

# Inteligencia Artificial

Sistemas inteligentes artificiales



# Reconocimiento de patrones y Redes neuronales artificiales



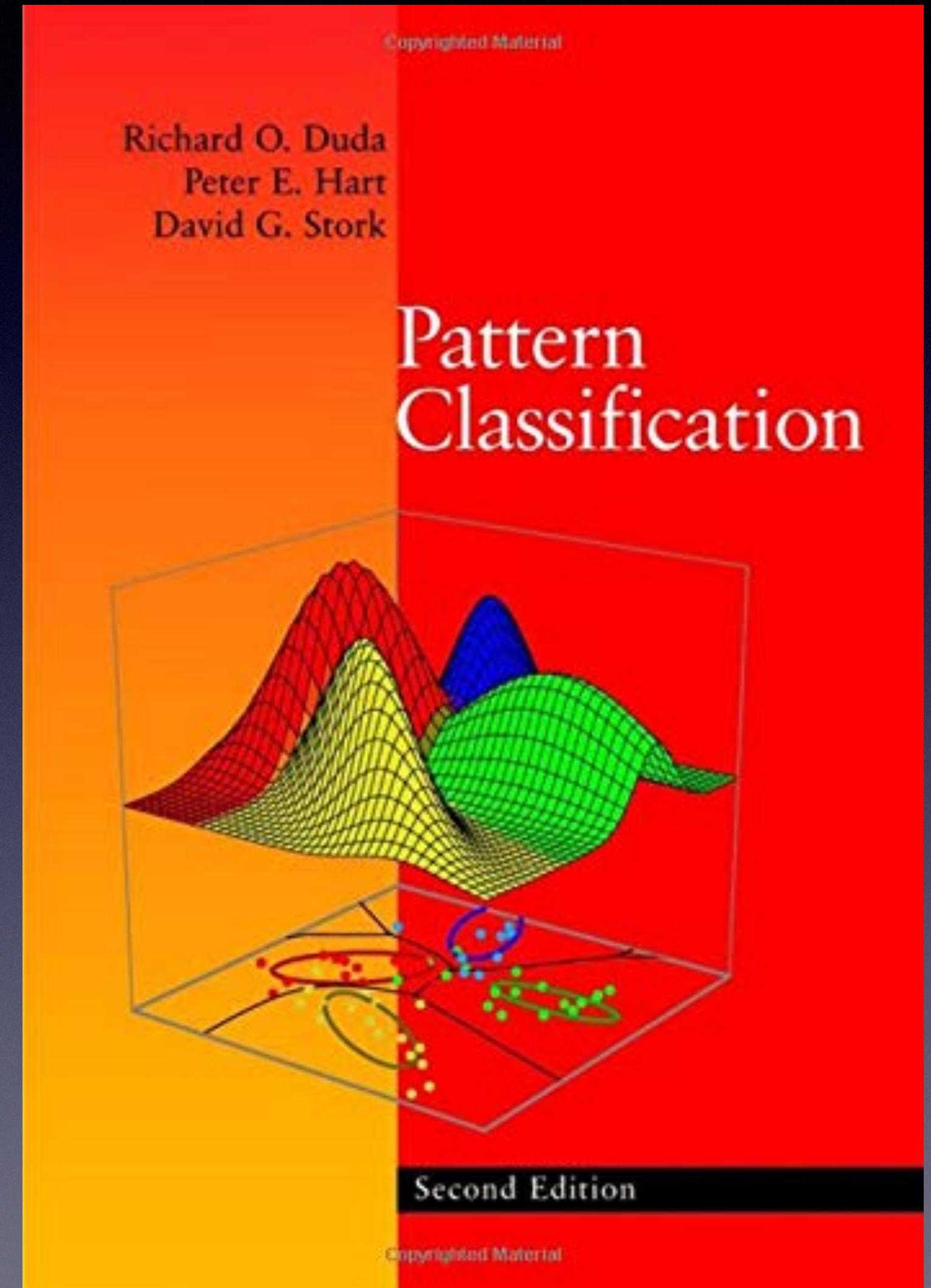
# Temas

- Fundamentos de reconocimiento de patrones
- Concepto de distancia
- Concepto de optimización
- Estructuras de redes neuronales
- Aprendizaje supervisado
- Aprendizaje NO supervisado (Clustering)



# Fundamentos de reconocimiento de patrones

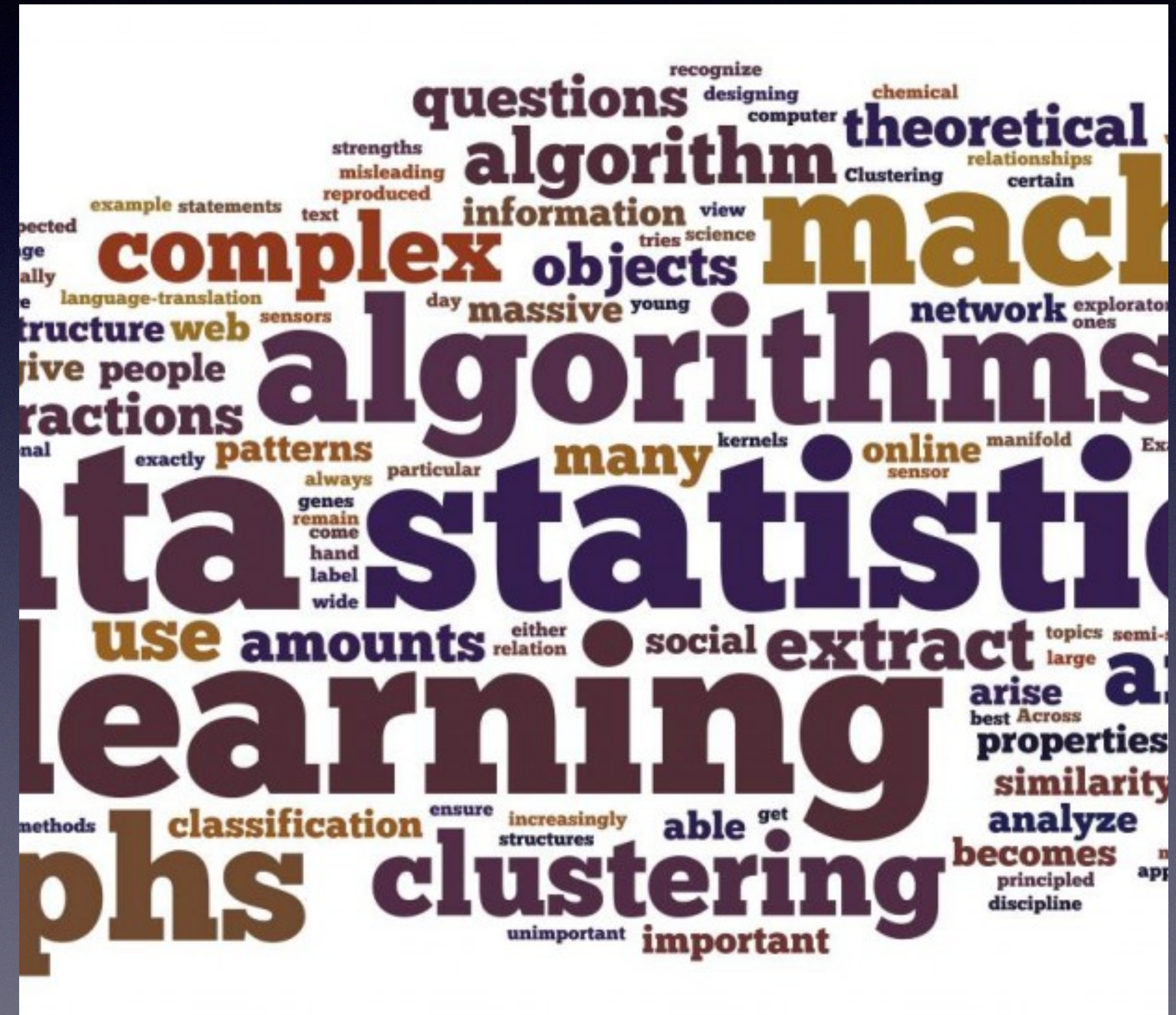
- Aprendizaje supervisado
- Aprendizaje NO supervisado
- Clustering
- Sistema recomendador
- Aprendizaje por refuerzo





# ¿Qué tipo de aprendizaje necesitas?

1. Definir bien el problema. ¿Es de ML?
2. Definir funcionalidad a implementar
  - A. Predecir una métrica var continua
  - B. Predecir una etiqueta var discreta
  - C. Buscar grupos
  - D. Optimizar por prueba y error





# Define el problema

- Tipos de variables
  - A. Continuas
  - B. Discretas
- Rangos aceptables para predicción
- Tipo de datos por área de conocimientos
- Intuición ¿Es posible hacer una predicción?





