# EL PERCEPTRON SIMPLE COMO CLASIFICADOR

## Nuestra primera regla de aprendizaje

Hasta aquí vimos bajo qué condiciones existe una solución al problema de clasificación binaria, pero no tenemos ninguna regla para construir un vector sináptico viable para resolver el problema. Dijimos que es necesario que el problema sea LINEALMENTE SEPARABLE.

Si en nuestro problema tenemos un número grande de entradas $(N \gg 1)$ y un número grande de entradas etiquetadas en el conjunto de entrenamiento $(p \gg 1)$, el problema de encontrar un $\bar{w}$ posible se torna muy difícil.

Ahora introduciremos el primer y más simple algoritmo de aprendizaje.

# REGLA DE APRENDIZAJE DEL PERCEPTRON

Supongamos que tenemos un PERCEPTRON simple con $N$ entradas $\xi_i$ $(i=1,\dots,N)$ y una única neurona de salida $O=\pm 1$. Suponemos que el umbral $\eta$ (tambien llamado bias) es tratado como una neurona extra.

Supongamos que tenemos $p$ entradas etiquetadas (a las cuales alguien le asigno la salida correcta)

$$\vec{\xi}^{\mu} \longrightarrow \zeta^{\mu} \qquad \mu = 1, 2, \dots, p$$

Finalmente, suponemos que el problema es linealmente separable. Nos preguntamos:

¿Cómo encontramos un $\bar{w}$ adecuado?

Ante la falta de otro criterio, elegimos inicialmente los acoplamientos $\bar{w}$ de forma aleatoria. Por ejemplo, podemos elegir los $w_i$ en forma independiente entre sí, y con distribución gaussiana de media $0$ y desviación $\sigma$.

$$P(\bar{w}) = \prod_{i=1}^{N} p(w_i) \quad \leftarrow \text{independencia}$$

$$p(w_i) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{w_i^2}{2\sigma^2}}$$

A partir de este valor inicial de los componentes del vector $\vec{w}$ le presentamos secuencialmente cada uno de los ejemplos etiquetados, uno tras otro.

$\mu = 1$

repetir, hasta que $\mu = ?$

presento $\xi^\mu$

evaluo la salida $O^\mu$

si $O^\mu = \zeta^\mu$ entonces

no hago nada

y si $O^\mu = -\zeta^\mu$ entonces

cambio un poco $\vec{w}$

$\mu = \mu + 1$

fin de repetir

Supongamos $\mu = 1$. Calculamos $O^\mu$

$$O^\mu = g(h^\mu) = g\left(\sum_{k=1}^{N} w_k \xi_k^\mu\right)$$

Si $O^\mu = \zeta^\mu$ no cambiamos nada, pero si

$O^\mu = -\zeta^\mu$ castigamos el perceptron cambiando levemente $\vec{w}$

La regla que usamos es la siguiente

$$\vec{w}^{\,nuevo} = \vec{w}^{\,anterior} + \Delta\vec{w}$$

o, en términos de componentes:

$$w_k^{nuevo} = w_k^{anterior} + \Delta w_k \qquad k=1,2,\dots,N$$

donde:

$$\Delta w_k^\mu = \begin{cases} 2\eta\,\zeta^\mu \xi_k^\mu & \text{si } O^\mu = -\zeta^\mu \\ 0 & \text{si } O^\mu = \zeta^\mu \end{cases}$$

Esto se puede escribir, para cada componente como:

$$\Delta w_k^\mu = \eta\,(1 - \zeta^\mu O^\mu)\,\zeta^\mu \xi_k^\mu$$
$$= \eta\,(\zeta^\mu - O^\mu)\,\xi_k^\mu$$

o, para el vector:

$$\Delta\vec{w} = \eta\,(\zeta^\mu - O^\mu)\,\vec{\xi}^\mu$$

La regla establece que si la red clasifica mal el ejemplo $\mu$ del conjunto de entrenamiento el vector debe "desplazarse" hacia $\zeta^\mu \bar{\xi}^\mu$

- Si $\zeta^\mu = 1$ y se equivocó, $\bar{w}$ se corre un poco hacia $\bar{\xi}^\mu$.

$$\Delta \bar{w} = 2\eta \bar{\xi}^\mu$$

- Si $\zeta^\mu = -1$ y se equivocó, $\bar{w}$ se corre un poco hacia $-\bar{\xi}^\mu$.

$$\Delta \bar{w} = 2\eta \bar{\xi}^\mu$$

- Si no se equivocó, $\bar{w}$ no cambia

$$\Delta \bar{w} = \bar{0}$$

El parámetro $\eta$ se llama razón de aprendizaje y hoy lo presentamos para que siempre nos acompañe.

