Almacenamiento de memoria y transiciones de fase en una red de Hopfield

5 de mayo de 2019

Índice

1. Red de Hopfield

La red de Hopfield es un modelo de red neuronal propuesto por Hopfield en 1982 [1]. Se trata de una red formada por N neuronas, las cuales sólo pueden tener dos estados: $\sigma_i = +1$ ("encendido") o $\sigma_i = -1$ ("apagado"), los cuales evolucionan en el tiempo, de forma discreta con pasos temporales finitos Δt. Estas neuronas interaccionarán entre sí con pesos J_{ij}, de modo que el potencial sobre una de las neuronas debido a su interacción con las demás es:

$$\Phi_i(t) = \sum_{j \neq i} J_{ij}\sigma_j$$

El fin de estas redes es el almacenamiento y la reproducción de patrones. Podemos definir un patrón μ como una configuración dada de la red, $\mathcal{P}_i^\mu = \pm 1, i \in [1, N]$. Entonces, considerando que la evolución de los estados neuronales se da en pasos temporales finitos Δt, la red habrá reproducido correctamente el patrón μ si $\sigma_i(t) = \sigma_i(t + \Delta t) = \mathcal{P}_i^\mu$; o dicho de otro modo, los patrones deberán ser puntos fijos de la dinámica. En general, consideraremos que la red almacena un número K de patrones, de modo que $\mu \in [1, K]$. También consideraremos que los patrones sean ortogonales, de modo que:

$$\frac{1}{N}\sum_{i=1}^{N} \mathcal{P}_i^\mu \mathcal{P}_i^\nu = \delta^{\mu\nu} \tag{1.1}$$

1.1. Analogía con sistemas magnéticos. Hamiltoniano de Ising

Este modelo es muy similar al modelo de Ising para un sistema magnético. Acogiéndonos a esta analogía, bajo la restricción de simetría en las interacciones ($J_{ij} = J_{ji}$), podemos utilizar el hamiltoniano del modelo de Ising (a campo magnético $\vec{B} = 0$, situación para la que se puede observar la transición ferromagnético-paramagnético) para la dinámica de un sistema de espines σ_i a una temperatura T, para describir la dinámica de nuestra red de Hopfield.

$$H = \frac{1}{2}\sum_i \sum_{j \neq i} J_{ij}\sigma_i\sigma_j \tag{1.2}$$

dada la utilidad de esta analogía, por comodidad y abusando del lenguaje, llamaremos "espines" a los estados de las neuronas σ_i que hemos definido anteriormente.

Podemos cuantificar la similitud entre el estado en el que se encuentra la red σ_i y el patrón que buscamos reproducir $\mathcal{P}_i^\mu$ definiendo el solapamiento:

$$m^\mu(t) = \frac{1}{N}\sum_i \mathcal{P}_i^\mu \sigma_i(t) \tag{1.3}$$

que toma el valor máximo $m^\mu = 1$ cuando la red reproduce exactamente el patrón, es decir, $\sigma_i = \mathcal{P}_i^\mu \ \forall i$.

Estos estados estacionarios vienen determinados por la forma de los pesos de la interacción J_{ij}, es decir, son los responsables del almacenamiento de los patrones $\mathcal{P}_i^\mu$. Recordemos que hemos elegido el caso simétrico, $J_{ij} = J_{ji}$. Estos pesos pueden interpretarse como elementos de una matriz. Si queremos que nuestra red sea capaz de reproducir (recordar) un número K de patrones $\mathcal{P}_i^\mu$ previamente almacenados, buscamos que los J_{ij} establezcan una correlación entre los estados y el patrón. Para ello, la forma más simple que pueden tomar es:

$$J_{ij} = \frac{1}{N}\sum_{\mu=1}^{K} \mathcal{P}_i^\mu \mathcal{P}_j^\mu \tag{1.4}$$

dónde N es el número de neuronas (o espines) en la red; en el límite termodinámico (LT), $N \to \infty$ mientras que para una red finita D-dimensional de lado L, $N = L^D$. Recordemos que $\mathcal{P}_i^\mu = \pm 1$, y que los patrones ortogonales de modo que los $\mathcal{P}_i^\mu$ satisfacen (1.1).

