurrent artificial neural network popularized by

John Hopfield in 1982, lies at the heart of this capability. Hopfield Networks, originating from the convergence of statistical physics, neuroscience, and computer science, have established themselves as pivotal models for associative memory. With the advent and proliferation of deep learning, their relevance has been accentuated, particularly with the attention mechanism's introduction in the Transformer model. This network is unique due to its symmetric connections, ensuring the weight between any two nodes is mutual. Originally devised as a memory model, its unmatched potential lies in its capacity to reconstruct entire patterns from partial inputs without backpropagation and multi layers involved.

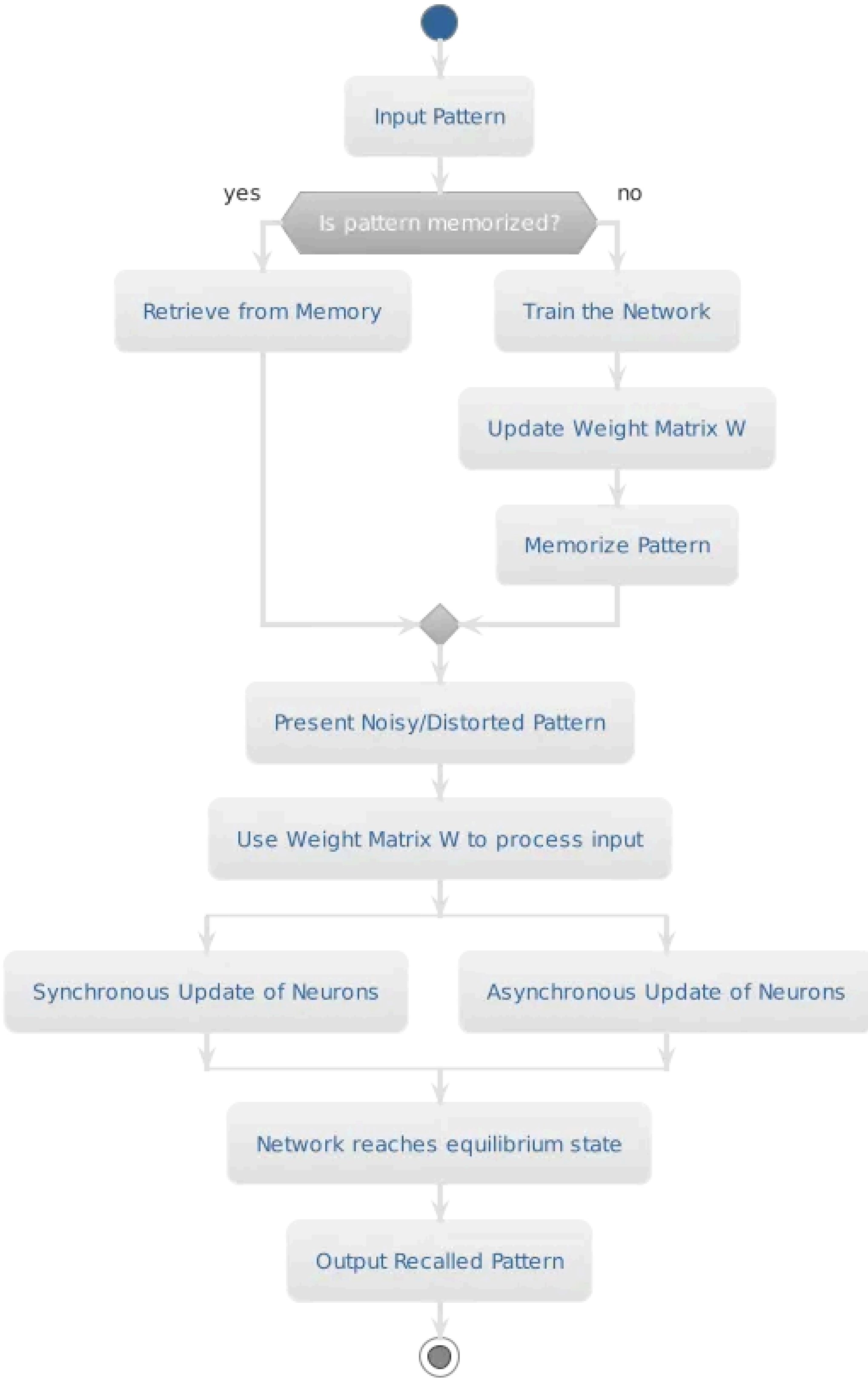
Introduction

The Hopfield network is a fully connected recurrent neural network that can be used for associative memory tasks, such as pattern recognition and noise reduction in input data .



It was popularized by John Hopfield in 1982 and is characterized by symmetric connections between neurons, allowing for bidirectional information flow. The network can memorize specific patterns, and when a noisy or distorted pattern is input, it can recall the original pattern. The structure of a Hopfield network consists of N neurons, with each pair of neurons having connections, resulting in N*N connections, including selfconnections . The connection weight from neuron i to neuron j is represented by a weight matrix W . Each neuron combines inputs from other connected neurons to compute a value and holds a binary value (1 or -1) . The output of the network is an N-dimensional vector containing 1 or -1 values . To store and recall patterns, Hopfield networks use a learning process that involves updating the weight matrix W based on the patterns to be memorized . Once