$\eta$ regula cuán bruscos serán los cambios durante el proceso de aprendizaje. Veremos que es un parámetro MUY DELICADO y que nos dará mucho trabajo

Para que haga bien su trabajo con el ejemplo $\mu$ debe cumplirse que:

$$O^\mu = \text{signo}(h^\mu) = \zeta^\mu$$

o:

$$\zeta^\mu \text{signo}(h^\mu) = \zeta^\mu \zeta^\mu = 1$$

$$\text{signo}(\zeta^\mu h^\mu) = 1$$

$$\zeta^\mu h^\mu > 0 \quad \left\{ \begin{array}{l} \text{con esto} \\ \text{alcanzaría} \end{array} \right.$$

Vamos a pedir más que eso:

$$\zeta^\mu h^\mu > N\kappa$$

$$\boxed{\Delta w_k = \eta\, \Theta\left(N x - \zeta^\mu h^\mu\right) \zeta^\mu \xi_k^\mu}$$

donde

$$\Theta(z) = \begin{cases} 1 & \text{si } z \geq 0 \\ 0 & \text{si } z < 0 \end{cases}$$

Esta regla se llama Regla de Aprendizaje del Perceptron y se debe a Frank Rosenblatt (1962).

Se puede demostrar que si el problema es linealmente SEPARABLE, la regla converge a una solución $\bar{W}$ en un número finito de pasos.

Veamos el mismo problema en la representación

$$\bar{X}^\mu = \zeta^\mu \xi^\mu$$

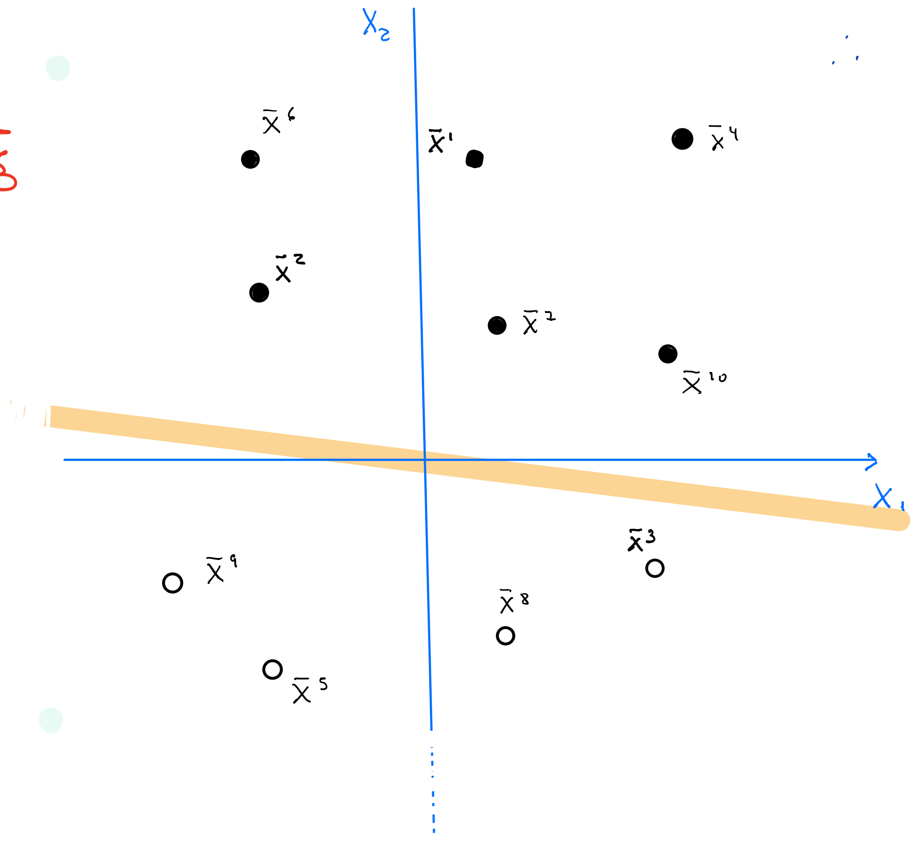
Ahora la regla del perceptron es:

$$\Delta \bar{w} = \eta\, \Theta\left(N \kappa - \bar{w} \cdot \bar{X}^\mu\right) . \bar{X}^\mu$$

