Redes Recurrentes parte 1

El modelo de Hopfielfd

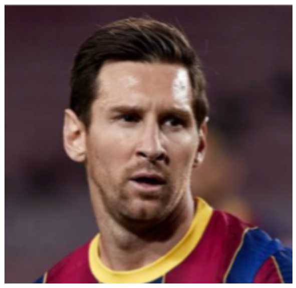
John Hopfield (1982)

En el problema de Memoria Asociativa queremos almacenar información y recuperarla asociando el input con la memoria más parecida, donde el concepto de parecido es muy complejo.

Esto es algo que los animales hacemos muy bien pero es difícil de programar.

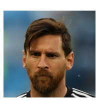
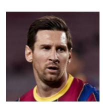
Ejemplo: supongamos que queremos hacer un programa que identifique fotos de Cristiano Ronaldo, Lionel Messi y Neymar.

Queremos almacenar 3 fotos, una por c/uno.

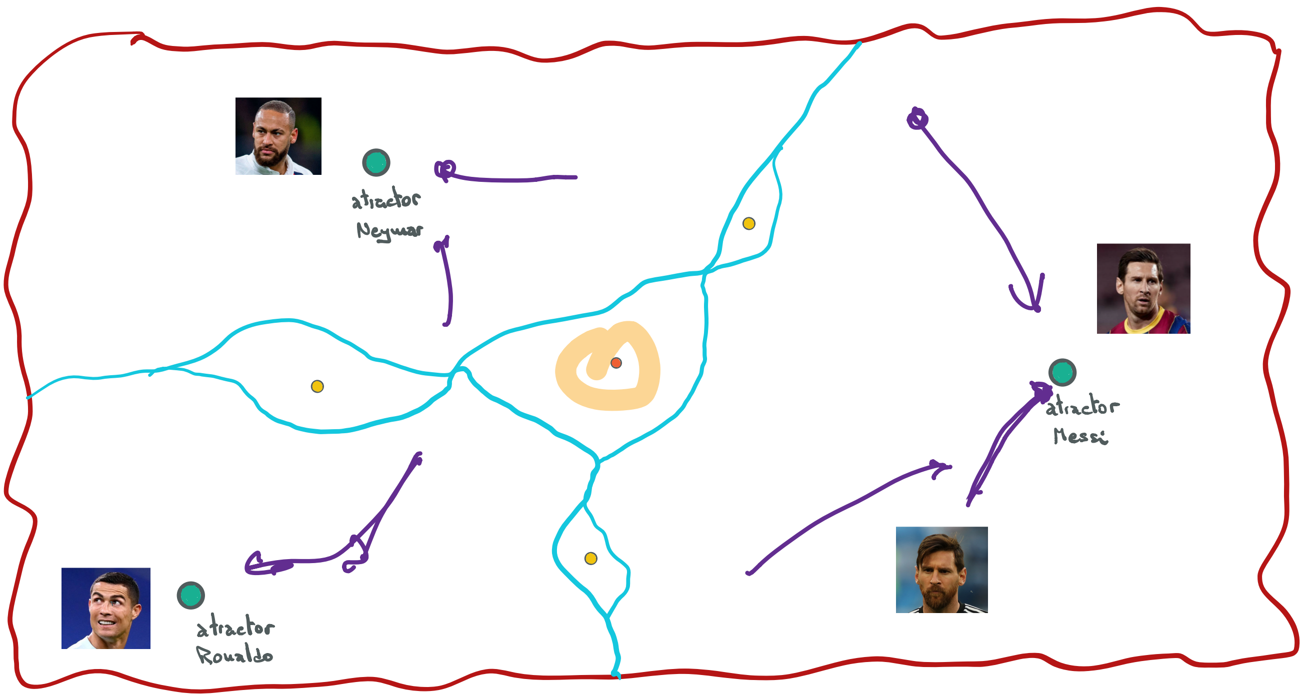
  

Si ahora le muestro esta nueva foto,

el programa debe decirnos de alguna forma que es Messi.

Elegimos un enfoque dinámico. Como la entrada  es de la misma naturaleza que la salida esperada  vamos a usar un enfoque dinámico.

Queremos que el MESSI almacenado sea un atractor de la dinámica, y que el input esté en la cuenca de atracción del MESSI almacenado



espacio donde viven TODAS las fotos posibles de Ronaldo, Messi y Neymar que son infinitas y fueron regularizadas

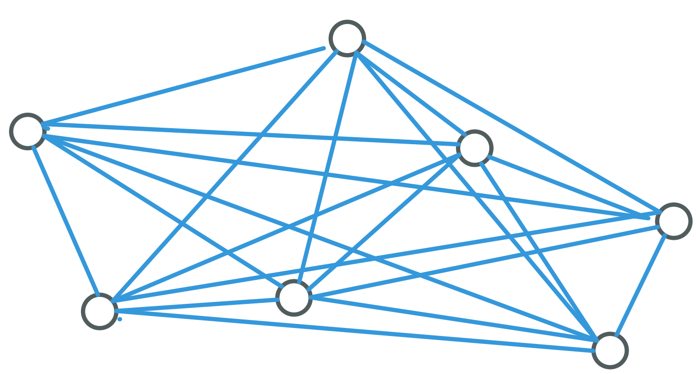
Esto fue resuelto por Hopfield en 1982 con pixeles binarios (aunque era más general que imágenes).