# Concepto de distancia

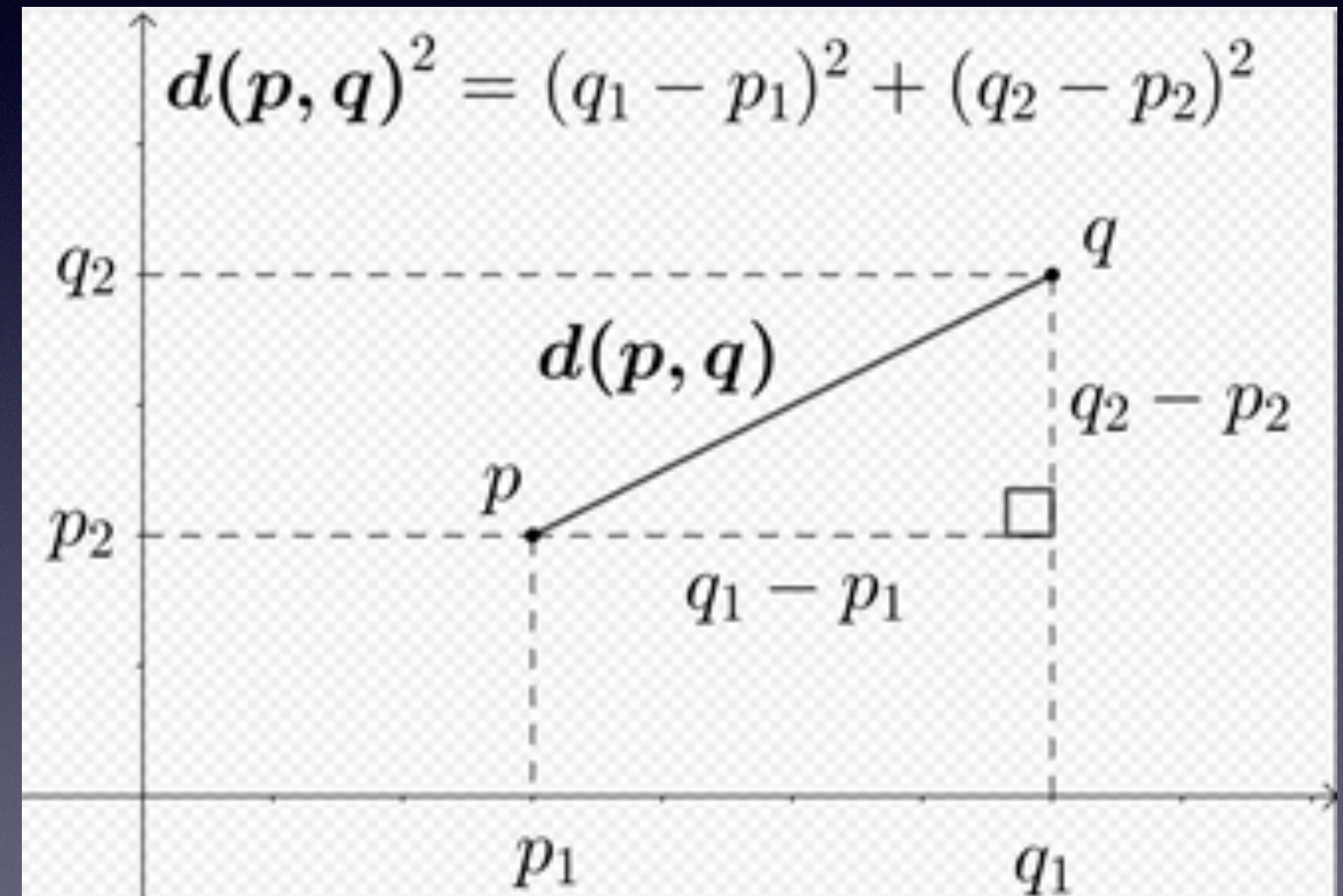
- Ya sea un algoritmo supervisado o no supervisado, las métricas de distancia juegan un papel importante en el aprendizaje automático.
- Se deben elegir diferentes medidas de distancia según los tipos de datos.





# Distancia euclidiana

- La distancia euclidiana es la distancia en línea recta entre dos puntos de datos en el espacio euclidiano.
- También se le llama como norma L2 o distancia L2.





$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

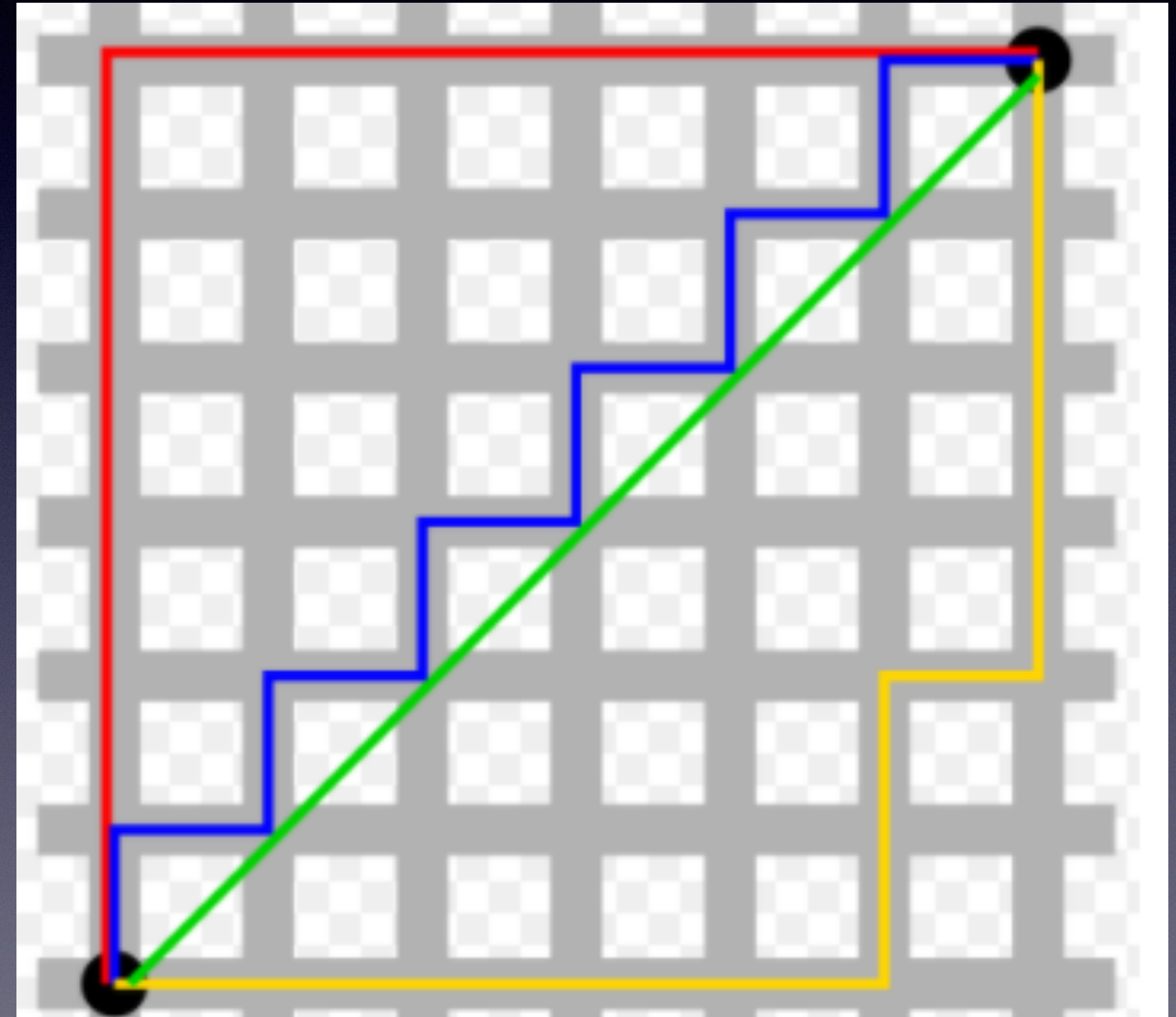
# Distancia euclidea n dimensiones

Si  $p = (p_1, p_2 \dots p_n)$  y  $q = (q_1, q_2 \dots q_n)$  son dos puntos en el espacio euclidiano, la distancia euclidiana es  $d(p, q)$