2. Generalización del modelo de Hopfield

Hasta ahora hemos visto que la red de Hopfield admite ser interpretada como un modelo de Ising, por su analogía con un sistema magnético de espines que sólo pueden tener dos orientaciones posibles. Además, la red de Hopfield sigue una dinámica dónde, empezando desde una configuración arbitraria, el sistema evoluciona a través de una secuencia de cambios de espín (que involucra aquellos espines desalineados con sus campos moleculares). En esencia, esta es una dinámica de Monte Carlo o Glauber, dónde el proceso produce un decrecimiento monótono del hamiltoniano (1.2), es decir, disminuye la energía. Sin embargo, hasta ahora hemos tratado un sistema sin ruido; en un sistema con ruido se pueden llegar a configuraciones fuera de la dinámica que estamos tratando. Para solventar este problema, se introduce una temperatura efectiva T que caracteriza el nivel de ruido en el sistema [4]. La generalización natural del modelo a un sistema con dicho ruido es adoptar la dinámica de espines de Glauber a una temperatura finita T, dónde la distribución de las configuraciones se relaja a una distribución de Gibbs.

$$P(\{\sigma\}) \propto e^{-H(\{\sigma\})/T}$$

Este modelo se conoce como el modelo de Hopfield generalizado. En este nuevo caso, la estabilidad de un estado a todos los cambios de espín individuales no es suficiente para la estabilidad dinámica a temperaturas finitas.

2.1. Teoría de campo medio para el modelo

Ahora, estudiemos el hamiltoniano (1.2) con los pesos de interacción (1.4), para un número de patrones o memorias K finito. Posteriormente tomaremos el LT, $N \to \infty$. La densidad de energía libre será:

$$f(T) = -\frac{T}{N} \langle \log Z \rangle_{\mathcal{P}} \tag{2.1}$$

dónde $\langle \cdot \rangle_{\mathcal{P}}$ es el promedio sobre la distribución de patrones $\{\mathcal{P}_i^\mu\}$ y Z es la función de partición.

$$Z = \sum_{\{\sigma\}} e^{-H(\{\sigma\})/T} = \left(\frac{N}{T}\right)^{\frac{K}{2}} e^{-\frac{K}{2T}} \int \prod_\mu \frac{dm^\mu}{\sqrt{2\pi}} \exp\left\{ -\frac{N\vec{m}^2}{2T} + \sum_i \log\left[2\cosh\left(\frac{\vec{m}\cdot\vec{\mathcal{P}}_i}{T}\right)\right]\right\} \tag{2.2}$$

dónde hemos introducido la cantidad m^μ, y la notación vectorial $\vec{m}$ y $\vec{\mathcal{P}}_i$, cuyas componentes son los K valores de ambas cantidades. $\vec{m}$ será el parámetro de orden. Si K es finito, la integral estará predominada por su valor de punto de silla, con lo que tenemos:

$$-\frac{T\ln Z}{N} = \frac{1}{2}\vec{m}^2 - \frac{T}{N}\sum_i \ln\left[2\cosh\left(\frac{\vec{m}\cdot\vec{\mathcal{P}}_i}{T}\right)\right] \tag{2.3}$$

y el parámetro de orden viene dado por la ecuación $\partial \ln Z/\partial m^\mu = 0$.

$$\vec{m} = \frac{1}{N}\sum_i \vec{\mathcal{P}}_i \tanh\left(\frac{\vec{m}\cdot\vec{\mathcal{P}}_i}{T}\right) \tag{2.4}$$

para N finito el lado derecho de estas ecuaciones depende de la distribución concreta de patrones, pero en el LT las fluctuaciones se suprimen y las cantidades $\ln Z$ y $\vec{m}$ están autopromediadas; por tanto, las sumas $(1/N)\sum_i$ se convierten en promedios sobre la distribución $\vec{\mathcal{P}}_i$, de dónde obtenemos las ecuaciones de campo medio:

$$f(T) = \frac{1}{2}\vec{m}^2 - T\left\langle \log\left[2\cosh\left(\frac{\vec{m}\cdot\vec{\mathcal{P}}_i}{T}\right)\right]\right\rangle_{\mathcal{P}} \tag{2.5}$$

$$\vec{m} = \left\langle \vec{\mathcal{P}}_i \cdot \tanh\left(\frac{\vec{m}\cdot\vec{\mathcal{P}}_i}{T}\right)\right\rangle_{\mathcal{P}} \tag{2.6}$$

si ahora introducimos el promedio térmico de los espines:

$$\langle\sigma_i\rangle = \tanh\left(\frac{\vec{m}\cdot\vec{\mathcal{P}}_i}{T}\right) \tag{2.7}$$

introduciendo (2.7) en (2.6) obtenemos:

$$m^\mu = \langle \mathcal{P}_i^\mu \langle\sigma_i\rangle\rangle_{\mathcal{P}} \tag{2.8}$$

comparando (2.8) con (1.3), vemos que el parámetro de orden $\vec{m}$ es el solapamiento promedio. De hecho, esta comparación es más fácil con la ecuación (2.6), antes de eliminar el sumatorio que también aparece en (1.3). Aquí podemos ver una analogía con el modelo de Ising. En él, el parámetro de orden –la magnetización– para campo medio tiene la siguiente forma:

$$\tilde{m} = \langle\sigma_i\rangle$$

que es similar a la forma que ha tomado nuestro parámetro de orden, el solapamiento, en (2.8). Veremos que esta analogía se puede extender más en el siguiente apartado.

2.2. Distribución de los patrones. Estados de Mattis

Estudiemos ahora la distribución de los estados que conforman los patrones. Cómo ya hemos establecido en la sección 1, un espín i que forma parte del patrón μ sólo puede tomar dos valores: $\mathcal{P}_i^\mu = +1$ o $\mathcal{P}_i^\mu = -1$. Entonces, sabemos inmediatamente que hay un 50 % de probabilidad de que $\mathcal{P}_i^\mu$ se encuentre en el primer caso, y otro 50 % de que se encuentre en el segundo. Esto nos da directamente la distribución para un único espín del patrón, que matemáticamente podemos escribir cómo:

$$p(\mathcal{P}_i^\mu) = \frac{1}{2}\left[\delta(\mathcal{P}_i^\mu + 1) + \delta(\mathcal{P}_i^\mu - 1)\right] \tag{2.9}$$

Con este resultado en mano, podemos escribir la distribución para los K patrones (o lo que es lo mismo, para $\vec{\mathcal{P}}_i$) cómo:

$$P(\{\mathcal{P}_i^\mu\}) = \prod_{\mu,i} p(\mathcal{P}_i^\mu) \tag{2.10}$$

Ahora, el desarrollo en serie de potencias de $\vec{m}$ de las ecuaciones (2.5) y (2.6) es:

$$f = -T\log 2 + \frac{1}{2}(1-\beta)\vec{m}^2 + o(\vec{m}^4) \tag{2.11}$$

$$m^\mu = \beta m^\mu + \frac{2}{3}\beta^3 (m^\mu)^3 - \beta^3 m^\mu \vec{m}^2 + o(\vec{m}^4) \tag{2.12}$$

con $\beta \equiv 1/T$ (trabajamos en unidades $k_B = 1$). Vemos en (2.11) que para $T = 1$, $f = -T \log 2$; y despreciando en el término cúbico en m^μ a la derecha podemos escribir una versión aproximada de (2.12):

$$m^\mu \approx (\beta - \beta^3 \vec{m}^2) m^\mu$$

Si evaluamos esta aproximación en $T = 1$, o lo que es igual, $\beta = 1$, tenemos que la única solución posible es $\vec{m} = 0$ (estado "paramagnético"), que implica de nuevo por la ecuación (2.11) $f = -T \log 2$. Para $T > 1$, $\beta < 1$, por lo que la aproximación que hemos tomado en (2.12) tiene mayor validez y por tanto podemos concluir que la única solución sigue siendo $\vec{m} = 0$, y $f = -T \log 2$. Alternativamente, podemos argumentar que sí $T >> 1$, $\beta << 1$ y directamente por (2.12) $m^\mu \approx 0 \ \forall \mu$, lo que implica $\vec{m} = 0$.

