REDES NEURONALES 2024

Clase 15 parte 2

Lunes 7 de octubre 2024

FAMAF, Universidad Nacional de Córdoba

Instituto de Física Enrique Gaviola (UNC-CONICET)

J12/10/23 c15p1

COMO DESHACERNOS DEL MOLESTO UMBRAL

El hecho de que los pesos y el umbral tengan un rol tan diferente en la dinámica de la neurona complica la generación de algoritmos. Pero es una diferencia ficticia. Hay una manera muy simple de equiparar al umbral a una eficacia sináptica. Para ellos imaginamos que tenemos una entrada extra a la que le asignamos el índice *k=0* y está fija en el valor *-1*.

$$h_i - \mu_i = \sum_{k=1}^{N} w_{ik}\xi_k - \mu_i$$

$$= \sum_{k=1}^{N} w_{ik}\xi_k + \mu_i \xi_0$$

$$= \bar{w}_i \cdot \bar{\xi} + \mu_i \xi_0$$

$$h_i - \mu_i = \sum_{k=0}^{N} w_{ik}\xi_k$$

siempre que $w_{i0} = \mu_i$ y $\xi_0 = -1$

Con esto nos olvidamos de darle un tratamiento diferenciado al umbral

A partir de ahora entonces, para nosotros ***aprender*** significa encontrar un vector en *N+1* dimensiones que resuelva los datos provistos en el conjunto de entrenamiento.

$$\overline{W} = (w_{i0}, w_{i1}, \dots, w_{iN}) \in \mathbb{R}^{N+1}$$

$$= (w_{i1}, w_{i2}, \dots, w_{i,N+1}) \in \mathbb{R}^{N+1}$$

o sea, que para cualquier entrada del conjunto de entrenamiento nos de

$$O_i^\alpha = g(h_i^\alpha) = g\left(\sum_{k=0}^{N} w_{ik}\xi_k^\alpha\right) = \zeta_i^\alpha$$

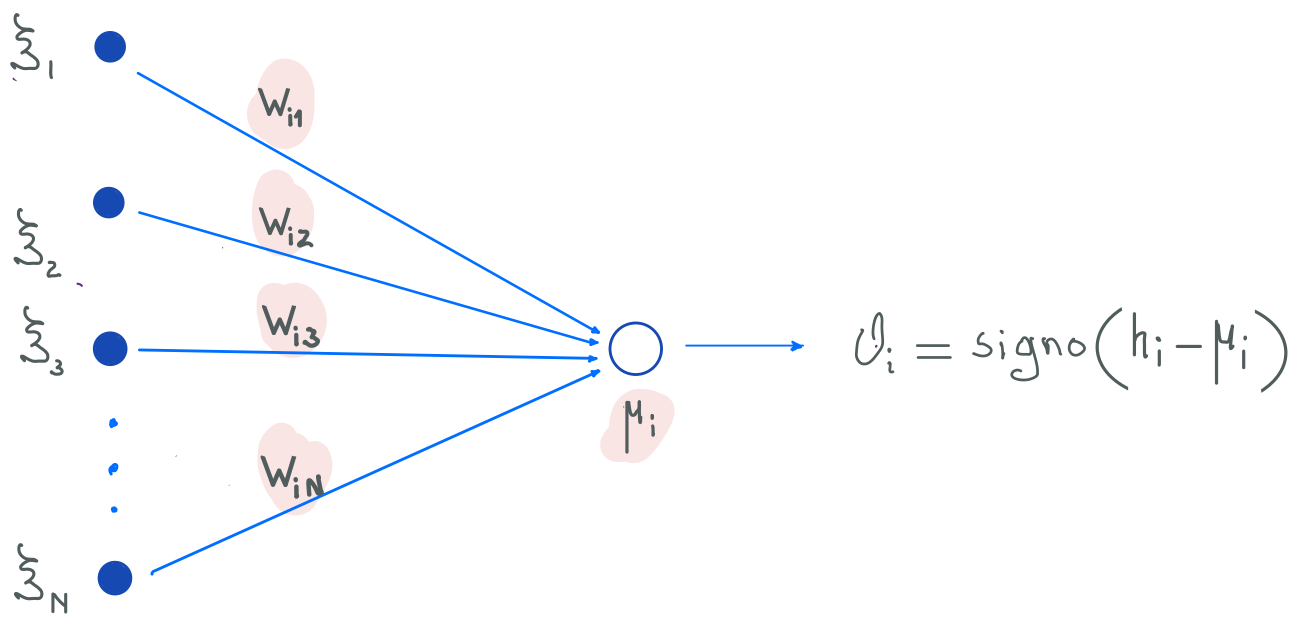
$$\boxed{O_i^\alpha = \zeta_i^\alpha} \leftarrow \text{Esto es lo deseado}$$

No siempre podremos aprender perfectamente el conjunto de entrenamiento. Es más, casi nunca. Nos conformaremos con fijar una tolerancia que aceptaremos en el error que comete la red con el conjunto de entrenamiento y dejaremos de buscar mejores soluciones cuando se cumpla que

$$\text{Error de entrenamiento} \leq \text{Tolerancia al error}$$

Más adelante veremos qué le exigimos a la red a la hora de generalizar, o sea, de resolver lo que nunca vió, pero lo dejamos para las próximas clases.