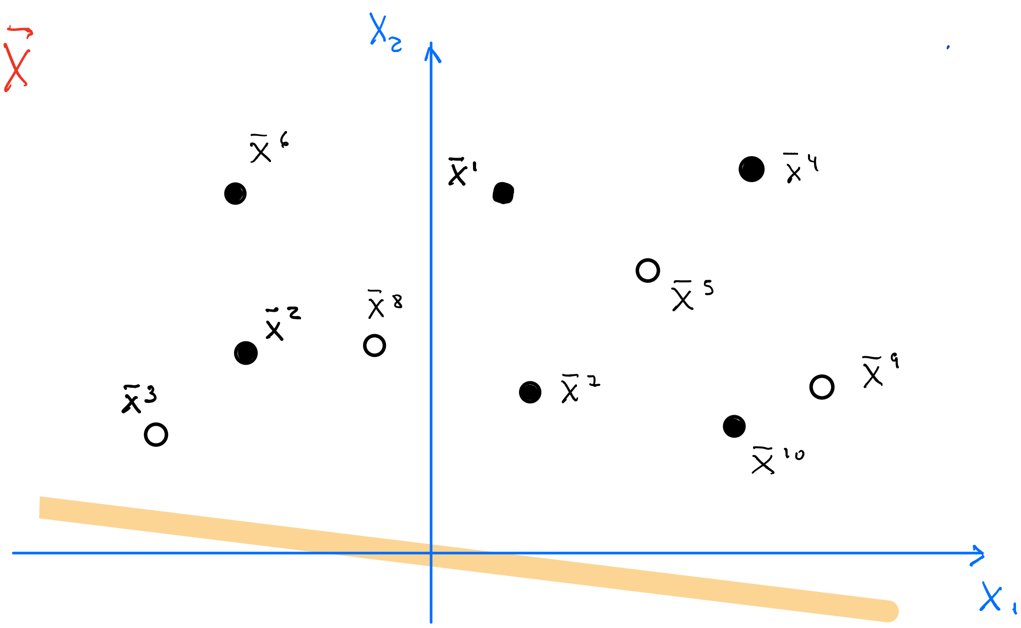
$$= \eta\, \Theta\left(N x - \sum_k w_k \xi_k \zeta^\mu\right) . \zeta^\mu \bar{\xi}^\mu$$

$\bar{w}$ cambia en la dirección de $\bar{x}^\mu$ cuando le presentamos $\bar{\xi}^\mu$.