Esta solución se vuelve inestable sin embargo, por debajo de la temperatura crítica $T_c = 1$, dónde aparecerán soluciones con m^μ no nulas. Por ello definiremos la dimensionalidad de $\vec{m}$, n: el número de componentes no nulas de $\vec{m}$ para una solución con $T < 1$. Estamos ahora observando un comportamiento crítico; sin embargo, esto no debe sorprendernos pues el modelo que estamos utilizando un modelo inspirado en el modelo de Ising, cuyo fin original es describir sistemas magnéticos donde efectivamente hay un comportamiento crítico, la transición ferromagnético-paramagnético. Se puede ver en (2.6) y (2.9) que alterar el orden o cambiar el signo de las n componentes no nulas no altera las soluciones.

Veamos ahora el caso $n = 1$. Este caso es especialmente interesante porque nos devolverá las ecuaciones del modelo de Ising para sistemas magnéticos en campo medio. Supongamos $m^1 = m \neq 0$ y $m^\mu = 0 \ \forall \mu > 1$. Entonces:

$$f = \frac{1}{2} m^2 - T \log \left[2 \cosh(\beta m) \right] \tag{2.13}$$

$$m = \tanh(\beta m) \tag{2.14}$$

Estas soluciones corresponden a un estado dónde:

$$\langle \sigma_i \rangle = \mathcal{P}_i^1 \tanh(\beta m) \tag{2.15}$$

si introducimos esta solución en (2.8) tenemos $m = \langle (\mathcal{P}_i^1)^2 \tanh(\beta m) \rangle_{\mathcal{P}}$. Cómo $\mathcal{P}_i^\mu = \pm 1$, esto se convierte en $m = \langle \tanh(\beta m) \rangle_{\mathcal{P}}$ y, por último, al no aparecer los estados de los patrones, el promedio desaparece. Hemos recuperado (2.14).

El estado (2.15) es equivalente termodinámicamente al estado ferromagnético del modelo de Ising. Existen $2K$ estados de este tipo correspondientes a distintas μ y a distintos signos de las componentes m^μ. Estos estados se denominan estados de Mattis.

3. Ecuación de Langevin para la dinámica de los espines

En la sección 1, hicimos un breve comentario sobre la dinámica de los espines, donde establecimos que si un estado es parcialmente similar a aquellos que forman el patrón, la dinámica llevará los espines a dichos estados estacionarios. Feigelman et. al. [3] proponen la ecuación de Langevin que gobierna esta dinámica.

El primer paso que tomaremos será añadir una modificación al hamiltoniano en forma de un término con un nuevo potencial $V(\sigma_i)$ para cada espín; lo que en realidad nos interesa no es este potencial en sí, si no su derivada, cuyo fin se hará vigente en las ecuaciones del movimiento. Nuestro hamiltoniano es entonces:

$$H = \sum_i V(\sigma_i) + \frac{1}{2} \sum_i \sum_{j \neq i} J_{ij} \sigma_i \sigma_j \tag{3.1}$$

Las ecuaciones de movimiento dadas por (3.1) serán equivalentes a las ecuaciones de Langevin para la evolución de los espines σ_i. Estas son [3]:

$$\frac{\partial \sigma_i}{\partial t} = -\frac{\partial V(\sigma_i)}{\partial \sigma_i} + \sum_j J_{ij}\sigma_i + h_i\sigma_i + \xi_i(t) \tag{3.2}$$

En esta dinámica gobernada por (3.2) los patrones serán estados estables con un rango de atracción en el espacio de configuración, de modo que si un estado inicial $\tilde{\sigma}_i$ es parcialmente similar a uno de los estados estacionarios que σ_i conforman el patrón, la dinámica llevará $\tilde{\sigma}_i$ a σ_i; este es el mecanismo mediante el cual la red recuerda el patrón.