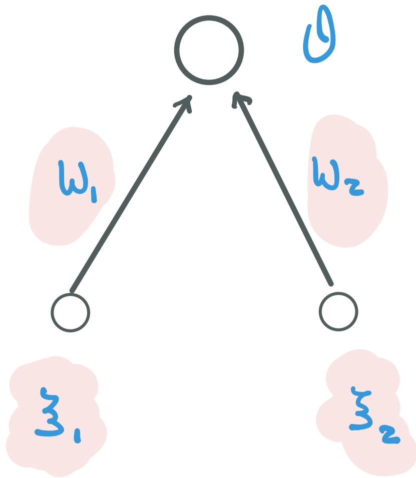
PERCEPTRON SIMPLE COMO UN CLASIFICADOR BINARIO



Vamos a analizar la red neuronal más simple que podemos construir para clasificar en dos categorías los valores de entrada. Supongamos que

$$\vec{\xi} = (\xi_1, \xi_2, \ldots, \xi_N)$$

y la salida es binaria, pudiendo tomar valores *-1* y *+1*. Para simplificarlo aún más, supongamos que solo hay dos entradas ξ_1 y ξ_2 .



PERCEPTRON CON $N=2$

Quitamos aquí el índice i que identifica a la neurona para no complicarnos con los índices en el cálculo que haremos, pero tengamos en mente que estas neuronas simples se pueden ensamblar y que cuando eso ocurre necesitamos preservar el índice.

Suponemos también que las variables de entrada pueden tomar cualquier valor real. Tanto la entrada como las sinapsis vive en el plano. No incluiremos por ahora el umbral.

$$\bar{\xi} \in \mathbb{R}^2 \qquad \bar{\xi} = (\xi_1, \xi_2)$$

$$\mu = 0$$

$$\bar{w} \in \mathbb{R}^2 \qquad \bar{w} = (w_1, w_2)$$

$$O = \text{signo}(h_i)$$

$$O = \text{signo}\left(W_1 \xi_1 + W_2 \xi_2\right)$$

$$= \text{signo}\left(\overline{W} \cdot \overline{\xi}\right) \in -1, +1$$

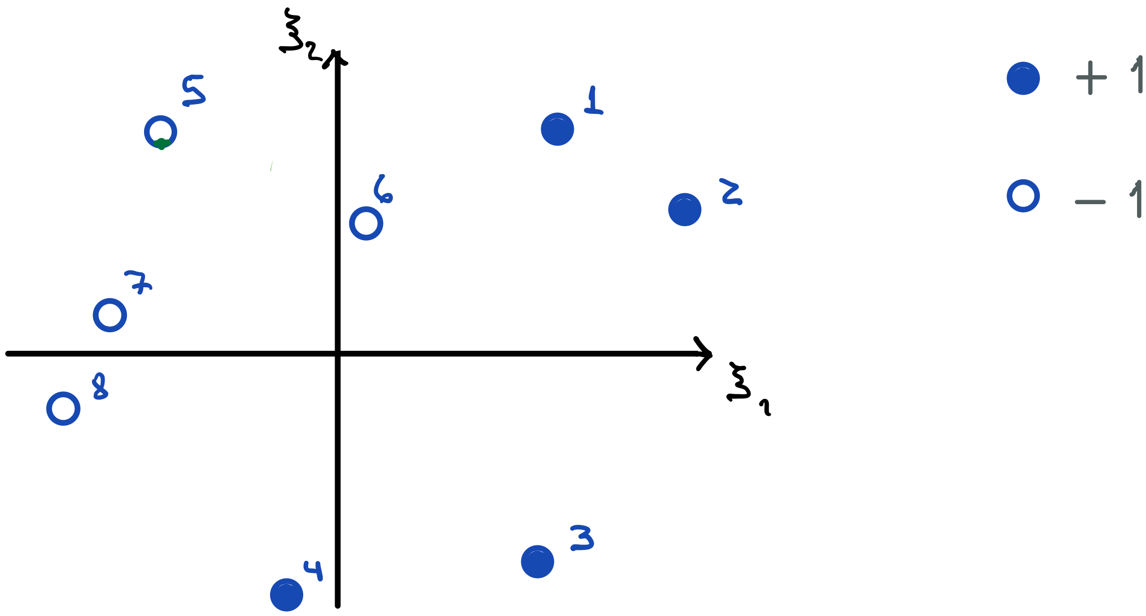
Tenemos relaciones **INPUT-OUTPUT** conocidas, las que serán utilizadas para entrenar. Supongamos que tenemos un conjunto pequeño, de solo **8** elementos en el conjunto de entrenamiento (poquísimas para un caso real)

$$\overline{\xi}^{\alpha} \longrightarrow \zeta^{\alpha} = \pm 1$$ Valor correcto de salida

$$\alpha = 1, 2, 3, \dots, 8$$

$$\zeta^{\alpha} = \begin{cases} +1 & \text{si } \overline{\xi}^{\alpha} \text{ pertenece a la categoría A} \\ -1 & \text{si } \overline{\xi}^{\alpha} \text{ pertenece a la categoría B} \end{cases}$$

Voy a representar los puntos en R^2



Cada uno de los 8 puntos representa a uno de los 8 elementos del conjunto de entrenamiento. En este caso en particular 4 elementos pertenecen a la categoría A *(α = 1, 2, 3 y 4)* y los representamos con círculos azules llenos, en tanto los otros 4 elementos pertenecen a la categoría B *(α = 5, 6, 7 y 8)* y los representamos con círculos azules vacíos.

