

## Práctica 5: Aprendizaje no supervisado

Evelyn G. Coronel  
Redes Neuronales - Instituto Balseiro

(3 de mayo de 2020)

### EJERCICIO 1: DISTRIBUCIÓN GAUSSIANA

En este ejercicio, la red consiste en una capa lineal de 4 entradas y una salida. Podemos considerar a la entrada como un vector en 4 dimensiones  $\xi$ . La salida  $O$  se obtiene mediante  $O = \xi \cdot w$ , donde  $w$  representa el vector de pesos, también de 4 componentes.

Los datos de la entrada tienen una distribución de probabilidad dada por la Ec. 1

$$P(\xi) = \frac{1}{(2\pi)^2 \sqrt{\det \Sigma}} e^{-\frac{1}{2} \xi^T \Sigma \xi}, \text{ con } \Sigma = \begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{bmatrix} \quad (1)$$

donde  $\Sigma$  representa la matriz de correlación. Esta matrix  $\Sigma$  tiene 2 autovalores, 5 y 1, donde el autovalor 1 tiene degeneración 3. El autovector correspondiente al autovalor 5 es  $\lambda = [1, 1, 1, 1]^T$ .

Para la ejecución del programa que simula la red, se inicializó el vector  $w$  con valores aleatorios entre  $[-0.01, 0.1]$ . La regla de actualización de los pesos es la regla de Oja, dada por la Ec. 2.

$$\Delta w = \eta O(\xi - Ow) \quad (2)$$

con  $\eta = 0.001$  para 1000 épocas de ejecución. Durante la evolución de la red, tal como se ve en la Fig. 1, los pesos convergen a un valor cercano a 0.5. Esto implica que  $\|w\| \approx 1$ , por lo que la predicción de la regla de Oja sobre la normalización de  $w$  es consistente con la simulación.

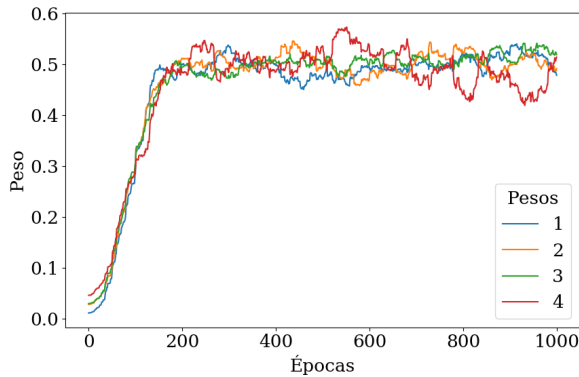


Fig. 1: Evolución en función de las épocas de los pesos de la red

Otro aspecto interesante de la evolución de los pesos, representados como puntos en la Fig. 2, donde  $w$  tiende a alinearse con el vector  $\lambda$ . Esto se debe a que la dirección predominante en la correlación es la de mayor autovalor, representada por la flecha de línea sólida en la Fig. 2. Esta figura es una proyección de la espacio de 4 dimensiones, pero tiene la misma apariencia tomando la componente  $w_4$  en vez de alguna de las graficadas.

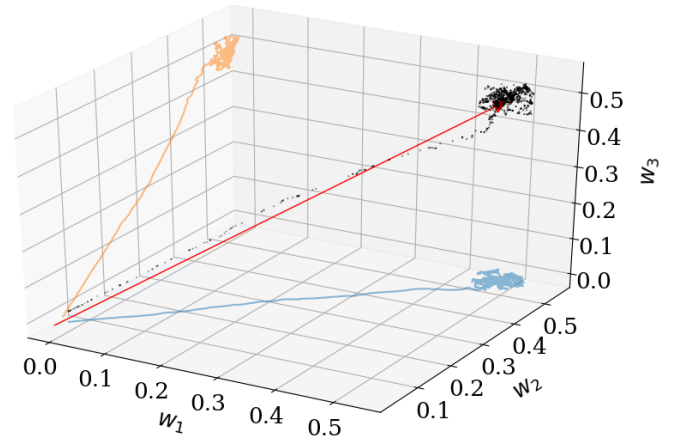


Fig. 2: Comparación de los pesos  $w_1$ ,  $w_2$  y  $w_3$ . Se observa los mismos se alinean con la proyección en el esquema tridimensional del autovector con mayor autovalor.

### EJERCICIO 2: DISTRIBUCIÓN UNIFORME EN UN SEMI-ANILLO

La red consiste en una entrada con dos neuronas y una salida de 10 neuronas, donde cada neurona de entrada está conectada a todas las neuronas de salida. Para los datos de entrada tiene una distribución uniforme en un anillo en un rango de  $r \in [0.9, 1.1]$  y  $\theta \in [0, \pi]$ , donde  $r$  y  $\theta$  representan el radio y el ángulo en coordenadas polares. Para obtener una distribución uniforme en el rango considerado, se implementó el siguiente algoritmo:

$$r^2 = (\text{random entre 0 y 1}) \times (r_{max}^2 - r_{min}^2) + r_{min}^2$$
$$\theta = (\text{otro random entre 0 y 1}) \times \pi$$

donde  $r_{min} = 0.9$  y  $r_{max} = 1.1$ . Una vez definido el punto en coordenadas polares, se alimenta a cada neurona de entrada con las coordenadas cartesianas de este punto, mediante la transformación  $x = r \cos \theta$  e  $y = r \sin \theta$ . La entrada  $\xi$  puede tratarse como un vector de dos com-

ponentes. La primera entrada es la coordenada  $x$  y la segunda es la coordenada  $y$ .

Por cada neurona de salida se tiene dos pesos, uno por cada entrada, para un total de 20 valores. Los mismos se organizan como 10 vectores de dos componentes  $w_i$ , donde la primera componente corresponde al peso de la primera entrada (coordenada  $x$ ) y lo mismo para la segunda (coordenada  $y$ ). Los pesos son inicializados tal que la componente en  $x$  sea 0.5 mientras que los valores en  $y$  varían entre  $(-1.1, 1.1)$ .

En cada época, se calcula la salida  $O_i$  mediante  $O_i = \xi_i \cdot w_i$ . Se dice que una neurona de salida  $i^*$  es la ganadora cuando  $\|\xi - w_{i^*}\|$  tiene el menor valor, en comparación a las otras neuronas  $i$ . La actualización de los pesos se realiza mediante la Ec.3

$$\Delta w_{i,j} = \eta \Lambda(i, i^*) (\xi_j - w_{i,j}) \quad (3)$$

donde  $\eta$  es la tasa de aprendizaje y  $\Lambda(i, i^*)$  es la función de inhibición. Para este problema en particular se utilizó la función dada en la Ec.4, que corresponde a una gaussiana.

$$\Lambda(i, i^*) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(i-i^*)^2}{2\sigma^2}} \quad (4)$$

La red evolucionó con  $\eta = 0.02$ , además se varió los parámetros  $\sigma$  y  $N$ , que corresponde a la dispersión de la Ec.4 y la cantidad de épocas. Se observaron distintos comportamientos en la convergencia de los pesos para distintos parámetros. En la Fig.3 se observa los pesos iniciales y finales de la red para  $\sigma = 0.6$  y distintos tiempos de evolución  $N$ . La distribución de los datos de entrada se representa con dos arcos concéntricos de radios 0.9 y 1.1, centrados en 0.

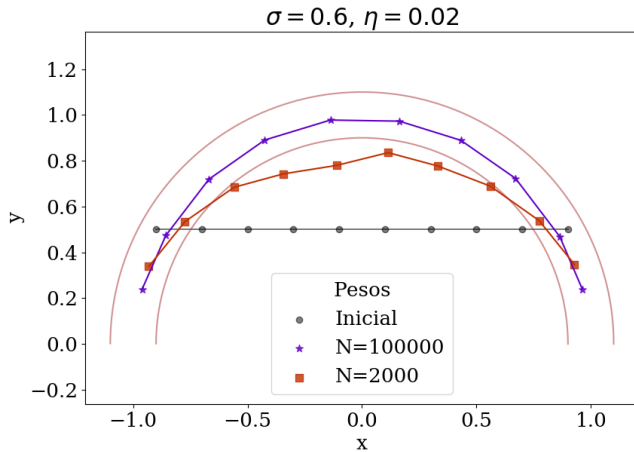


Fig. 3: Los valores finales de los pesos para una red con  $\sigma = 0.6$  y distintas cantidad de épocas. Se puede observar que los pesos copian la distribución de los datos de entrada.

Es claro que para  $N = 1\,000\,000$ , los pesos convergen a la distribución de los datos de la entrada, como es de

esperarse utilizando el algoritmo de Kohonen, dada por la Ec.3, mientras que para  $N = 2000$  el sistema aún no converge a los valores óptimos.

En cambio si el sistema evoluciona con  $\sigma = 0.3$ , como se muestra en la Fig.4, para  $N = 1\,000\,000$ , los pesos se distribuyen de forma desordenada, así como también para  $N = 2000$ . Esto se debe a que la función de inhibición inhibe fuertemente a las neuronas lejanas a la neurona ganadora.

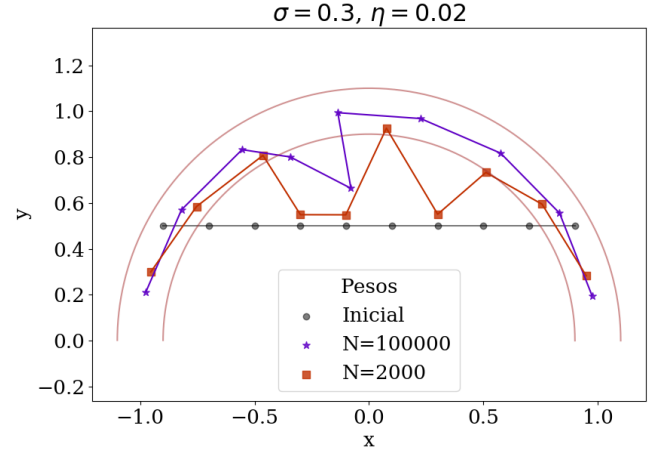


Fig. 4: Los valores finales de los pesos para una red con  $\sigma = 0.3$  y distintas cantidad de épocas.

Si relajamos  $\sigma$  hasta un valor mayor, como se muestra en la Fig.5 con  $\sigma = 2$ , el sistema sigue copiando la correlación de los datos de entrada, pero la convergencia se alcanzaría para mayores valores de  $N$  en comparación a la Fig.3.

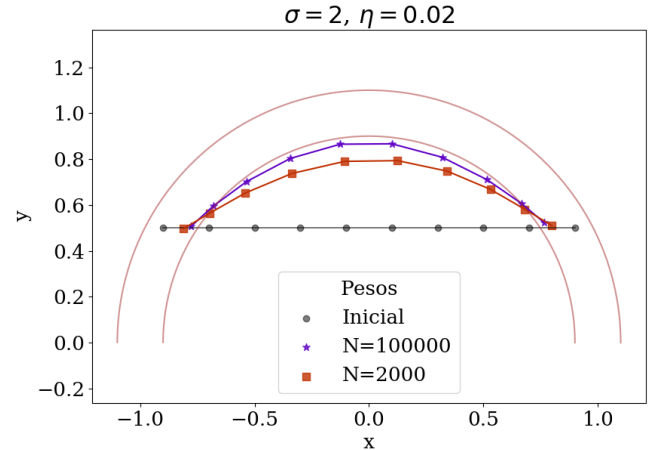


Fig. 5: Los valores finales de los pesos para una red con  $\sigma = 2$  y distintas cantidad de épocas.

Si seguimos aumentando el valor de  $\sigma$ , como se muestra en la Fig.6 con  $\sigma = 10$ , la inhibición ya es débil para forzar

a la red a copiar la correlación de los datos de entrada.

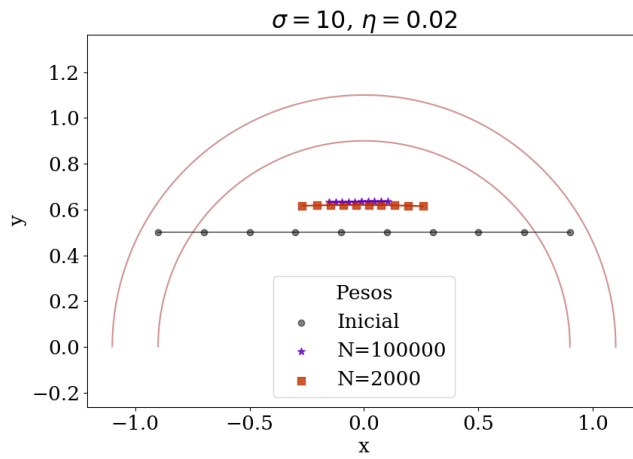


Fig. 6: Los valores finales de los pesos para una red con  $\sigma = 10$  y distinta cantidad de épocas.