# Distancia de Manhattan

- La distancia de Manhattan entre dos puntos en dos dimensiones es la suma de las diferencias absolutas de sus coordenadas cartesianas.
- La distancia de Manhattan también se llama con diferentes nombres, como distancia rectilínea, distancia L1, norma L1, distancia de serpiente, distancia de manzana, etc.





$$d_1(p, q) = \|p - q\| = \sum_{i=1}^n \|p_i - q_i\|$$

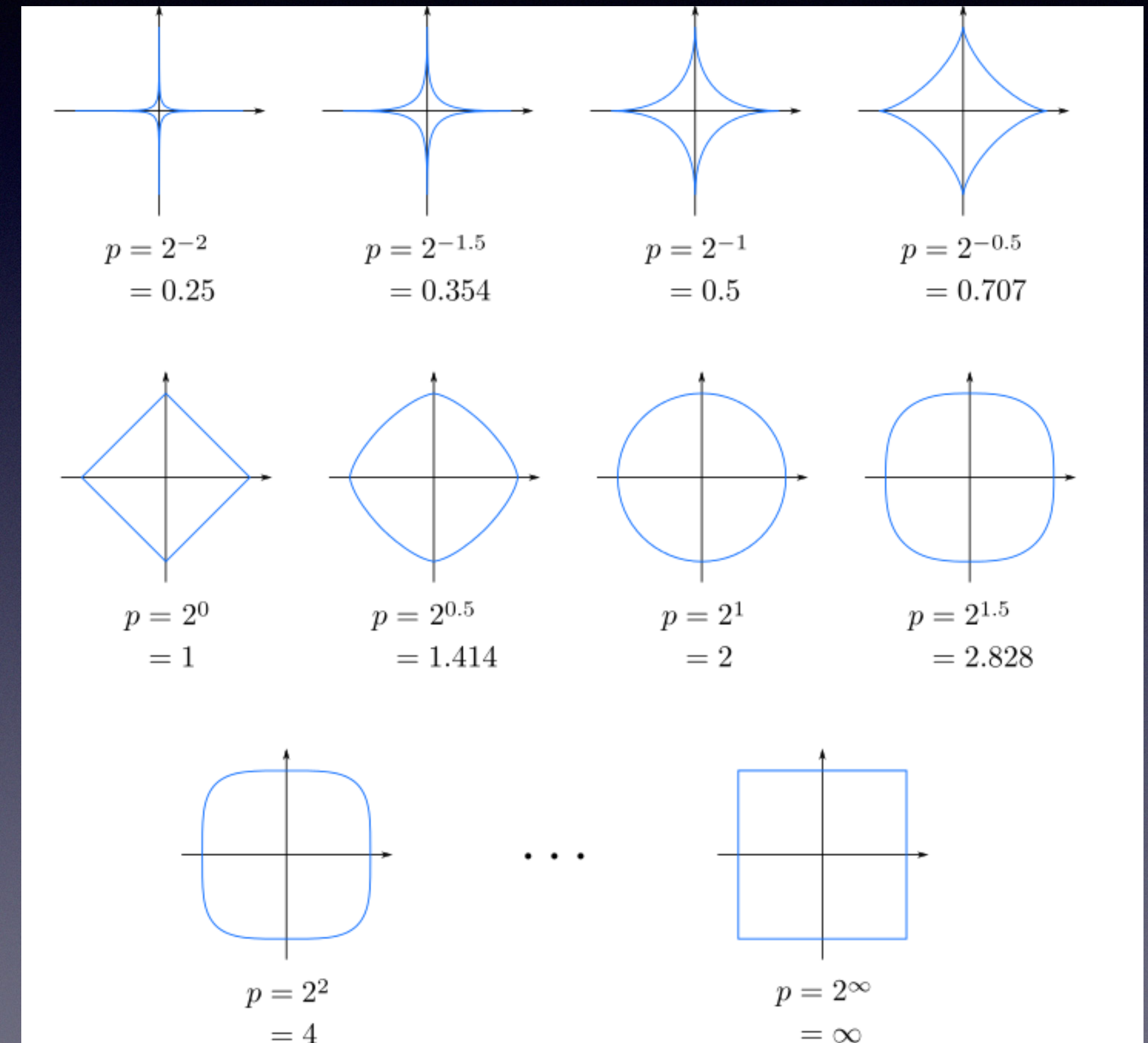
## Manhattan n dimensiones

Si  $p = (p_1, p_2, \dots, p_n)$  y  $q = (q_1, q_2, \dots, q_n)$  son dos vectores en el plano, la distancia de Manhattan nD



# Distancia de Minkowski

- La distancia de Minkowski se puede considerar como una forma generalizada tanto de la distancia euclidiana como de la distancia de Manhattan.





$$d(X, Y) = \left( \sum_{i=1}^n \|x_i - y_i\|^p \right)^{\frac{1}{p}}$$

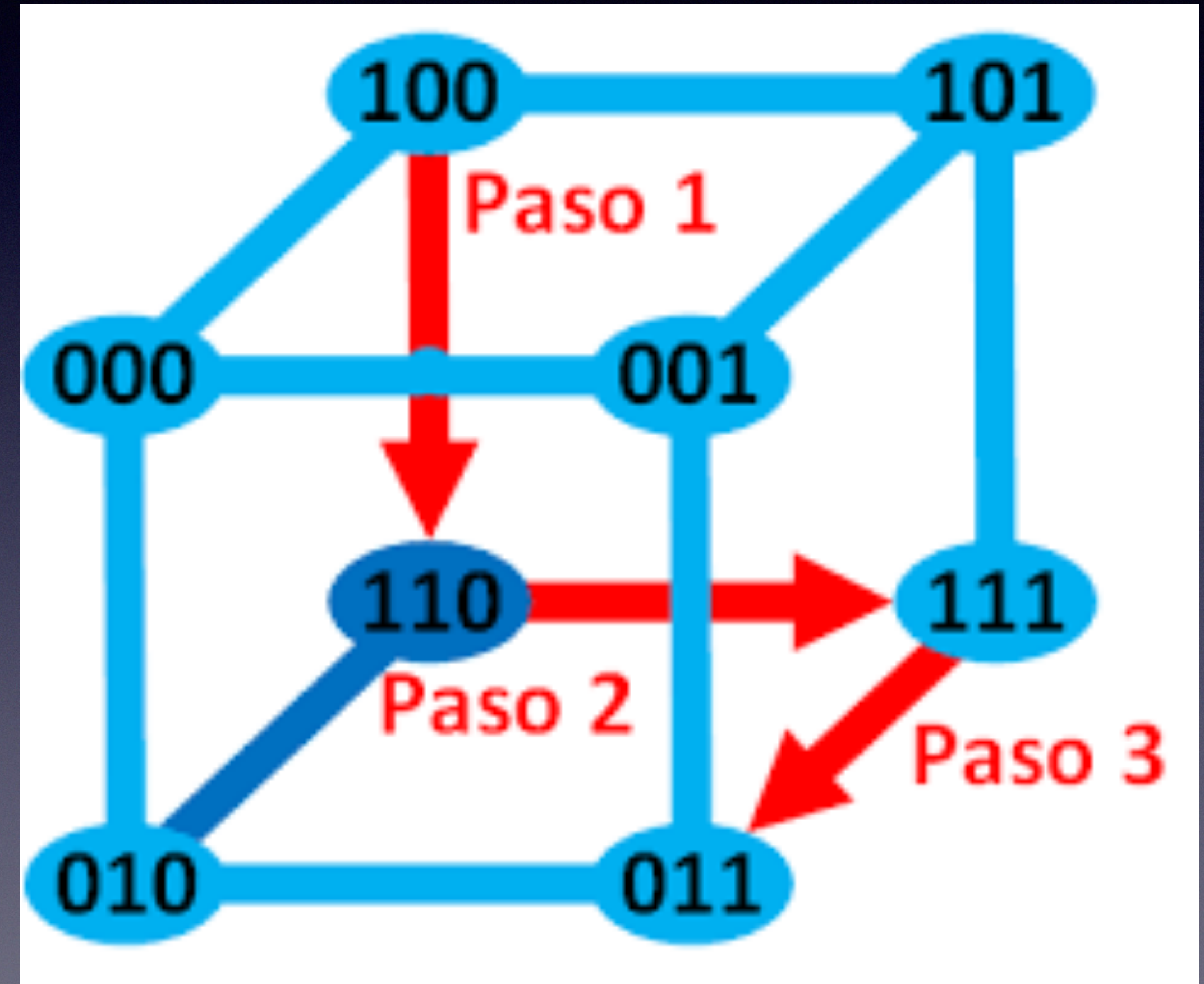
## Distancia de Minkowski

La distancia de Minkowski de orden  $p$  (donde  $p$  es un número entero) entre dos puntos  $X = (x_1, x_2 \dots x_n)$  e  $Y = (y_1, y_2 \dots y_n)$