Hagamos algunas cuentas simples:

$$\xi^\alpha \longrightarrow O^\alpha = \zeta^\alpha \qquad \text{(lo deseado)}$$

$$\textbf{signo}\left(\sum_{k=1}^{2} w_k \xi_k^\alpha\right) = \zeta^\alpha$$

$$\zeta^\alpha\,\textbf{signo}\left(\vec{w}\cdot\vec{\xi}^\alpha\right) = \zeta^\alpha \zeta^\alpha$$

$$\zeta^{\alpha} \ \textbf{signo}\left(\bar{w} . \bar{\xi}^{\alpha}\right) = \left(\zeta^{\alpha}\right)^{2} = +1$$

$$\textbf{signo}\left(\zeta^{\alpha}\right) \textbf{signo}\left(\bar{w} . \bar{\xi}^{\alpha}\right) = +1$$

$$\textbf{signo}\left(\left(\bar{w} . \bar{\xi}^{\alpha}\right) \zeta^{\alpha}\right) = +1$$

$$\textbf{signo}\left(\bar{w} . \left(\bar{\xi}^{\alpha} \zeta^{\alpha}\right)\right) = +1$$

Sea $\bar{X}^{\alpha} = \bar{\xi}^{\alpha} \zeta^{\alpha}$,

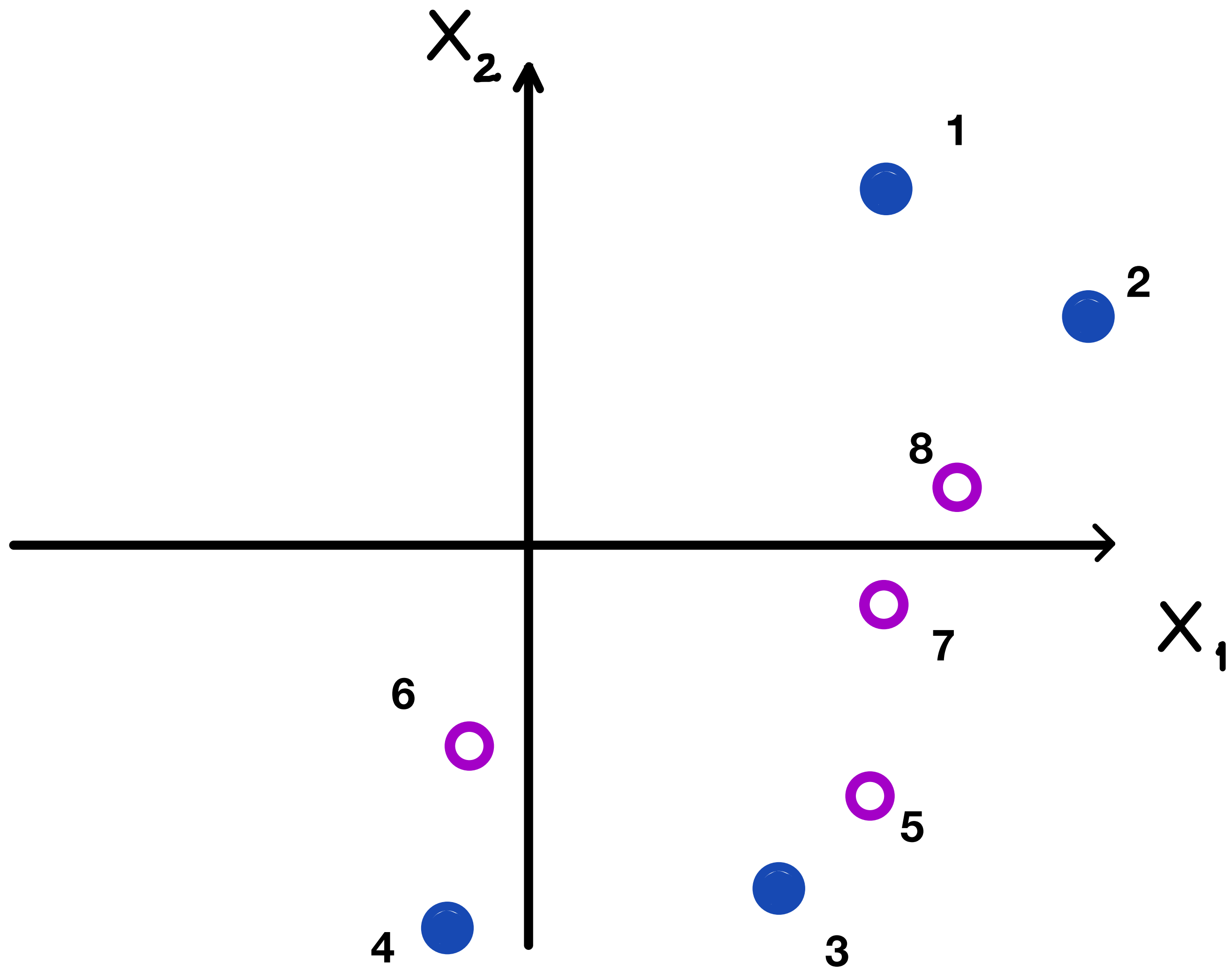
$$\textbf{signo}\left(\bar{w} . \bar{X}^{\alpha}\right) = +1$$