El propuso N neuronas binarias tipo McCulloch & Pitts sin umbrales:

$$S_i(t+\Delta t) = \text{signo}\left(h_i(t)\right) = \begin{cases} +1 & \text{si } h_i(t) \geq 0 \\ -1 & \text{si } h_i(t) < 0 \end{cases}$$

$$h_i(t) = \sum_{j \neq i}^{N} w_{ij} S_j(t)$$

Estas neuronas forman un grafo completo (cada nodo se conecta con todos los otros nodos) o fully connected (totalmente conectado) con conexiones (aristas) bidireccionales y simétricas



Si quieren los ponen en un reticulado cuadrado, pero aun así, cada neurona "recibe" y "manda" señales a todas las otras.

En cada paso de tiempo t el sistema dinámico discreto se define por un punto en el hiper espacio de la 2 centrado en el origen y en dimensión N

$$\bar{S}(t) = (S_1(t), S_2(t), \ldots, S_N(t))$$

Nota: si fuesen fotos, serían en blanco (-1) y negro (+1).

Almaceno las 3 fotos en blanco y negro.
Ponemos inicialmente a la red en un dado estado y vemos a que atractor va.

Decimos que el atractor es el concepto del ejemplo que pusimos en el inicio

$$\bar{\xi}(0) \longrightarrow \bar{\xi}^{\mu}$$

Hopfield lo fue resolviendo de a poco.

Supongamos que queremos hacer un único atractor, o sea, guardar un único concepto. Parece simple y lo es. Pero es también didáctico. Sea

$$\xi_i^1 = +1 \text{ o } -1$$

1 → 1° concepto

i → neurona i

el valor de la neurona i cuando estamos en el atractor correspondiente al concepto almacenado.

$$W_{ij} = \frac{1}{N} \xi_i^1 \xi_j^1$$

regla de Hopfield para un concepto ó memoria almacenada

Noten que $W_{ij} = W_{ji}$ (son simétricos)

Hacemos "a mano" $W_{ii} = 0 \quad \forall\, 0 \le i \le N$,

o sea no permitimos auto interacción.

Veamos qué sucede con la dinámica.

Ponemos la red inicialmente en un estado arbitrario $\vec{S}(0) = (S_1(0), \dots, S_N(0))$

$$
\begin{aligned}
S_i(1) &= \text{signo}\left(h_i(0)\right) \\
&= \text{signo}\left(\sum_{\substack{j=1 \\ j\neq i}}^{N} w_{ij} S_j(0)\right) \\
&= \text{signo}\left(\sum_{\substack{j=1 \\ j\neq i}}^{N} \frac{1}{N} \xi_i^1 \xi_j^1 S_j(0)\right) \\
&= \text{signo}\left(\xi_i^1 \sum_{\substack{j=1 \\ j\neq i}}^{N} \xi_j^1 S_j(0)\right) \\
&= \text{signo}\left(\xi_i^1\right) \text{signo}\left(\sum_{j=1}^{N} \frac{\xi_j^1 S_j(0) - \xi_i^1 S_i(0)}{N}\right) \\
&= \xi_i^1 \,\text{signo}\left(\frac{2n}{N} - 1 - \frac{1}{N}\right) \longrightarrow \text{pequeño} \\
&\simeq \xi_i^1 \,\text{signo}\left(\frac{2n}{N} - 1\right)
\end{aligned}
$$

$$\frac{n - (N-n)}{N}$$

$$\frac{n}{N} - \frac{N}{N} + \frac{n}{N} \pm \cancel{\frac{1}{N}}$$

$$\left(2\frac{n}{N} - 1\right) \pm \cancel{\frac{1}{N}}$$

$$\frac{n - (N-n)}{N} = \frac{2n}{N} - 1$$

Si $n > \frac{N}{2}$, $\frac{2n}{N} - 1 > 0$ y

$$S_i(1) = \text{signo}(h_i(t)) = \xi_i^1$$

Si $n < \frac{N}{2}$, $\frac{2n}{N} - 1 < 0$ y

$$S_i(1) = \text{signo}(h_i(t)) = -\xi_i^1$$

O sea, en 1 paso llega al atractor, sin importar de donde sale.

El espacio de los estados se divide en 2.

Una cuenca va a ξ y la otra a $-\xi$