# Distancia de Hamming

- La distancia de martillo entre dos cadenas de igual longitud es el número de posiciones en las que los símbolos correspondientes son diferentes.
- Las cadenas pueden ser letras, bits o dígitos decimales, etc.





# Distancia del coseno y similitud del coseno

- El coseno de dos vectores distintos de cero se obtiene utilizando la fórmula del producto escalar euclidiano

$$A \cdot B = \|A\| \|B\| \cos \theta$$



$$\textit{Similarity} = \cos \theta = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

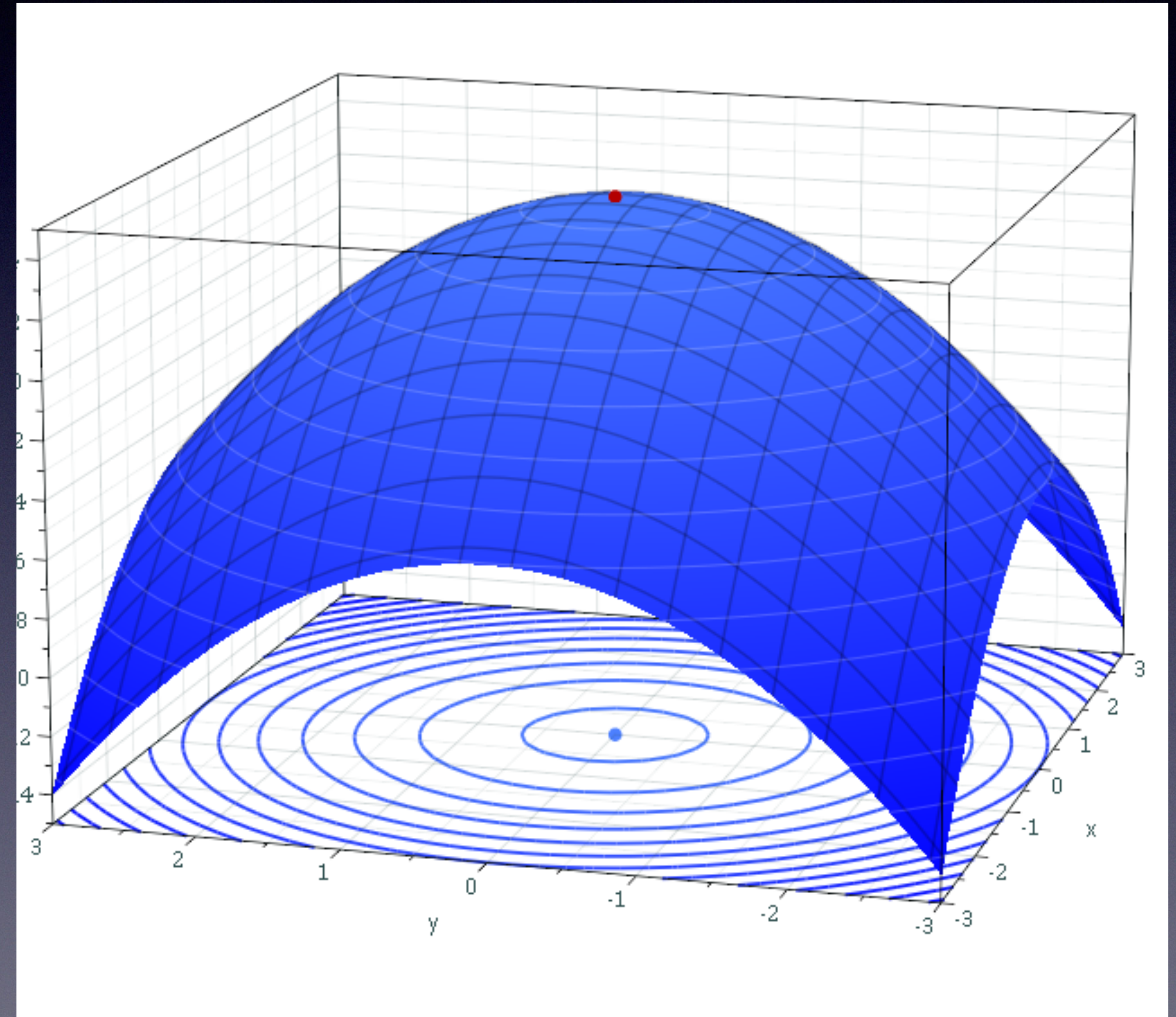
## Distancia del coseno y similitud del coseno

Dados dos vectores A y B , la similitud del coseno,  $\cos(\theta)$ , se representa utilizando un producto escalar y una magnitud como se muestra



# Concepto de optimización

- La optimización matemática consiste en el uso de ecuaciones matemáticas y algoritmos para resolver problemas con el propósito de encontrar la mejor solución posible entre todas las soluciones viables.
- Significa encontrar una solución que maximice o minimice algo en particular, de acuerdo con sus propósitos.





# Componentes de un problema de optimización

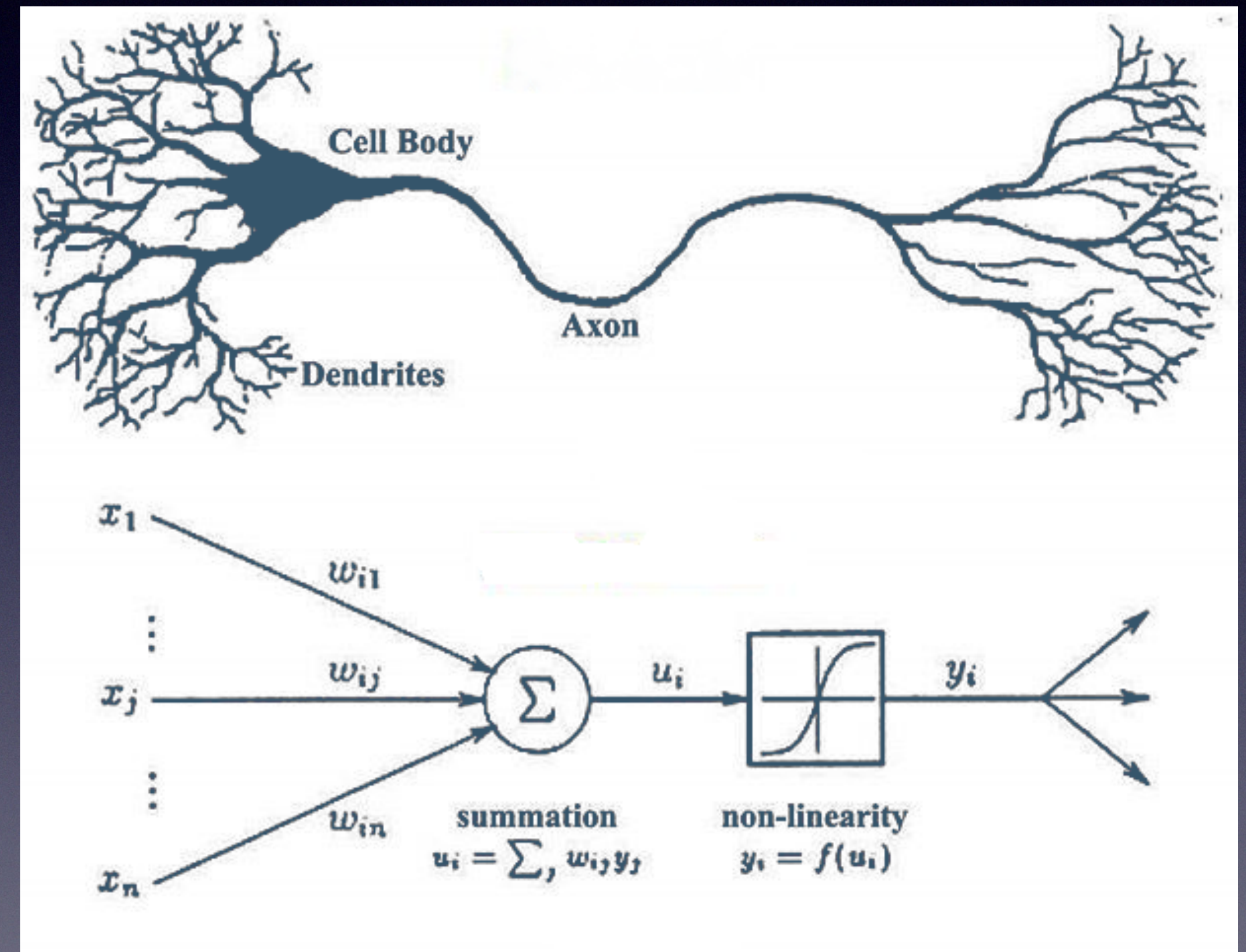
1. Variables de decisión: variables independientes que tienen un impacto significativo sobre la función objetivo
2. Restricciones: ecuaciones o inecuaciones las relaciones existentes entre las variables de decisión. Pueden ser de igualdad o de desigualdad
3. Función objetivo: índice de rendimiento o criterio de elección. La función objetivo permite determinar los mejores valores para las variables de decisión

$$\left\{ \begin{array}{l} \text{minimizar } f(x) \\ \text{sujeto a} \\ g_i(x) \leq 0 \quad i = 1, \dots, m \\ x \in S \subset \mathfrak{R}^n. \end{array} \right.$$



# Estructuras de redes neuronales

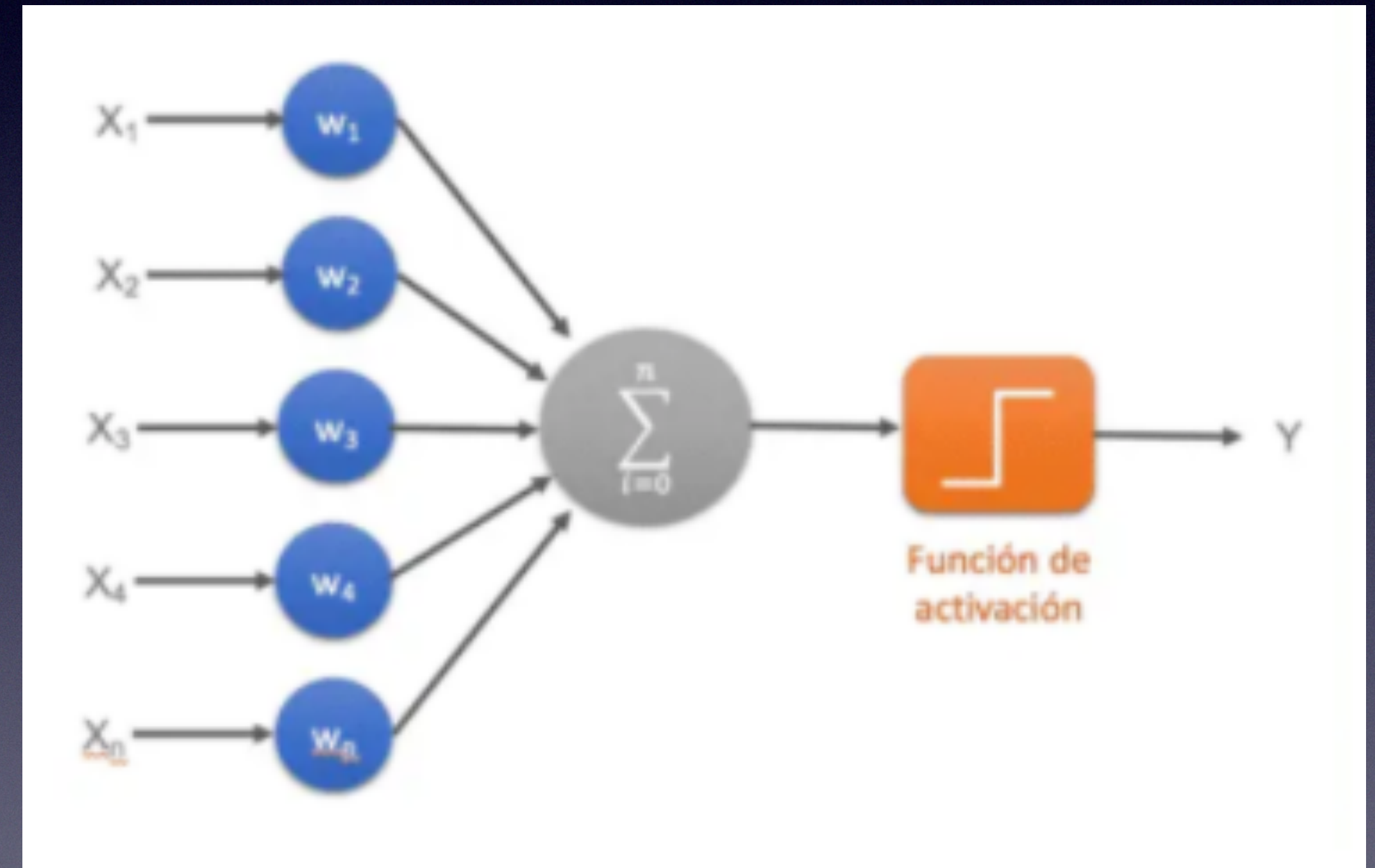
- Una red neuronal es un modelo simplificado que emula el modo en que el cerebro humano procesa la información
- Funciona mediante la creación de conexiones entre muchos elementos de procesamiento diferentes, cada uno similar a una sola neurona





# Perceptrón simple

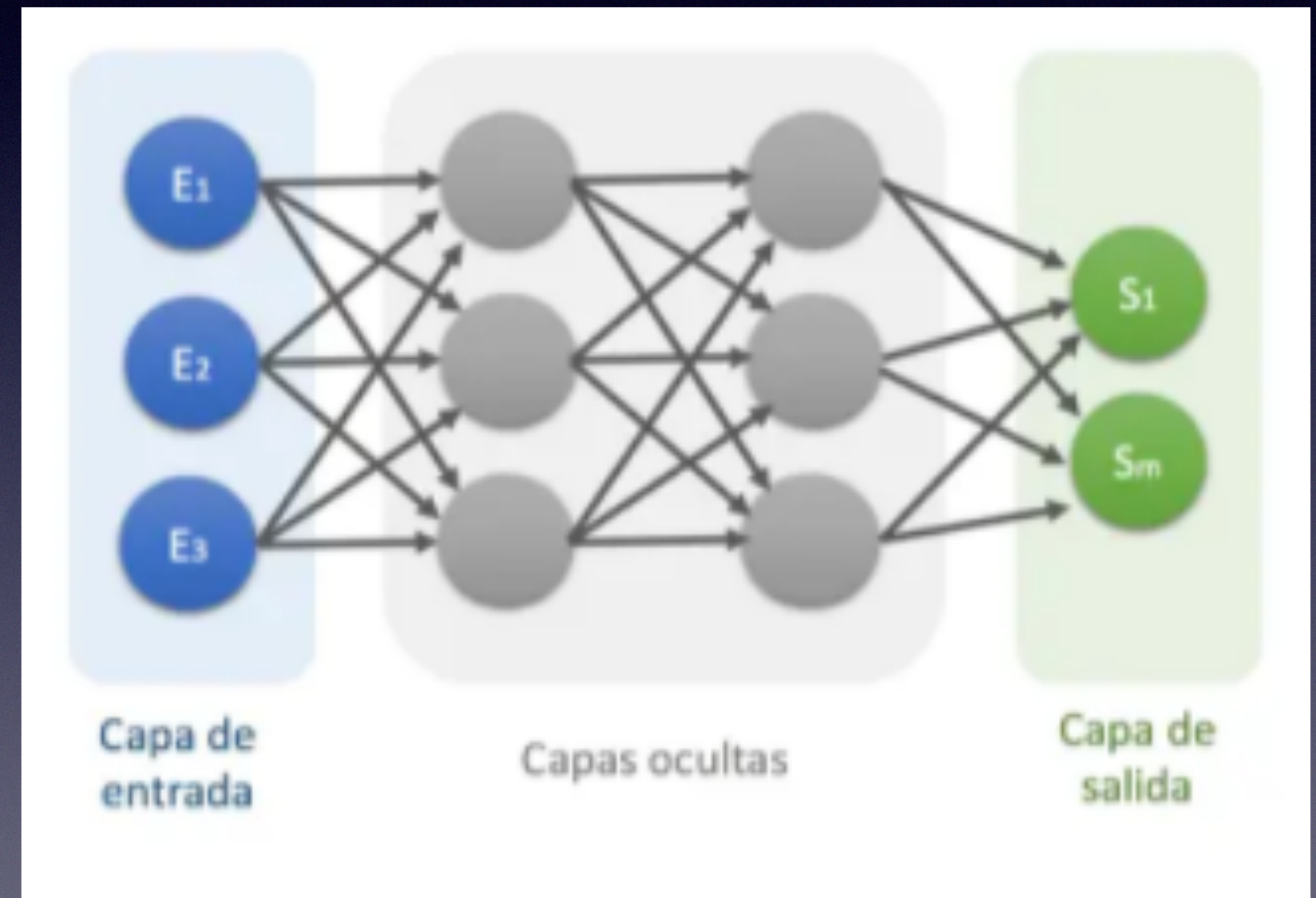
- La red neuronal monocapa se corresponde con la red neuronal más simple, está compuesta por una capa de neuronas que proyectan las entradas a una capa de neuronas de salida donde se realizan los diferentes cálculos





# Perceptrón multicapa

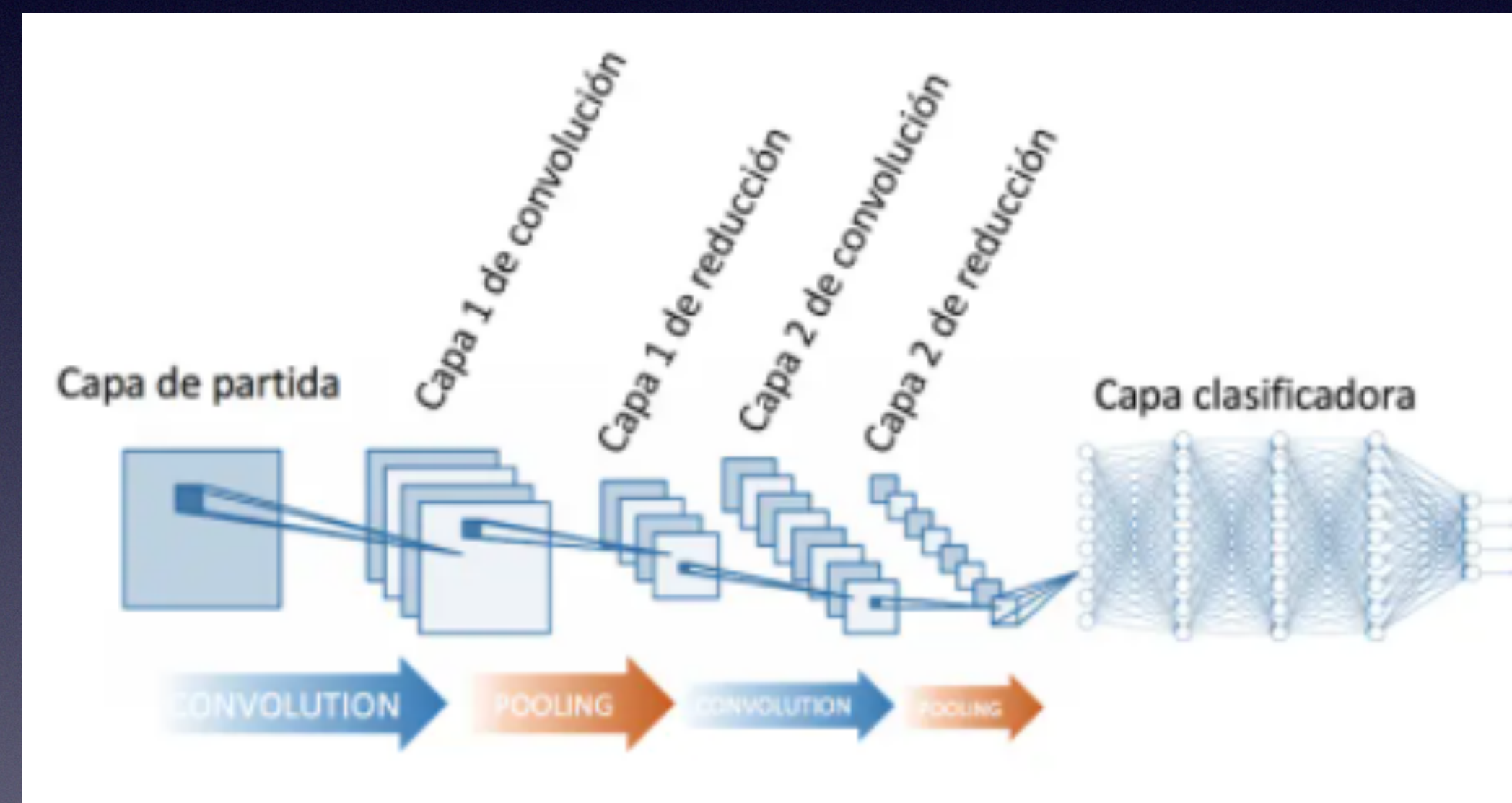
- Es una generalización de la red neuronal monocapa,
- Dispone de un conjunto de capas intermedias (capas ocultas) entre la capa de entrada y la de salida.





# Red neuronal Convolutiva (CNN)

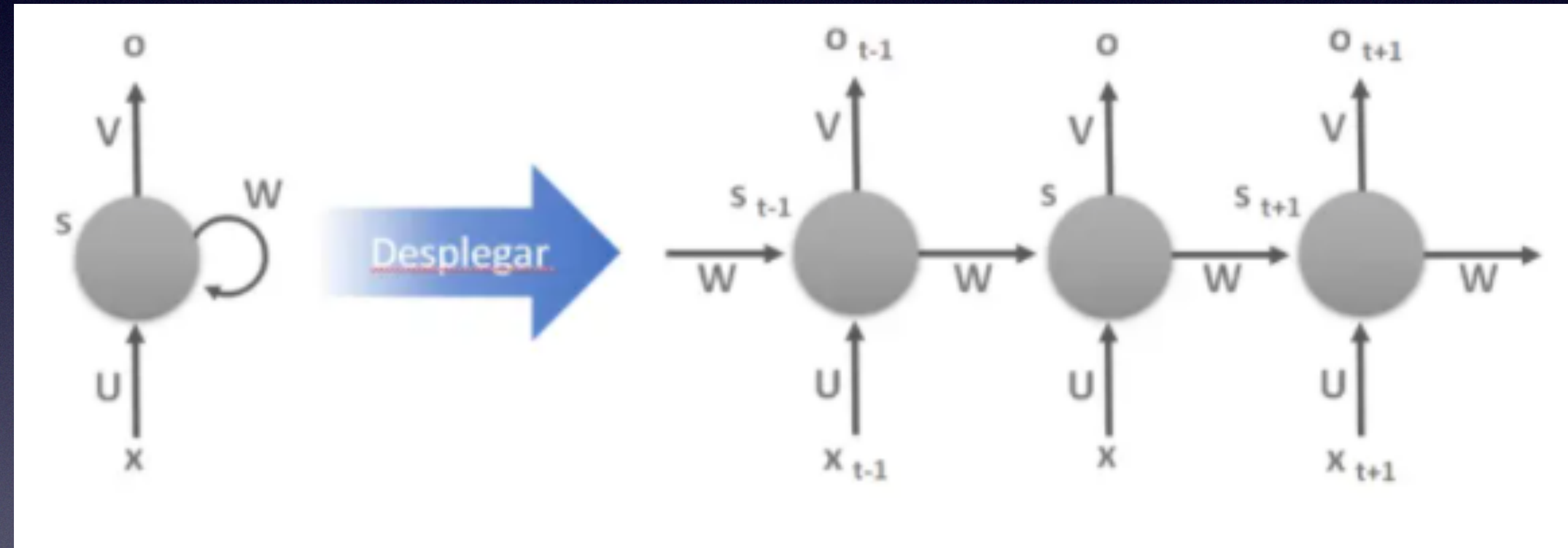
- La principal diferencia de la red neuronal convolutiva con el perceptrón multicapa viene en que cada neurona no se une con todas y cada una de las capas siguientes sino que solo con un subgrupo de ellas (se especializa), con esto se consigue reducir el número de neuronas necesarias y la complejidad computacional necesaria para su ejecución.





# Red neuronal recurrente (RNN)

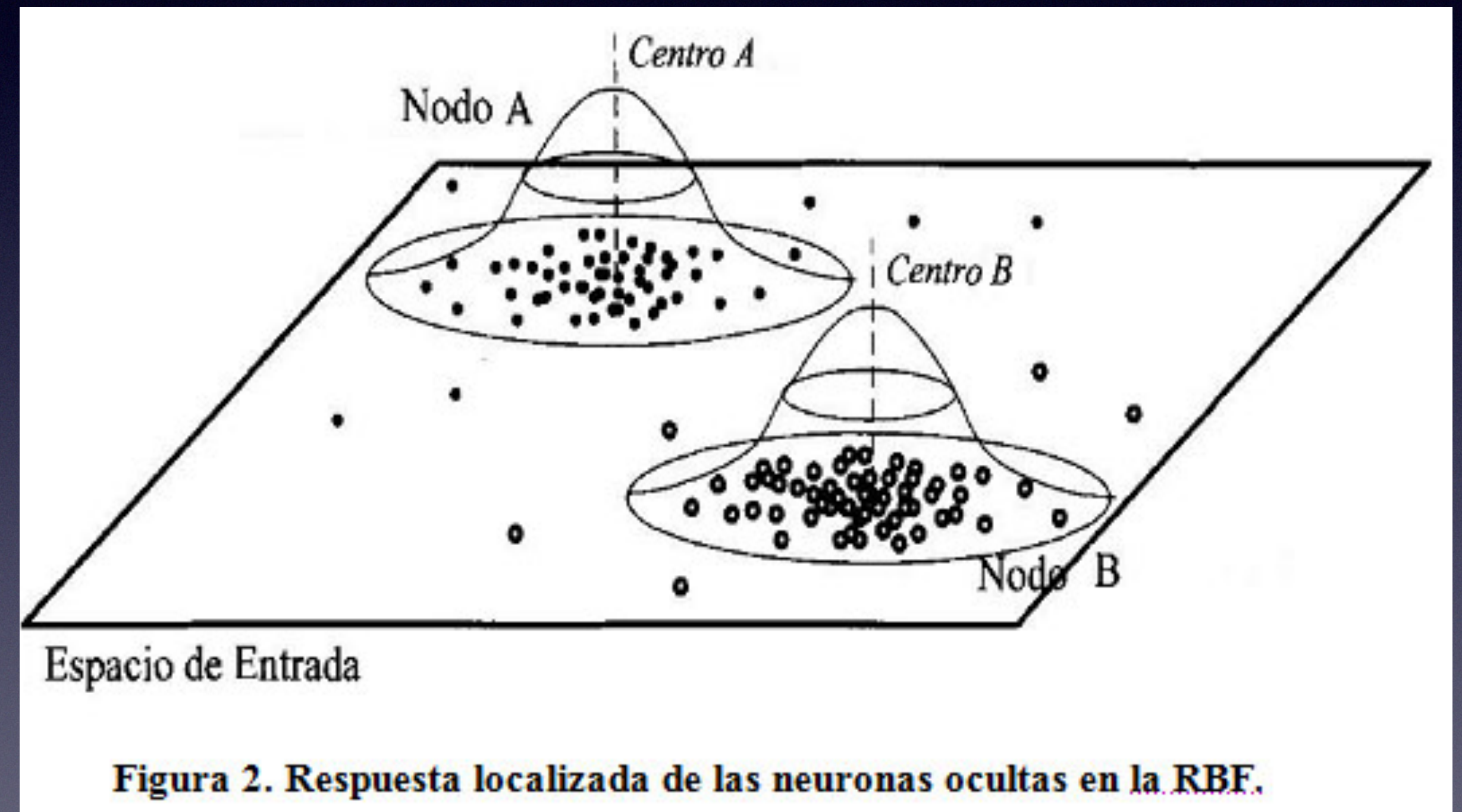
- Las redes neuronales recurrentes no tienen una estructura de capas, sino que permiten conexiones arbitrarias entre las neuronas, incluso pudiendo crear ciclos, con esto se consigue crear la temporalidad, permitiendo que la red tenga memoria.
- Los datos introducidos en el momento  $t$  en la entrada, son transformados y van circulando por la red incluso en los instantes de tiempo siguientes  $t + 1$ ,  $t + 2$ , ...



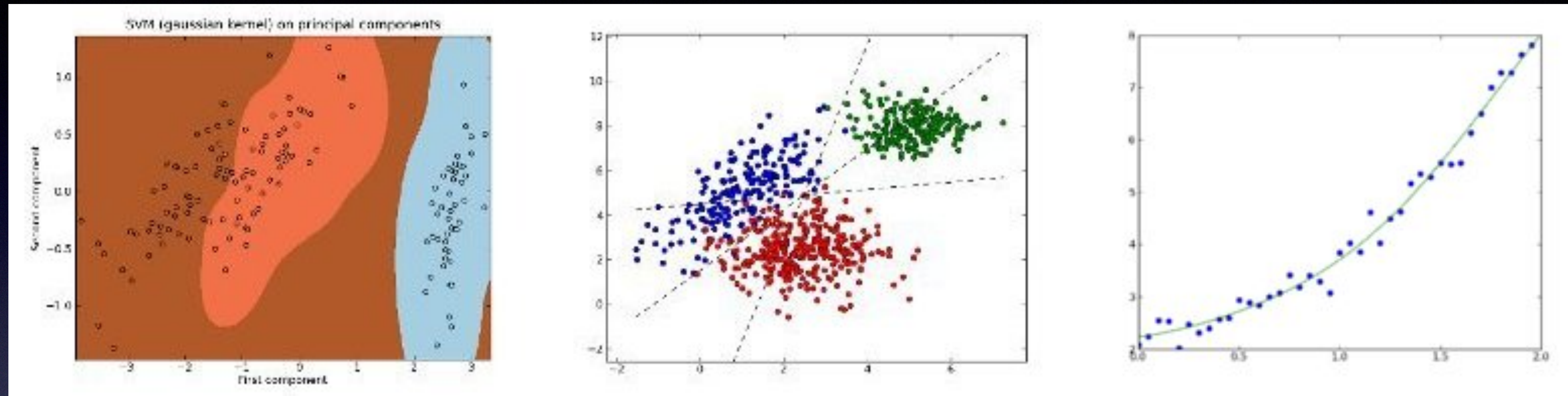


# Redes de base radial (RBF)

- Las redes de base radial calculan la salida de la función en función de la distancia a un punto denominado centro. La salida es una combinación lineal de las funciones de activación radiales utilizadas por las neuronas individuales.
- Las redes de base radial tienen la ventaja de que no presentan mínimos locales donde la retropropagación pueda quedarse bloqueada.
- Las redes RBF tienen una construcción rígida de tres capas: Capa de entrada, capa oculta y capa de salida (a diferencia de otras redes backpropagation).







# Aprendizaje inductivo

Machine learning



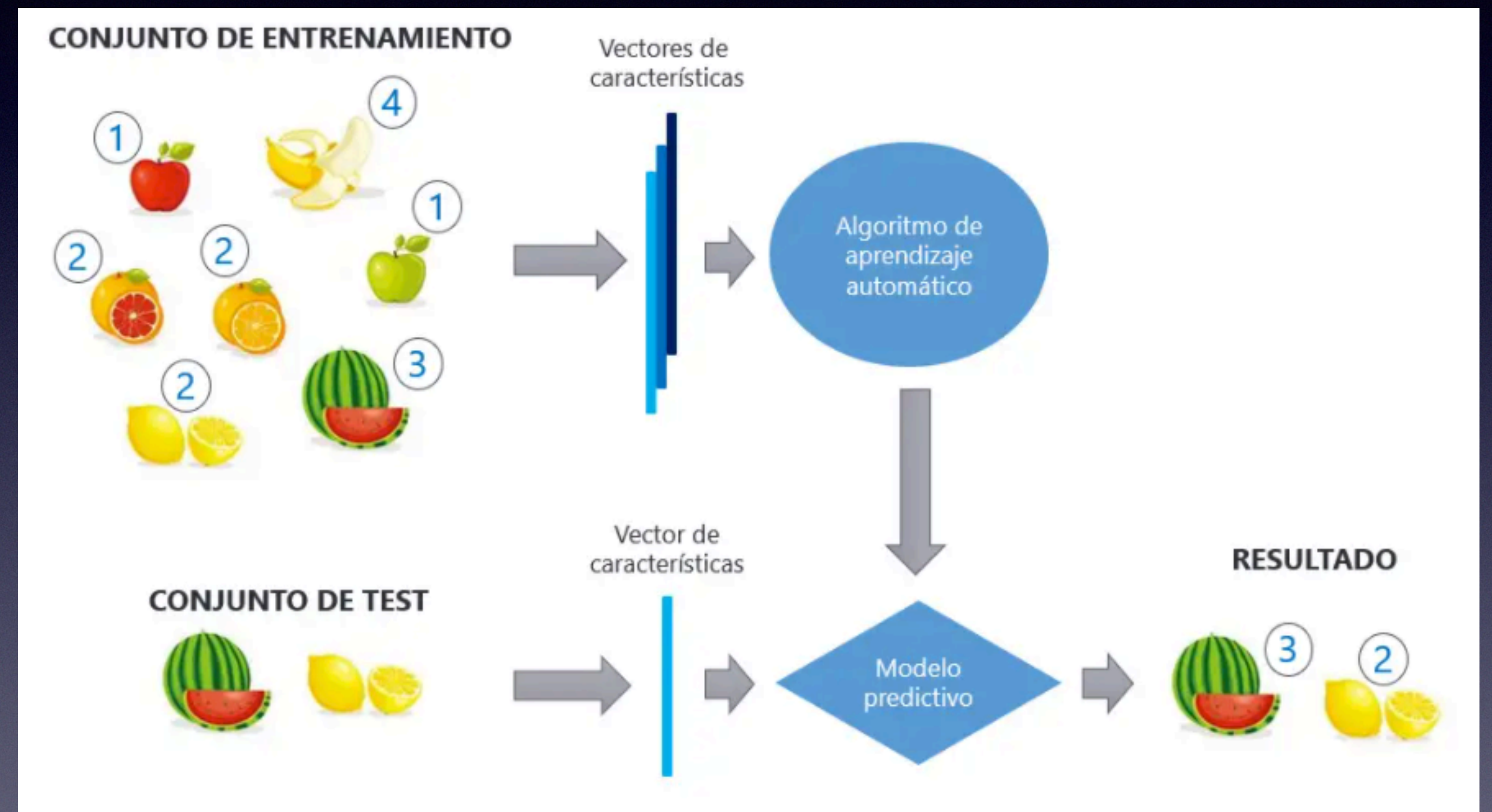
# Aprendizaje inductivo

- Regresión: Intentan predecir un valor real. Por ejemplo, predecir la nota de un alumno en el examen final basándose en las notas obtenidas en las diversas tareas realizadas durante el curso.
- Clasificación (binaria o multiclase): Intentan predecir la clasificación de objetos sobre un conjunto de clases prefijadas. Por ejemplo, clasificar si una determinada noticia es de deportes, entretenimiento, política, etc. Si solo se permiten 2 posibles clases, entonces se llama clasificación binaria; si se permiten más de 2 clases, estamos hablando de clasificación multiclase.
- Ranking: Intentar predecir el orden óptimo de un conjunto de objetos según un orden de relevancia predefinido. Por ejemplo, el orden en que un buscador devuelve recursos de internet como respuesta a una búsqueda de un usuario.



# Aprendizaje supervisado

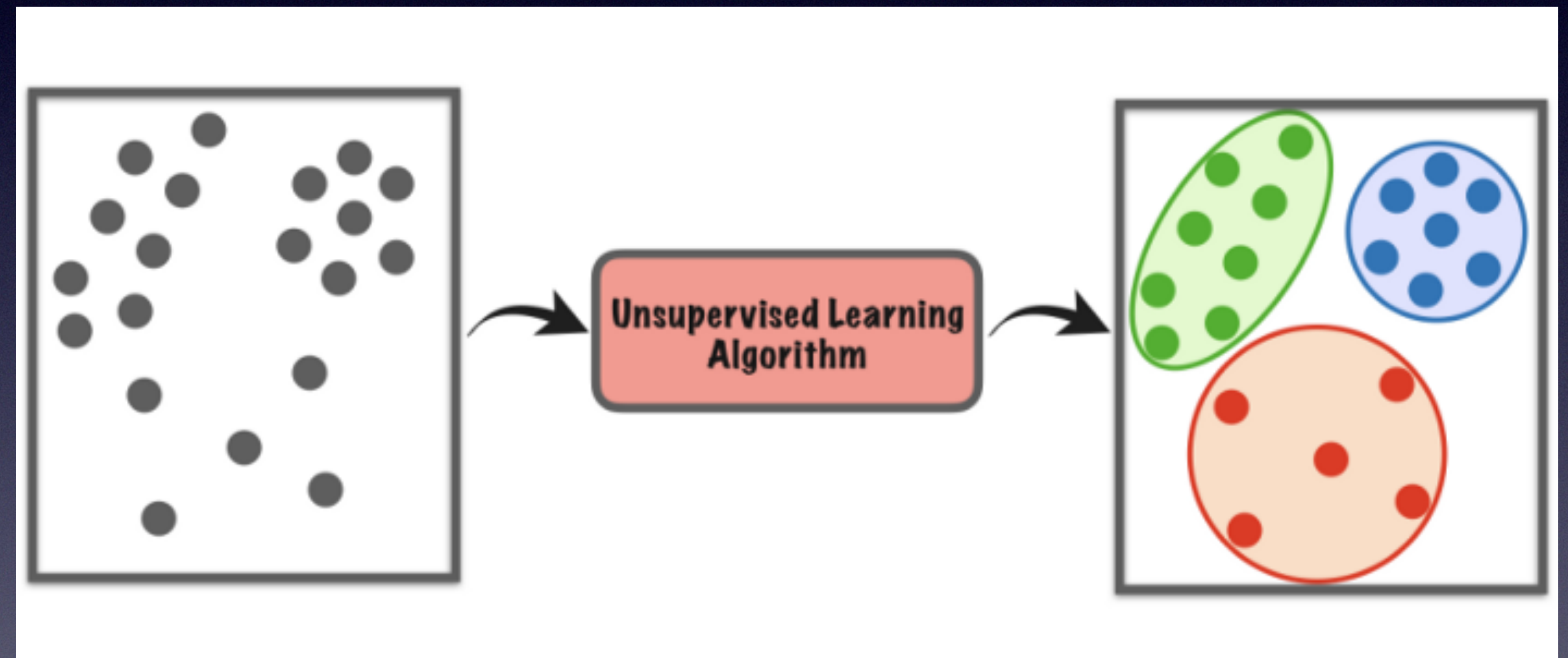
- Son aquellos en los que se aprenden funciones, relaciones que asocian entradas con salidas
- Se ajustan a un conjunto de ejemplos de los que conocemos la relación entre la entrada y la salida deseada.





# Aprendizaje NO supervisado (Clustering)

- Son un conjunto de técnicas que permiten inferir modelos para extraer conocimiento de conjuntos de datos donde a priori se desconoce.
- Se pueden aplicar sin necesidad de tener los datos etiquetados para el entrenamiento.
- Sólo se puede describir la estructura de los datos y con ello intentar encontrar algún tipo de organización que simplifique el análisis
- Tiene carácter exploratorio.





# Práctica con Python