Estudiemos ahora los distintos términos de (3.2). En estas ecuaciones aparece la derivada del potencial que hemos introducido anteriormente, $\partial V(\sigma_i)/\partial\sigma_i$. La razón de ser de este término es asegurar que (al menos casi siempre) el valor dado por esta ecuación para los espines sea $\sigma_i = \pm 1$. Una posible elección de este potencial es $V(\sigma_i) = \lambda\left(\sigma_i^2 - 1\right)^2$. Cuando el potencial toma esta forma, resolver la dinámica nos lleva al resultado:

$$\sigma_i = \frac{\pm 1}{\sqrt{1 - e^{-2\lambda t}}} \xrightarrow[t\to\infty]{} \pm 1$$

es decir, vemos que en el límite asintótico de tiempos infinitos, los valores de los espines son los esperados, ± 1. Si se utilizan otros potenciales (por ejemplo cuadráticos en vez de cuárticos) se puede comprobar que este comportamiento puede no replicarse, por eso es necesario introducir el potencial para que la dinámica sea coherente con las reglas del modelo. Por otro lado, el parámetro λ debe ser positivo; en el caso de que fuera negativo los espines convergerían a 0, y en el caso de que fuera $\lambda = 0$ (no hay potencial) los valores crecen sin control. Pero también necesitamos que tenga valores grandes respecto a 1: se puede comprobar que el valor al que convergen los espines para un valor de λ dado es:

$$|\sigma_i|_{estacionario} = \sqrt{1 + \frac{1}{\lambda}}$$

por tanto necesitamos que, como mínimo, $\lambda >> 1$, o preferiblemente $\lambda \to \infty$. Este hecho cobrará importancia cuando posteriormente llevemos a cabo simulaciones numéricas con la dinámica gobernada por (3.2).

En efecto, si consideramos que los espines han convergido al patrón $\mathcal{P}_i^\nu$, tenemos que $\sigma_i = \sqrt{1 + \frac{1}{\lambda}}\mathcal{P}_i^\nu$. Imponiendo que el estado sea estacionario en (3.2), a $T = 0$ obtenemos:

$$\sum_{j=1}^{N} J_{ij}\sigma_j - \lambda(\sigma_i^2 - 1)\sigma_i = \sqrt{1 + \frac{1}{\lambda}}\left(\sum_{\mu=1}^{K} \mathcal{P}_i^\mu \frac{1}{N}\sum_{j=1}^{N} \mathcal{P}_j^\mu \mathcal{P}_j^\nu - \mathcal{P}_i^\mu\right) = 0$$

Por otro lado, hemos introducido el ruido blanco gaussiano $\xi_i(t)$, que satisface $\langle \xi_i(t)\xi_j(t')\rangle = 2T\delta_{ij}\delta(t - t')$, dónde T es la temperatura efectiva que se introduce en el modelo de Hopfield generalizado que hemos comentado en la sección 2, como representante del ruido de fondo en el sistema.

Por último, tomaremos como pesos de interacción J_{ij} los definidos en (1.4). Son la forma más simple que pueden tomar para reproducir la dinámica deseada.

En el siguiente apartado utilizaremos esta dinámica para, mediante métodos numéricos, examinar como la red (a una temperatura fija dada T) evoluciona hacia un patrón dado externamente desde un estado inicial aleatoriamente desordenado, siguiendo la dinámica dada por (3.2). En la figura 1 se muestra un ejemplo.