entonces $\boxed{\bar{w} . \bar{X}^{\alpha} > 0}$

Notemos que:

si $\zeta^{\alpha} = +1$ entonces $\bar{X}^{\alpha} = \bar{\xi}^{\alpha}$

si $\zeta^{\alpha} = -1$ entonces $\bar{X}^{\alpha} = -\bar{\xi}^{\alpha}$



X_2
X_1
1
2
8
7
6
5
3
4

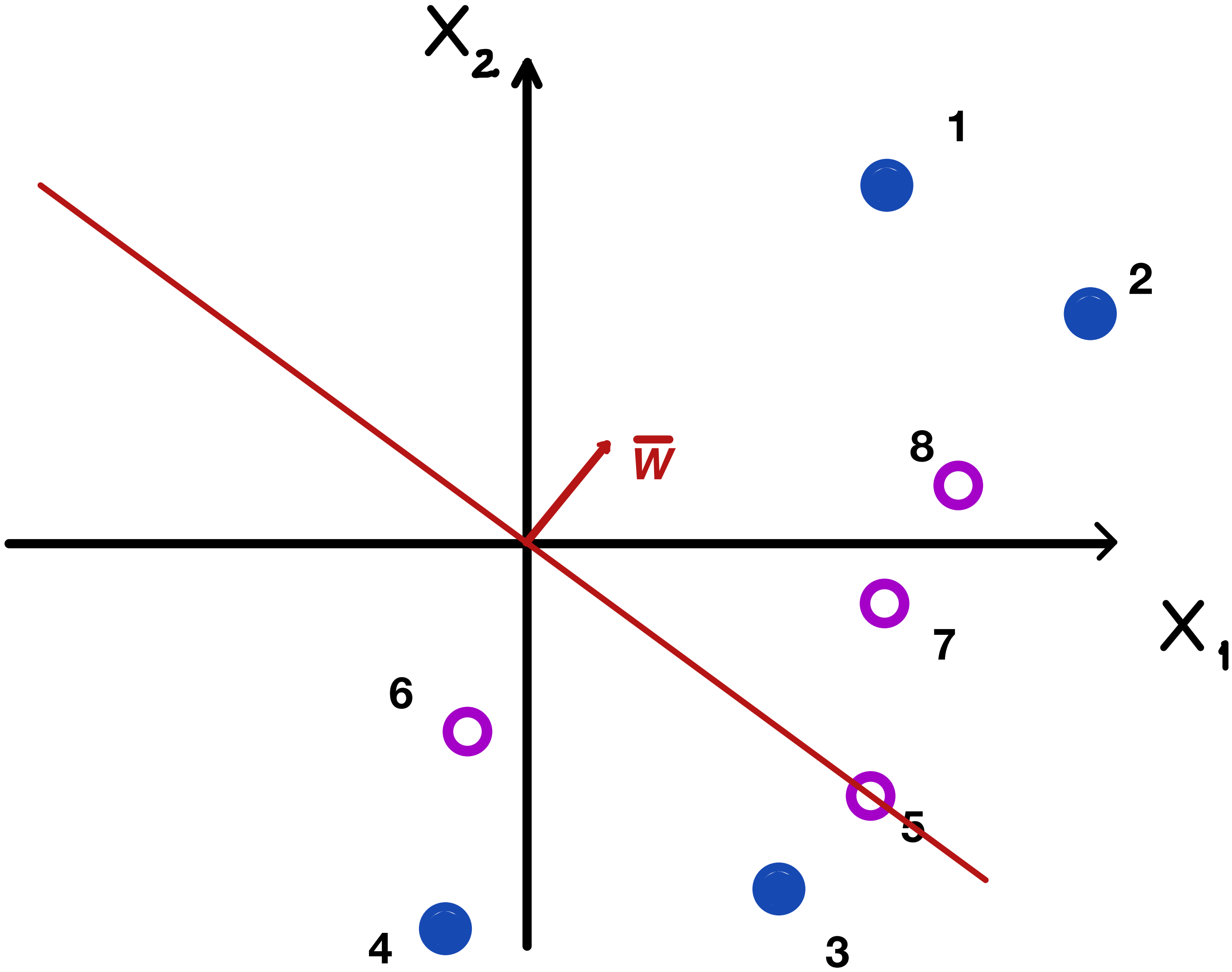
No perdamos de vista que necesitamos que, para los ocho elementos del conjunto de entrenamiento, se cumpla que:

$$\overline{W} \cdot \overline{X}^{\alpha} > 0$$

Sabemos que $\overline{W}^{\alpha} \cdot \overline{X} > 0$ Implica que la proyección de $\overline{W}$ sobre cada uno de los elementos del conjunto de entrenamiento debe ser positiva. O sea, de TODOS LOS EJEMPLOS.