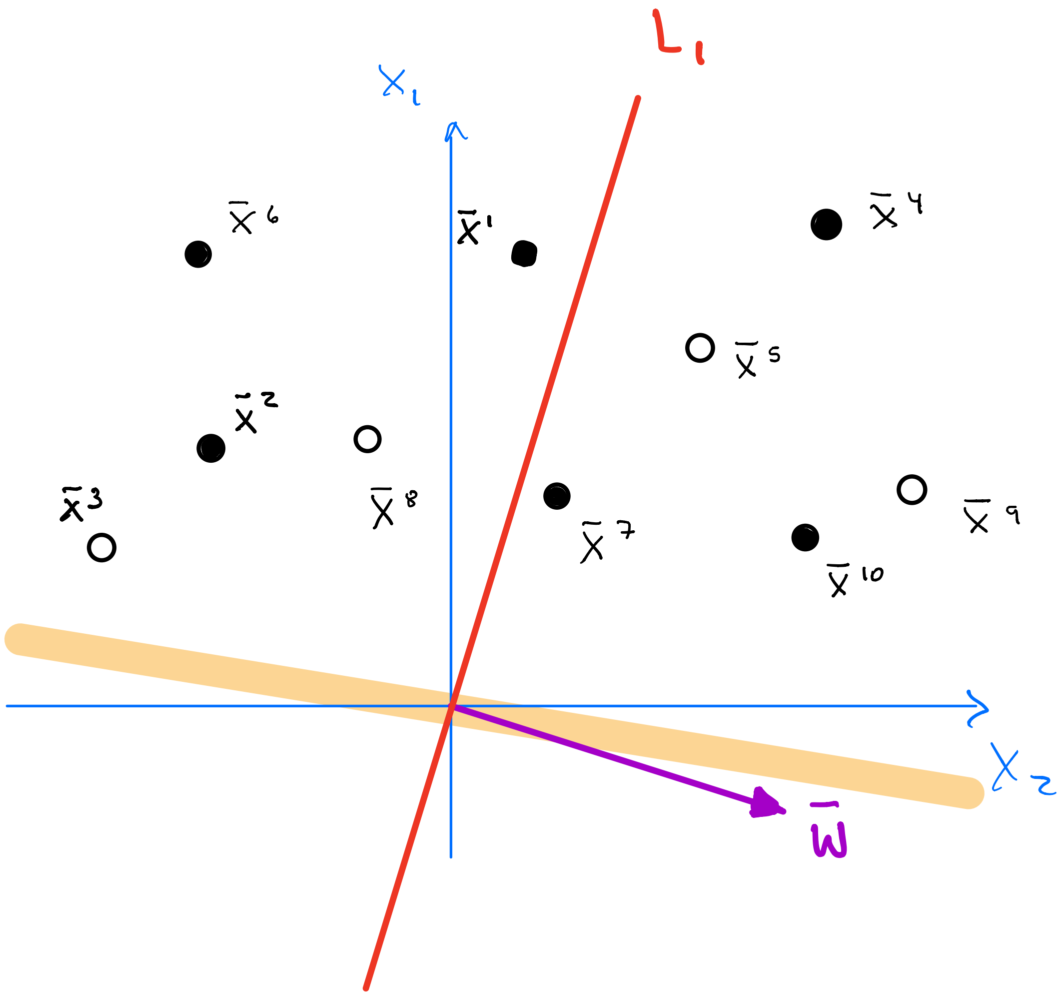
Representación $\bar{\xi}$




Representación $\vec{X}$



Ahora elegimos un valor inicial de $\bar{w}$ y este vector define una recta $L_1$. Vemos que este $\bar{w}$ resuelve bien los ejemplos 4, 5, 7, 9 y 10 pero resuelve mal los ejemplos 1, 2, 3, 6 y 8.



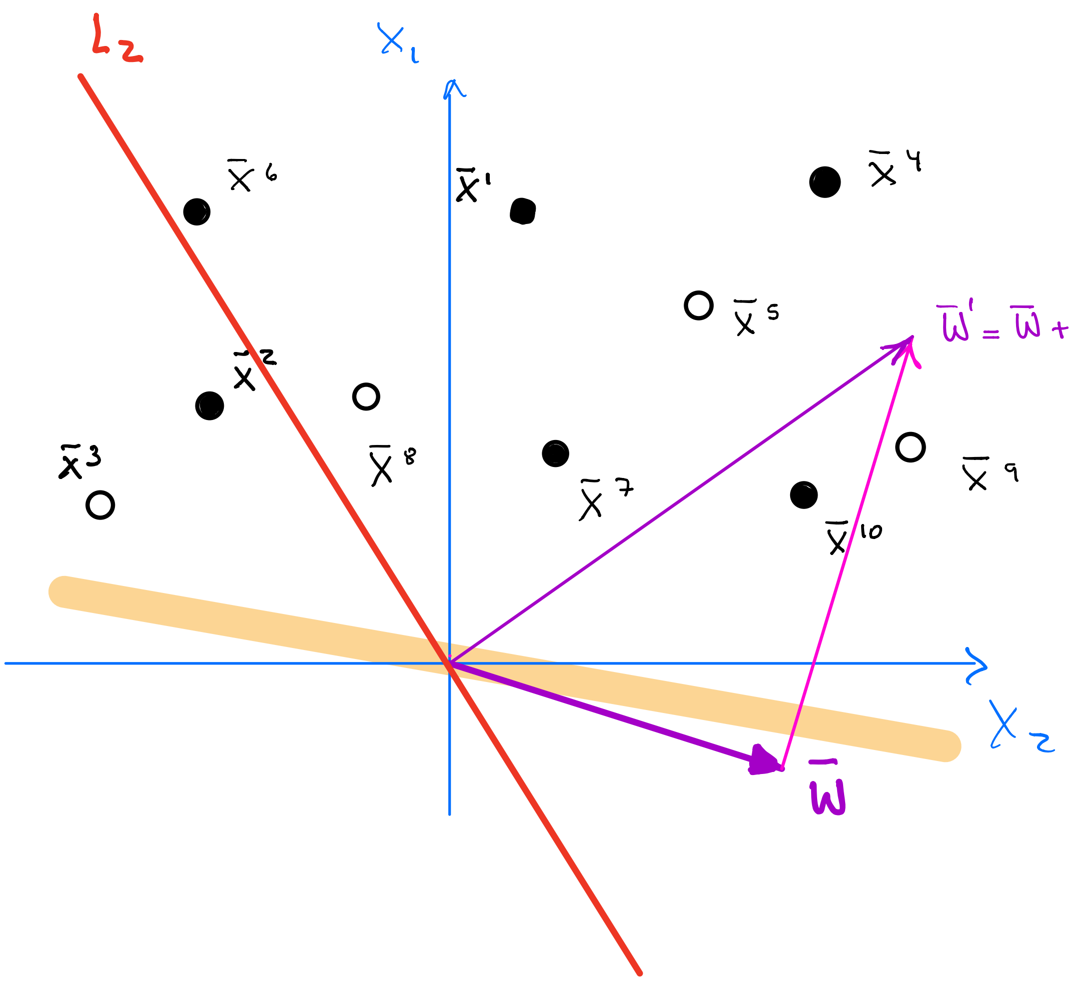
Le mostramos el ejemplo 1, que tiene que dar $\mathcal{O}' = \mathcal{Y}' = 1$ (porque está relleno el punto).

Pero como el ángulo entre $\bar{w}$ y $\bar{x}^1$ es obtuso nos da

$$\underset{+1}{\mathcal{Y}'}\mathcal{O}' = \mathcal{O}' = \text{signo}\,(\bar{w}\cdot\bar{x}) = -1$$

Tomamos $\eta = \frac{1}{2}$ (exageradamente alto)

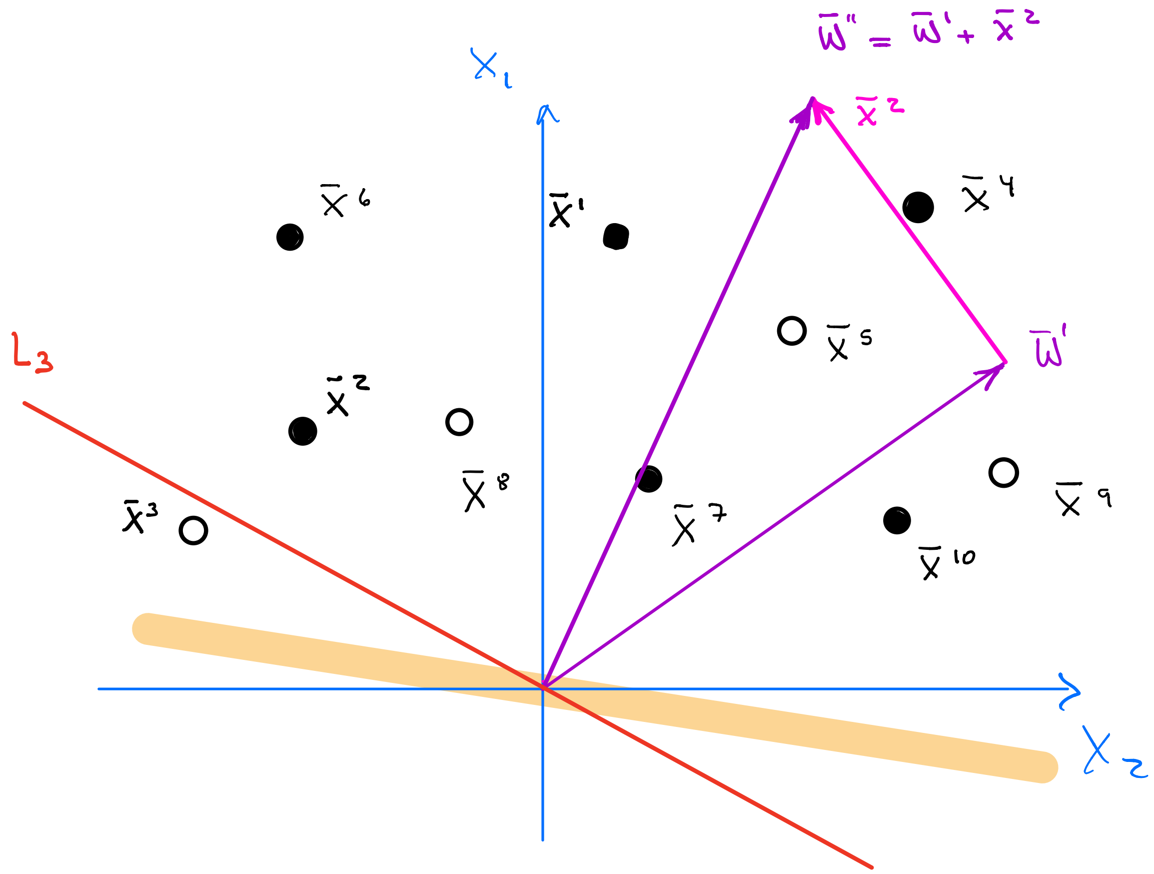
$$\bar{w}' = \bar{w} + \bar{x}^1$$



El nuevo vector $\bar{w}'$ define una nueva recta $L_2$ que ahora resuelve bien los ejemplos 1, 4, 5, 6, 7, 8, 9 y 10 y resuelve mal los ejemplos 2 y 3.

