Hopfield Neural Network and State Transition Diagram

1 What is the Hopfield neural network? What is a state transition diagram for Hopfield Neural Network? Explain how to derive it in Hopfield model.

A **Hopfield Neural Network** is a form of recurrent artificial neural network that serves as a content-addressable memory system. It was introduced by John Hopfield in 1982. The network is designed to store one or more stable patterns (or states) and can retrieve these patterns when presented with a noisy or incomplete version of them. It operates by converging to a stable state that corresponds to one of the stored patterns.

1.1 Key Features

- **Recurrent Architecture**: Each neuron is connected to every other neuron, and there are no self-connections.
- **Symmetric Weights**: The weight matrix is symmetric, i.e., $w_{ij} = w_{ji}$, and $w_{ii} = 0$.
- **Binary States**: Neurons typically have binary states (e.g., $+1$ or -1).
- **Energy Function**: The network has an associated energy function that decreases over time, ensuring convergence to a stable state.

2 State Transition Diagram for Hopfield Neural Network

A **state transition diagram** for a Hopfield network represents all possible states of the network and the transitions between them. Each state corresponds to a specific configuration of neuron activations (e.g., $[1, -1, 1, -1]$). The diagram shows how the network evolves from one state to another based on the update rule until it reaches a stable state (a local minimum of the energy function).

2.1 Steps to Derive the State Transition Diagram

1. **Define the Network**:
   - Specify the number of neurons N.
   - Define the weight matrix W (symmetric and zero diagonal).
   - Define the threshold values for each neuron (if any).
2. **List All Possible States**:
   - For N neurons, there are 2^N possible states (since each neuron can be in one of two states, e.g., $+1$ or -1).
3. **Apply the Update Rule**:
   - For each state, determine the next state using the update rule:

     $$s_i(t+1) = \text{sgn}\left(\sum_{j=1}^{N} w_{ij} s_j(t)\right)$$

     where sgn is the sign function, and $s_i(t)$ is the state of neuron i at time t.
4. **Draw Transitions**:
   - For each state, draw an arrow to the state it transitions to based on the update rule.
   - If a state transitions to itself, it is a stable state (attractor).
5. **Identify Stable States**:
   - Stable states are those where no further transitions occur (i.e., the network remains in that state).

2.2 Example

Consider a simple Hopfield network with 2 neurons:

- Neurons: s_1, s_2
- Possible states: $[1,1], [1,-1], [-1,1], [-1,-1]$
- Weight matrix W:

  $$W = \begin{bmatrix} 0 & w_{12} \\ w_{21} & 0 \end{bmatrix}$$

  where $w_{12} = w_{21}$.

Using the update rule, compute the next state for each state and draw the transitions. Stable states will have no outgoing arrows.

2.3 Energy Function and Convergence

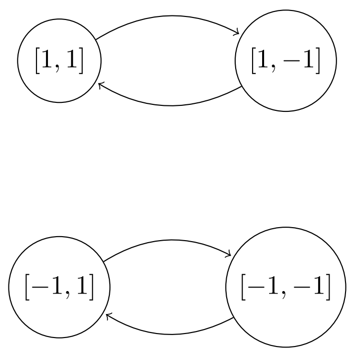
The Hopfield network minimizes an energy function:

$$E = -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} s_i s_j$$

The state transition diagram reflects the network's movement toward lower energy states until it reaches a stable state.

3 State Transition Diagram Example

Below is an example of a state transition diagram for a 2-neuron Hopfield network:



In this example, the states $[1, 1]$ and $[-1, -1]$ are stable states, while $[1, -1]$ and $[-1, 1]$ transition between each other.