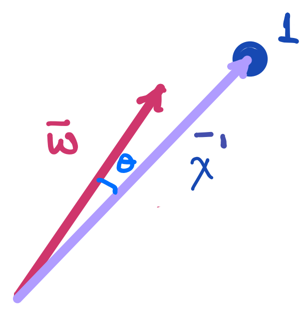
Como vimos, el $\overline{W}$ que buscamos vive en el plano.

Comencemos poniendo un vector $\overline{W}$ en el plano y veamos si procesa bien el conjunto de entrenamiento.



¿Resuelve este vector ***W*** los elementos del conjunto de entrenamiento?

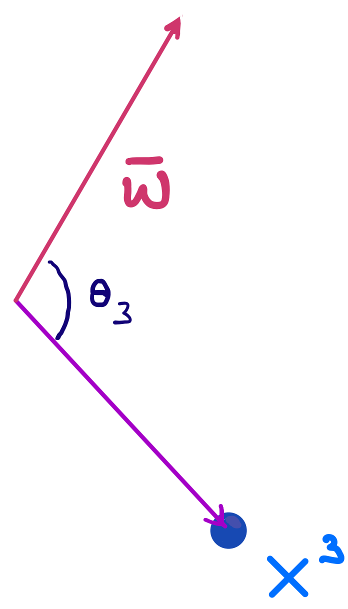
Verifiquemos que para el elemento **1** funciona bien:



$$\overline{W} \cdot \overline{X}^1 = |\overline{W}|\,|\overline{X}^1| \cos(\theta_1) > 0$$

pues $0 < \theta_1 < \frac{\pi}{2}$ y su coseno es entonces positivo.

Miremos ahora qué pasa con el elemento **3**:



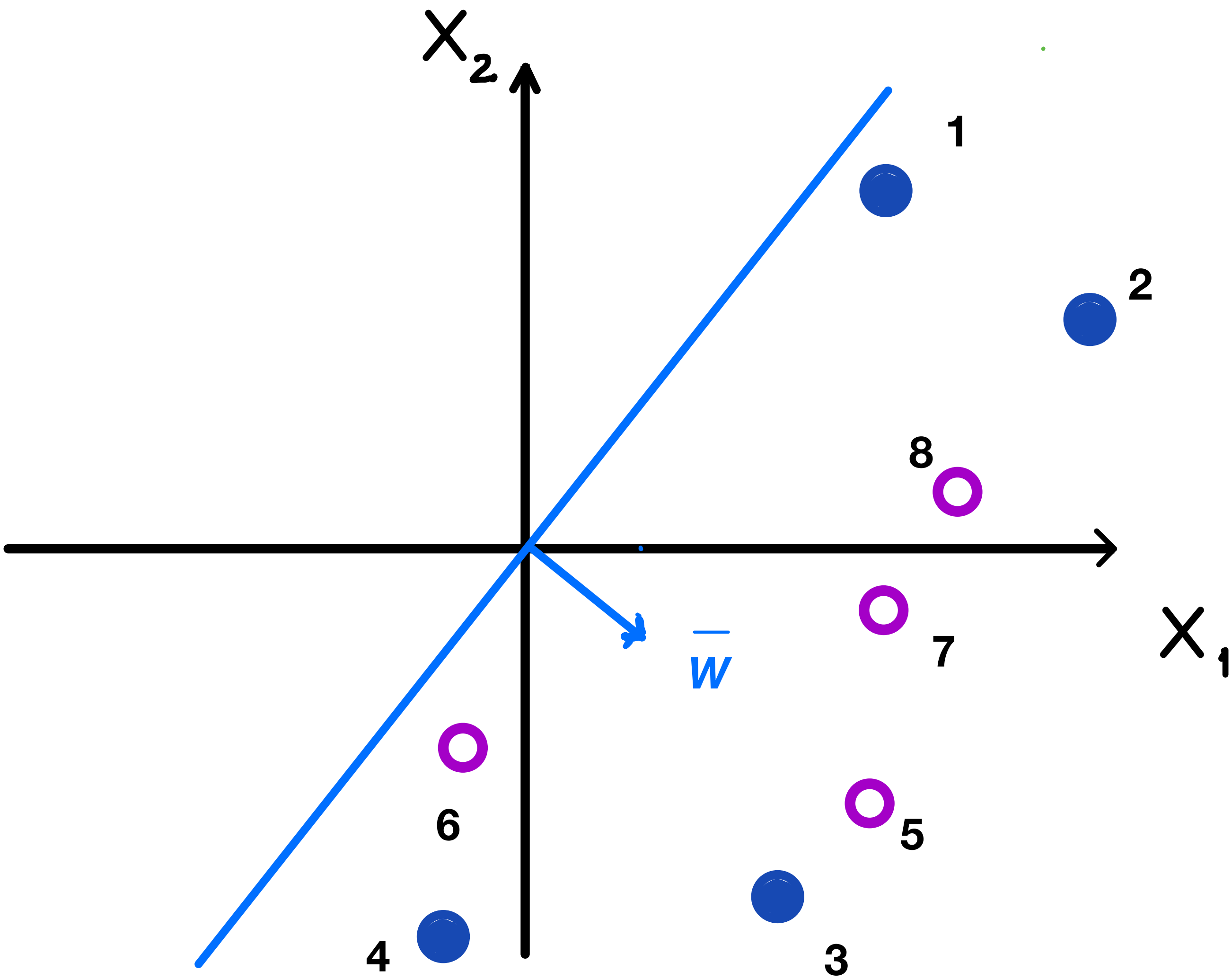
$$\overline{W}\cdot\overline{X}^{3} = |\overline{W}|\,|\overline{X}^{3}|\cos(\theta_3) < 0$$

pues $\frac{\pi}{2} < \theta_1 < \pi$ y su coseno es entonces negativo.