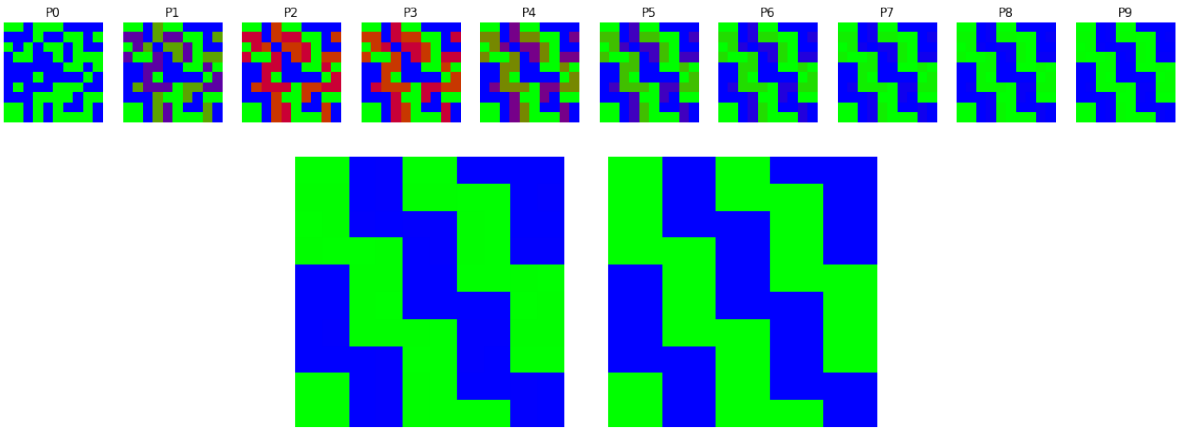

Figura 1: Arriba, la evolución de la red desde el estado desordenado P0 hasta el estado final P9. Abajo, el estado final de la red (izquierda) junto al patrón almacenado (derecha). Vemos que la red ha conseguido reproducir exitosamente el patrón.

3.1. Simulación numérica de la dinámica

En esta sección presentamos una simulación numérica llevada a cabo en Python, utilizando el paquete `neurodynex` proporcionado en [6].

Para poder tratar el problema de este modo, el primer paso que hemos de tomar es traducir la ecuación diferencial (3.2) a una de diferencias finitas. Como ya comentamos en la sección 1, la dinámica de la red neuronal evoluciona por pasos temporales finitos, la dinámica de decaimiento exponencial debe ser rápida (en tiempos de orden 1), mientras que la dinámica debida al modelo de Ising debe ser lenta.

$$\sigma_i(t+dt) - \sigma_i(t) = \left(-4\lambda(\sigma_i^2 - 1)\sigma_i + \sum_j J_{ij}\sigma_j + h_i\sigma_i \right) dt + \sqrt{2T}dW$$

Absorbiendo el 4 en el parámetro λ y despejando, obtenemos:

$$\sigma_i(t+dt) = g(\sigma_i(t)) = \left(-\lambda(\sigma_i^2 - 1)\sigma_i + \sum_j J_{ij}\sigma_j + h_i\sigma_i \right) dt + \sqrt{2T}dW + \sigma_i(t) \qquad (3.3)$$

de forma que $g(\sigma_i(t))$ será la función dinámica que utilicemos en el cálculo. Para convertir la ecuación de Langevin en una ecuación diferencial estocástica, hemos introducido el diferencial estocástico dW que cumple:

$$\mu(dW) = 0, \qquad \sigma^2(dW) = dt$$

Referencias

[1] J. J. Hopfield, *Neural networks and physical systems with emergent collective computational abilities*, Proc. Natl. Acad. Sci. USA, **79**, 2554-2558, abril de 1982.

[2] W. Gerstner, W. M. Kistler, R. Naud, L. Paninski, *Neuronal Dynamics*, Cambridge University Press, 2014, https://neuronaldynamics.epfl.ch/online/index.html.

[3] M. V. Feigelman, L. B. Ioffe, *The Statistical Properties of the Hopfield Model of Memory*, Europhys. Lett., **1** (4), 197-201, febrero de 1986.

[4] D. J. Amit, H. Gutfreund, H. Sompolinsky, *Spin-glass models of neural networks*, Physical Review A, **32** (2), 1007-1018, agosto de 1985.

[5] D. J. Amit, H. Gutfreund, H. Sompolinsky, *Storing Infinite Numbers of Patterns in a Spin-Glass Model of Neural Networks*, Physical Review Letters, **55** (14), 1530-1533, septiembre de 1985.

[6] W. Gerstner, W. M. Kistler, R. Naud, L. Paninski, *Neuronal Dynamics: Python Exercises*, https://neuronaldynamics-exercises.readthedocs.io/en/latest/index.html.