Ahora pasamos a tratar $\mu = 2$ $(\mu = \mu + 1)$.
Vemos que resuelve mal este segundo ejemplo

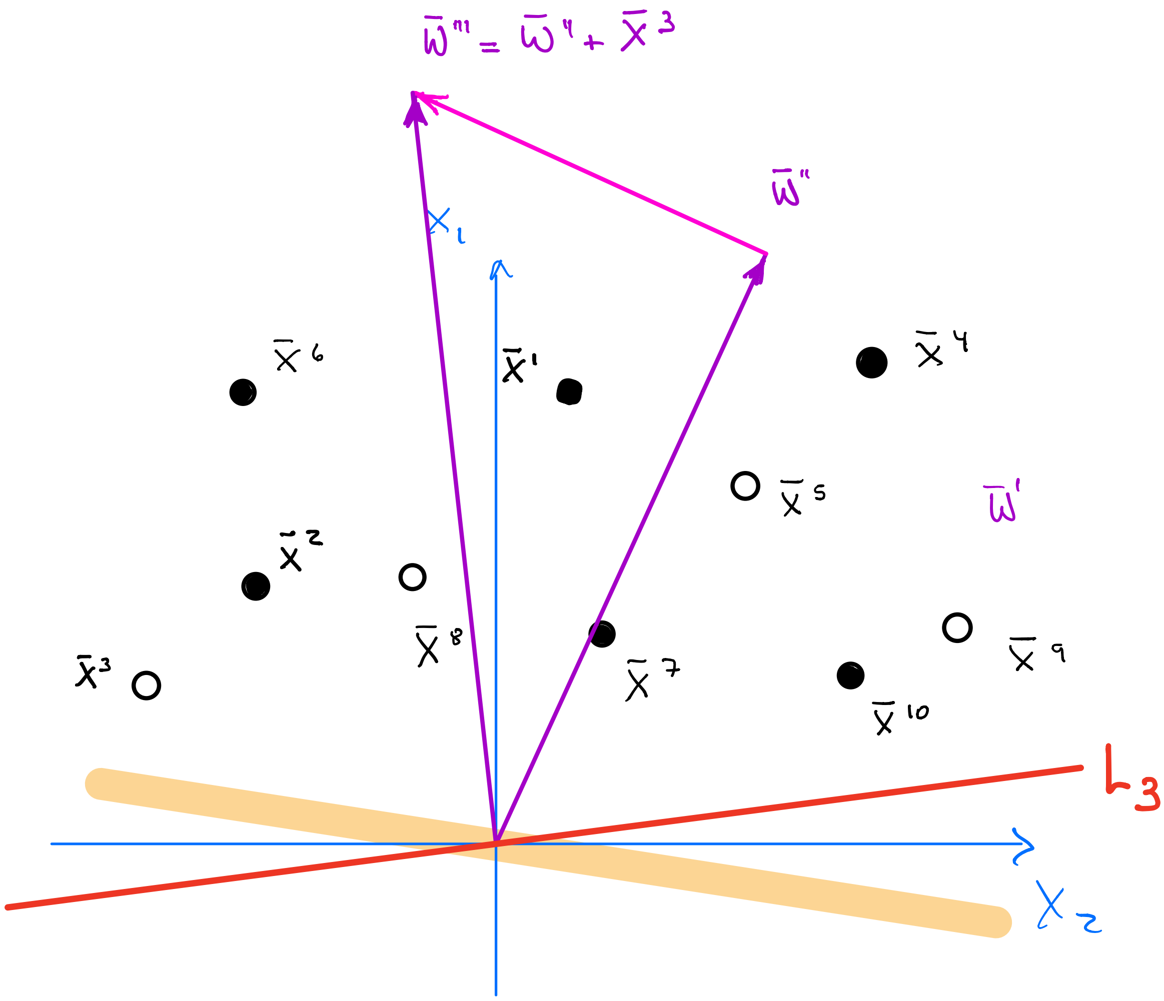
$$\bar{w}'' = \bar{w}' + \bar{x}^2$$



El nuevo vector $\bar{w}''$ define una nueva recta $L_3$ que ahora resuelve bien los ejemplos 1, 2, 4, 5, 6, 7, 8, 9 y 10 y resuelve mal los ejemplos 3.