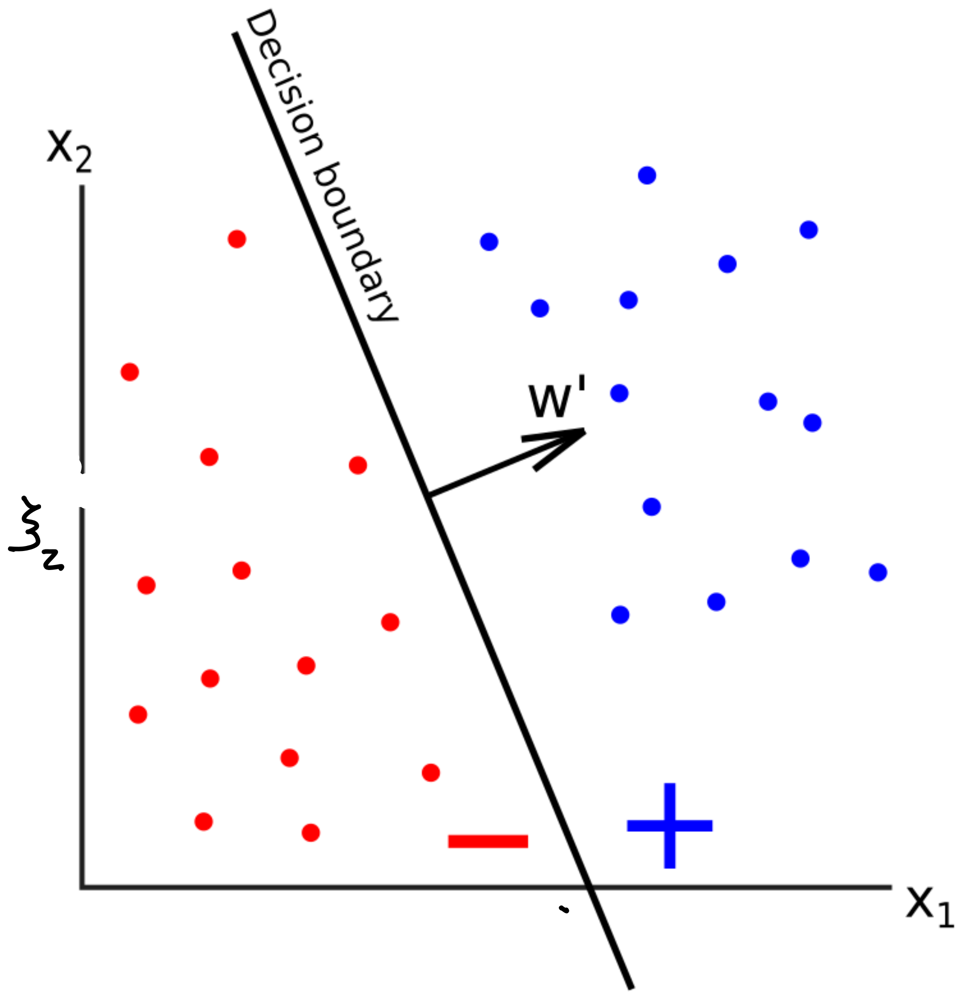
Esto significa el $\overline{W}$ que introdujimos no nos sirve. Procesa bien los ejemplos 1, 2, 7 y 8 pero procesa mal los ejemplos 3, 4 y 6. Además deja al elemento 5 del conjunto de entrenamiento en el borde.

O sea, debemos encontrar otro $\overline{W}$ que forme un ángulo agudo con los ocho elementos del conjunto de entrenamiento en su representación $\overline{X}$.

Introducimos un nuevo vector posible $\overline{W}$ en el plano.



Verificamos ahora que este nuevo vector de sinpsis ***W*** divide el plano en dos y deja a todos los valores de ***X*** del conjunto de entrenamiento en un semiplano al cual él pertenece. O sea, el tiene proyección positiva con los ocho elementos del conjunto de entrenamiento en la representación ***X***.



