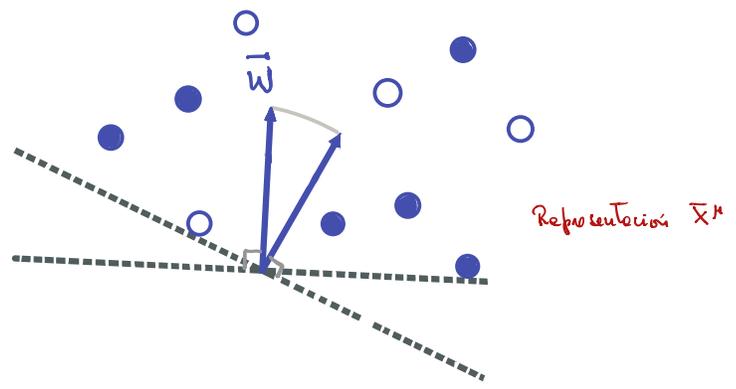
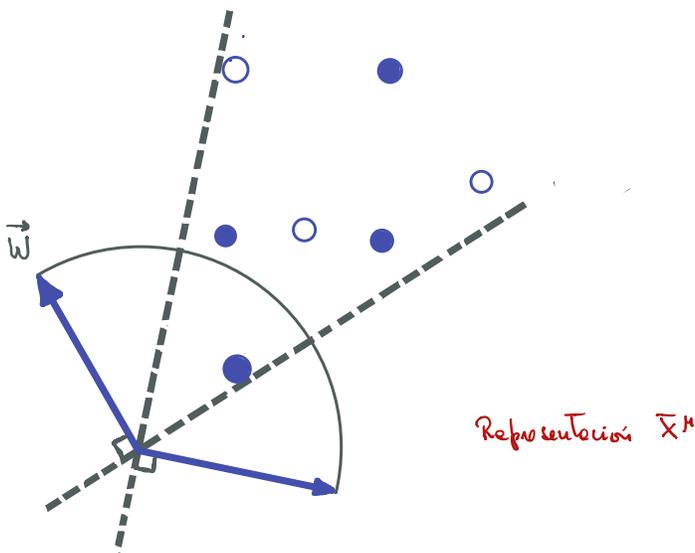


EL PERCEPTRON SIMPLE COMO CLASIFICADOR 2

Nuestra primera regla de aprendizaje parte 2

Claramente, salvo condiciones muy patológicas, no existe un único vector de sinapsis \vec{w} , sino infinitos. Miremos dos buenos ejemplos.



Cualquiera sea el valor de x , \bar{w} puede variar en magnitud y en dirección. Podemos definir la cantidad

$$D(\bar{w}) = \frac{1}{|\bar{w}|} \min_{\mu} \bar{w} \cdot \bar{x}^{\mu}$$

para cada posible \bar{w} . Es la mínima proyección de \bar{w} sobre los \bar{x}^{μ} del conjunto de entrenamiento

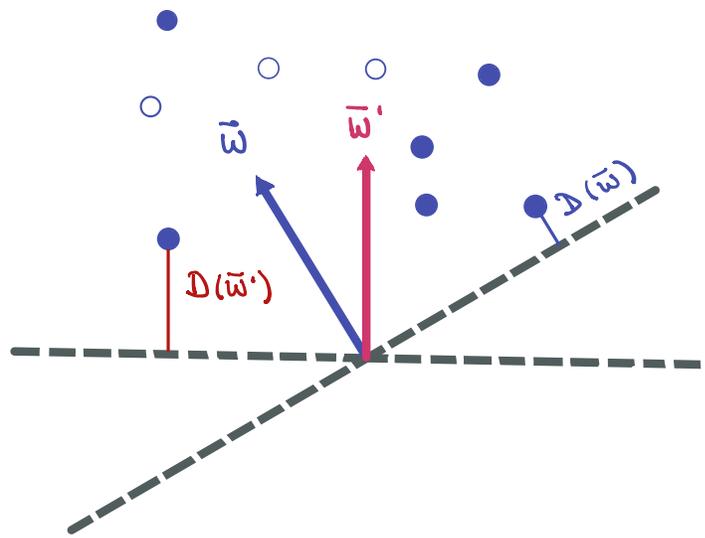
Como está dividido por $|\bar{w}|$ solo depende de la dirección de \bar{w} pero no de su magnitud.

$$\begin{aligned} D(\bar{w}) &= \min_{\mu} \frac{1}{|\bar{w}|} \cdot |\bar{w}| \cdot |\bar{x}^{\mu}| \cos(\theta_{\mu}) \\ &= \min_{\mu} |\bar{x}^{\mu}| \cos(\theta_{\mu}) \end{aligned}$$

Si $D(\bar{w})$, que es un escalar (un número real) es positivo, entonces todos los vectores \bar{x}^{μ} están en el lado correcto. Si es negativo, hay al menos un \bar{x}^{μ} para el cual no funciona el método.

Ahora podemos definir el mejor valor de \bar{w} de entre todos los posibles de la siguiente forma:

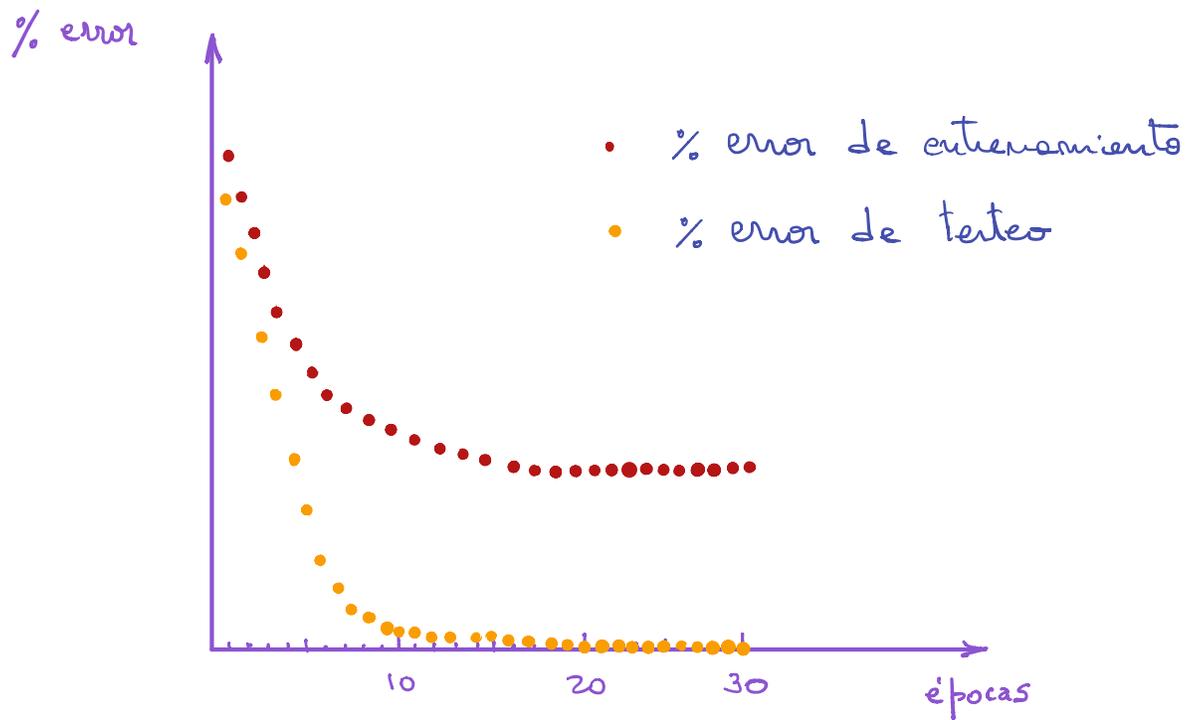
$$D_{\max} \equiv \max_{\bar{w}} D(\bar{w})$$



Como ya dijimos, presentemos cada ejemplo en
secuencia $\mu = 1 \rightarrow 2 \rightarrow 3 \rightarrow \dots \rightarrow P-1 \rightarrow P$.

Cuando presentemos una vez cada elemento de conjunto
de entrenamiento, transcurrió UNA ÉPOCA.

Luego de cada época graficamos el % de
errores en el entrenamiento y en el test,
usando cada conjunto previamente definidos
de P y T elementos



Vemos que en 20 épocas el error de entrenamiento llegó a cero y el error de testeo alcanzó un valor estacionario

Vemos que hoy se usan muy buenas estrategias para elegir el conjunto de entrenamiento y de testeo, en particular **CROSS-VALIDATION**.

Yo les dije que después de presentar cada ejemplo, debemos actualizar cada peso sináptico si la red lo resolvió mal, pero esto no es necesariamente cierto.

Si después de cada presentación y resolución errónea modificamos el vector \bar{w}

$$\bar{w} \xrightarrow{\mu} \bar{w} + \Delta \bar{w}^{\mu}$$

decimos que aprendemos **ON LINE**.

Otra forma posible es ir guardando las correcciones en cada época y corregir al final de la época

$$\bar{w} \xrightarrow{\text{época}} \bar{w} + \Delta \bar{w}^{\text{época}}$$

$$\begin{aligned} \Delta \bar{w}^{\text{época}} &= \Delta \bar{w}^1 + \Delta \bar{w}^2 + \dots + \Delta \bar{w}^p \\ &= \sum_{\mu=1}^p \Delta \bar{w}^{\mu} \end{aligned}$$

Esta forma de aprender se denomina en **BATCH**.
(por lote).

Vemos que esta elección no es única, que hoy se usan otras estrategias (mini batches), y que es un punto muy importante del aprendizaje profundo.

