Guía de Redes de Hopfield: Fundamentos y Aplicación a Recuperación de Claves

Proyecto `neurocipher`

Introducción

Las redes de Hopfield son un tipo de red neuronal recurrente que puede almacenar patrones binarios y recuperarlos a partir de versiones incompletas o ruidosas. Este documento actúa como una guía teórica para su uso en el proyecto `neurocipher`, orientado a la recuperación de claves binarias.

1. Modelo matemático de la red de Hopfield

Una red de Hopfield clásica está formada por N neuronas binarias $x_i \in \{-1, +1\}$, completamente conectadas entre sí, con pesos simétricos $w_{ij} = w_{ji}$ y sin autoconexiones $w_{ii} = 0$.

El estado global de la red es un vector:

$$\mathbf{x} = (x_1, x_2, \ldots, x_N)$$

1.1. Regla de Hebb para el aprendizaje

Si se desea almacenar un conjunto de patrones $\{\xi^\mu\}$ con $\mu = 1, \ldots, P$, cada patrón $\xi^\mu \in \{-1, +1\}^N$, los pesos sinápticos se definen como:

$$w_{ij} = \frac{1}{N} \sum_{\mu=1}^{P} \xi_i^\mu \xi_j^\mu \quad \text{con } w_{ii} = 0$$

Esta fórmula garantiza que los patrones almacenados sean puntos de equilibrio del sistema.

1.2. Dinámica de actualización

La red evoluciona asíncronamente (o sincrónicamente) según la siguiente regla de actualización para cada neurona i:

$$x_i(t+1) = \text{sign}\left(\sum_{j=1}^{N} w_{ij} x_j(t) \right)$$

La función signo se define como:

$$\text{sign}(z) = \begin{cases} +1 & \text{si } z > 0 \\ -1 & \text{si } z < 0 \\ x_i(t) & \text{si } z = 0 \text{ (sin cambio)} \end{cases}$$

Este proceso se repite hasta alcanzar un estado estable (mínimo de energía).

1.3. Energía del sistema

La red minimiza una función de energía definida como:

$$E(\mathbf{x}) = -\frac{1}{2} \sum_{i \neq j} w_{ij} x_i x_j$$

El proceso de actualización garantiza que la energía del sistema decrece o se mantiene constante en cada iteración.

2. Capacidad de almacenamiento

La red puede almacenar correctamente hasta aproximadamente $0{,}138N$ patrones ortogonales sin errores de recuperación. Más allá de ese límite, pueden aparecer errores debidos a interferencias (efecto crosstalk).

3. Aplicación en recuperación de claves binarias

En este proyecto, una clave binaria (por ejemplo, de 128 bits) puede representarse como un patrón $\mathbf{x} \in \{-1, +1\}^{128}$. La red de Hopfield se entrena con esa clave y, tras añadir ruido al patrón (invertir un porcentaje de bits), se prueba si la red es capaz de recuperar la clave original.

Ejemplo:

- Patrón original: clave de 128 bits convertida a $\{-1, +1\}$
- Patrón ruidoso: mismo vector con 20 % de los bits invertidos
- Patrón recuperado: resultado tras evolución de la red

La recuperación se evalúa comparando coincidencias exactas entre patrón original y recuperado.

4. Referencias básicas

- J.J. Hopfield, “Neural networks and physical systems with emergent collective computational abilities”, *PNAS*, vol. 79, no. 8, pp. 2554–2558, 1982.
- Hertz, Krogh, Palmer, *Introduction to the Theory of Neural Computation*, Addison-Wesley, 1991.
- Haykin, S. *Neural Networks and Learning Machines*, Pearson, 2008.