Decision boundary
x_2
x_1
w'
ξ_2
−
+

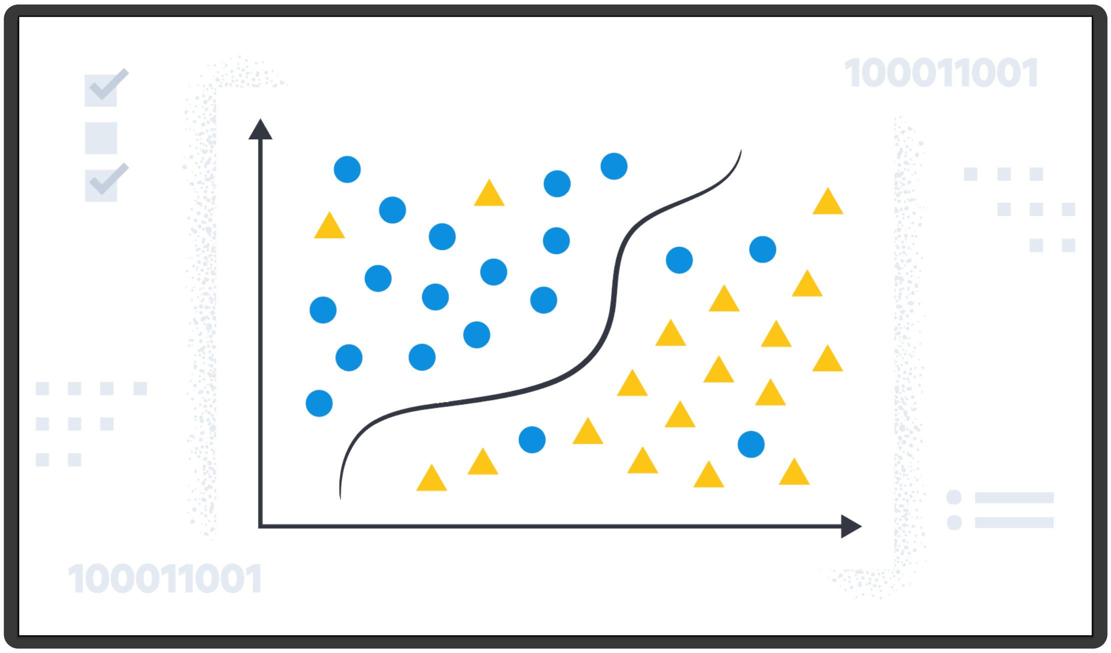
Que exista uno o infinitos $\overline{W}$ depende de que exista una o infinitas rectas que pasan por el origen que separe $\mathbb{R}^2$ en dos semiplanos, uno que contenga todos los ejemplos y otro que no tenga ningún ejemplo.

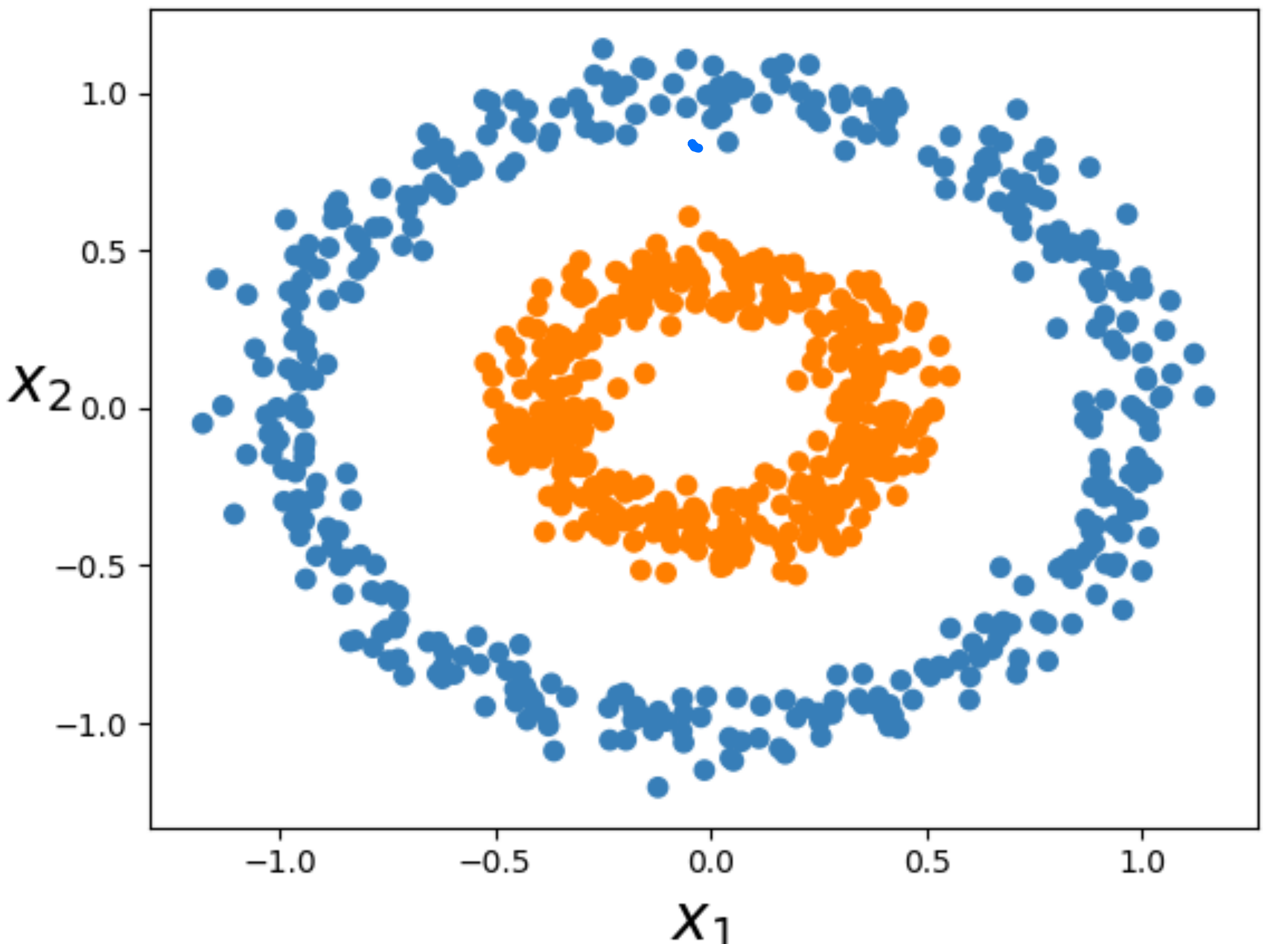
Si existe una o infinitas rectas que hacen posible esta deseada separación, diremos que el problema es