Ahora pasamos a tratar $\mu = 3$ $(\mu = \mu + 1)$.
Vemos que resuelve mal este segundo ejemplo

$$\bar{w}''' = \bar{w}'' + \bar{x}^3$$

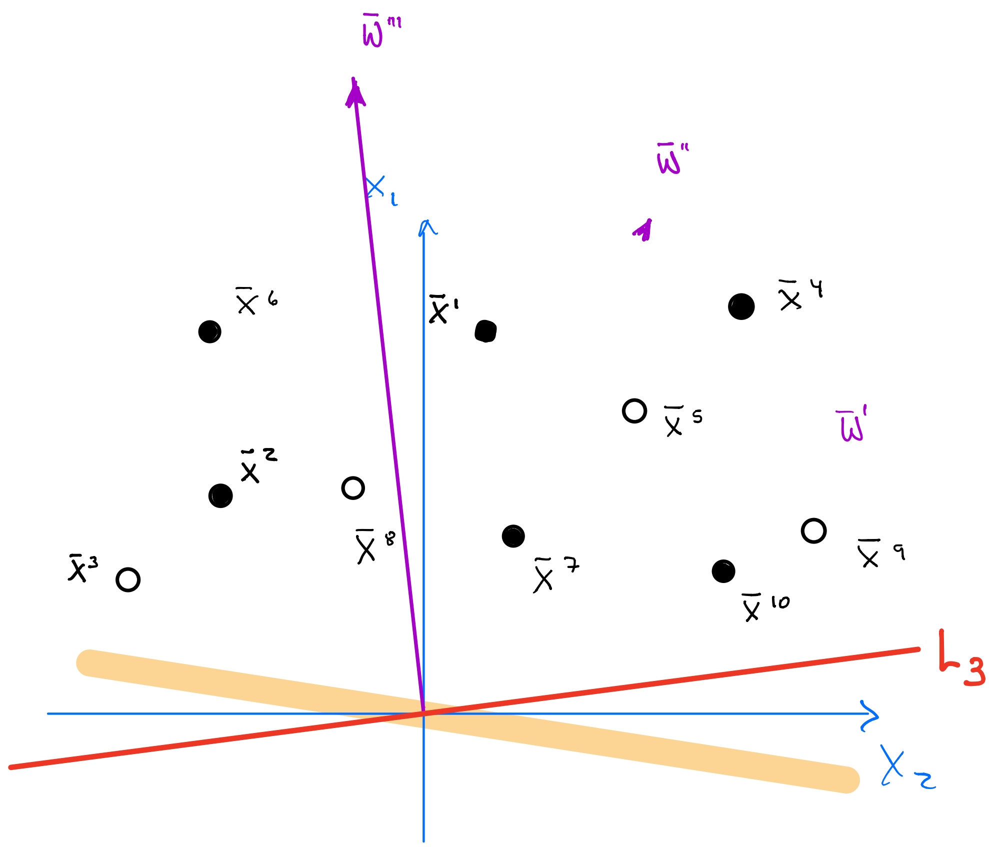


Ahora le muestro $\bar{x}^4$, y lo resuelve bien. Luego voy visitando $\bar{x}^5$, $\bar{x}^6$, $\bar{x}^7$, $\bar{x}^8$, $\bar{x}^9$ y $\bar{x}^1$.

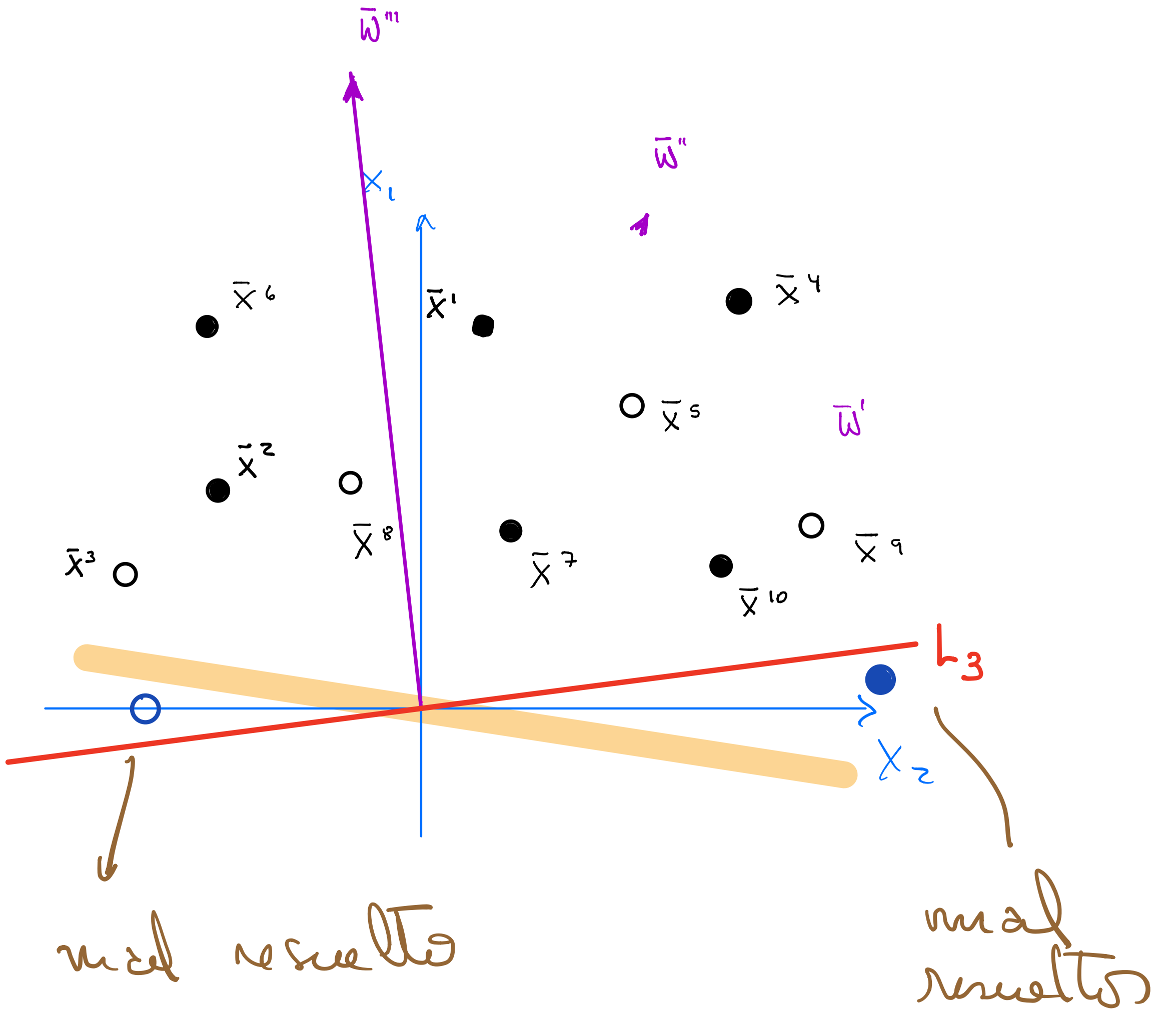
En todos los casos el resultado fue el correcto, gracias al vector $\bar{w}'''$.

Al terminar con último ejemplo de la lista de entrenamiento COMPLETAMOS UNA ÉPOCA

Repito el procedimiento comenzando con $\mu=1$ y terminando en $\mu=10=P$. Como resuelve todos los ejemplos de entrenamiento bien, el final $\bar{w}$ sigue siendo $\bar{w}'''$, y el algoritmo se detiene.



Observen que $\bar{w}'''$ resuelve bien los 10 ejemplos del conjunto de entrenamiento, pero no resuelve exactamente el problema real, que está definido por la línea naranja. Cuando le presentemos al perceptron casos de testeo que están entre la línea naranja y $L_3$, los resolverá mal.



Es muy importante dividir a todos los ejemplos etiquetados en 2 grupos

$\{\bar{X}^{\mu}\}$ con $\mu = 1, 2, \ldots, P$ conjunto de entrenamiento

$\{\bar{X}^{\nu}\}$ con $\nu = 1, 2, \ldots, T$ conjunto de testeo

en general $P > T$

El conjunto de entrenamiento es una muestra del universo de clasificación, y la solución que ajusta el conjunto de entrenamiento seguramente resolverá mal algunos casos fuera del conjunto de entrenamiento.