the network is trained, it can recall the original pattern when a noisy or distorted input pattern is presented . The recall process involves updating the neuron values iteratively until the network reaches an equilibrium state, at which point the recalled output is obtained . There are two update methods for Hopfield networks: synchronous (updating all neurons simultaneously) and asynchronous (updating neurons one by one).

Thus, Hopfield networks are a type of recurrent neural network that can be used for associative memory tasks. They have a fully connected structure with symmetric connections between neurons, allowing for bidirectional information flow. The network can memorize specific patterns and recall the original pattern when a noisy or distorted input is presented. The learning and recall processes involve updating the neuron values and weight matrix based on the patterns to be memorized and the input patterns, respectively.



Understanding the Hopfield Network

The Recursive Neural Framework

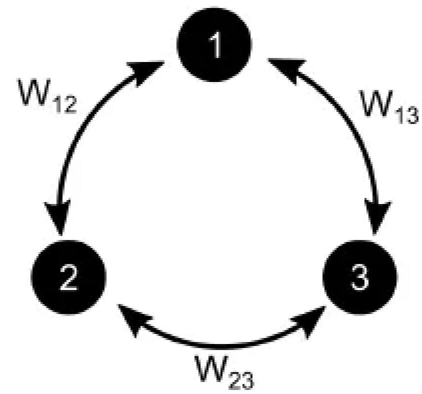
Hopfield Network is characterized as a fully connected recursive neural network. This implies that every neuron is interconnected, ensuring bidirectional information flow. The network operates predominantly based on an energy function. When updating individual neuron states, the energy function is steered in a decremental direction. The network's memory representation is in the form of "stable states," with specific network states gravitating towards these stable configurations.

The Classical Hopfield Networks

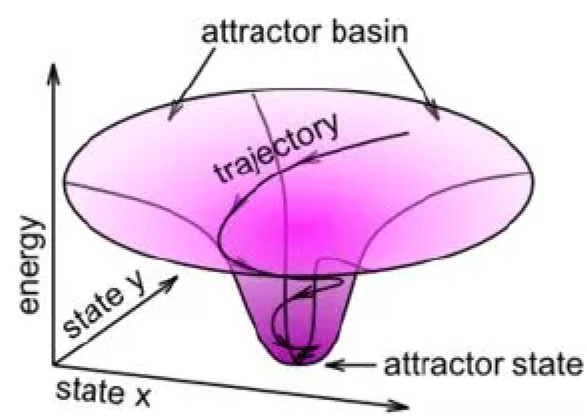
In this classical approach, each state within the network is signified by N neurons, each holding a binary value of ± 1. The energy function encompasses the state of each neuron, the inter-neuronal weights, and the neuron count designated for storage.

A central aspect of this model is the update rule, which ensures that the energy function consistently gets revised in a decremental direction. This network is adept at pattern recognition and error correction. However, there are constraints, such as the limited number of storable patterns. Exceeding this limit can cause the energy function to collapse.

Thus to explain this more without going into formulas, the physicist John Hopfield realized that the basic idea of the Ising model could be used to describe the collective behaviour of neurons, with certain states of activity representing low energy states while assuming that all neurons do not connect to themselves while the output of each neuron is a "spike"(1) or "no spike" (-1) depending on the neuron's input then an energy function can be defined over states. It can be proven that if we randomly select a neuron at each time-step and update its activity, then the energy term will always decrease or stay the same.



Thus, to store a memory, we only need to decrease the energy of that activity pattern to make it an attractor state. which done using gradient descent, we take the partial of the energy with respect to the weights and this is called the Hebbian learning rule.



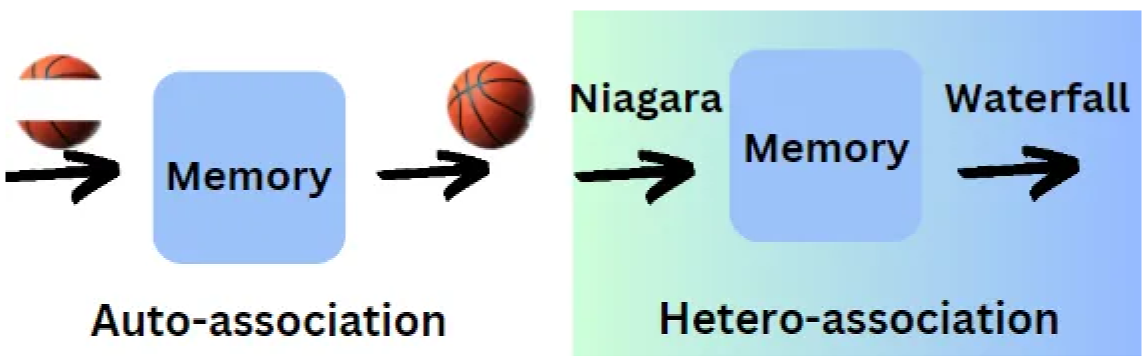
Using this weight update Hopfield networks can be trained to store multiple different memories, with a limit of roughly 0.138n memories, where n is the number of neurons.

Thus, the number of memories that can be stored increases with greater orthogonalization and sparsity in the patterns (making the representations a bit more local). Interestingly, this is what the dentate gyrus of the hippocampus seems to do.

Modeling Associative Memory

We as humans we associate the faces with names, letters with sounds, or we can recognize the people even if they have sunglasses or if they are somehow elder now. or when we listen to the first seconds of a melody we assign it to an old song we listened to. The Hopfield Network's prowess as an associative memory model is undeniable.

> *Associative memory is a data collectively stored in the form of a memory or weight matrix, which is used to generate output that corresponds to a given input, can be either **auto-associative** or **hetero- associative memory**. Classification of associative memory is such that while the memory in which the associated input and output patterns differ are called Hetero Associative Memory, it is called Auto Associative Memory if they are the same.*



Hopfield networks uses auto-association while hopfield networks have seen significant advancements and improvements over the years, particularly in terms of memory capacity and correlations with attention mechanisms. The deep learning revolution prompted a re-evaluation of the Hopfield network concept, leading to significant advancements in memory capacity, as documented in research works by Krotov+16 and Demircigil+17 . Furthermore, correlations between Hopfield networks and attention mechanisms have been established in studies by Ramsauer+20 and ICLR2021 . The idea of leveraging single-body complexes for enhancing neuronal