LINEALMENTE SEPARABLE

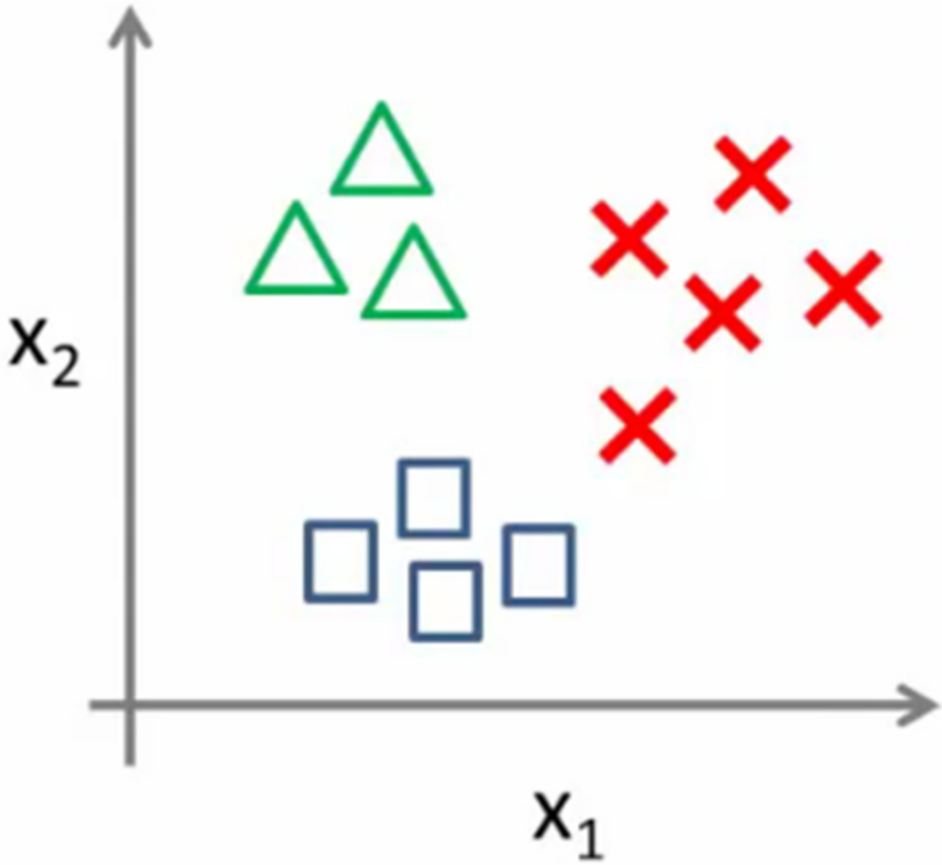
En el caso más general, separabilidad lineal significa que en $\mathbb{R}^N$ debe existir al menos un hiperplano en $\mathbb{R}^{N-1}$ que deje de un lado todos los vectores de entrenamiento y del otro no deje ninguno.

Aprenderemos más adelante que en la gran mayoría de los casos no lograremos un aprendizaje perfecto, pues siempre quedarán ejemplos del lado equivocado. Trataremos entonces de elegir el vector $\overline{W}$ que menos ejemplos deje del lado equivocado.

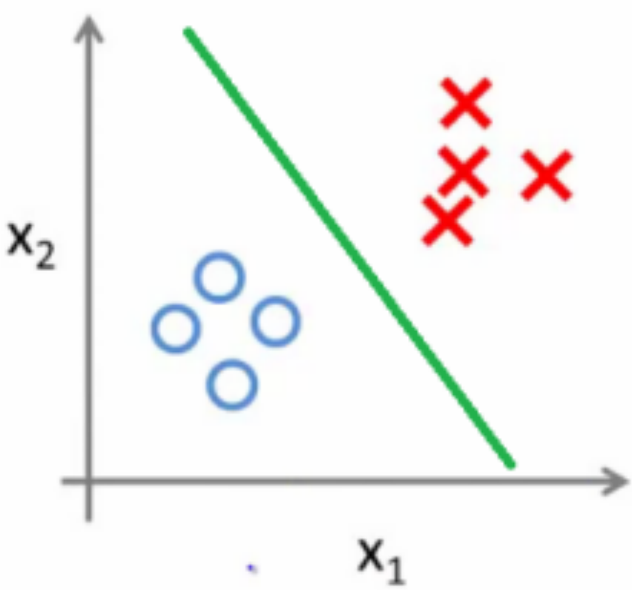

100011001
100011001


1.0
0.5
x_2 0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0
x_1

Multi-class classification:



Binary classification:



Multi-class classification: