# REDES NEURONALES 2024

**Clase 15 parte 1**

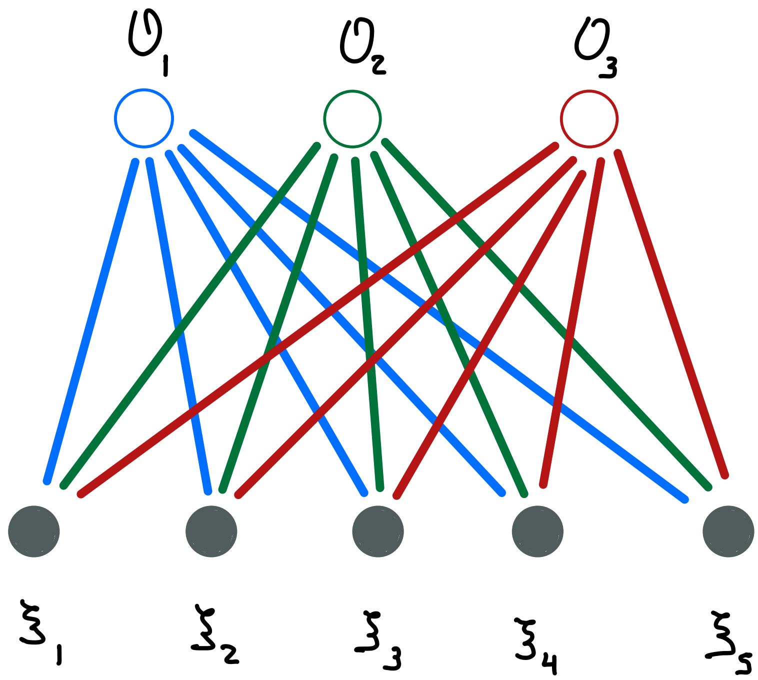
**Lunes 7 de octubre 2024**


**FAMAF, Universidad Nacional de Córdoba**

**Instituto de Física Enrique Gaviola (UNC-CONICET)**


M10/10/23 c14p1

# EL PERCEPTRON DE UNA ÚNICA CAPA

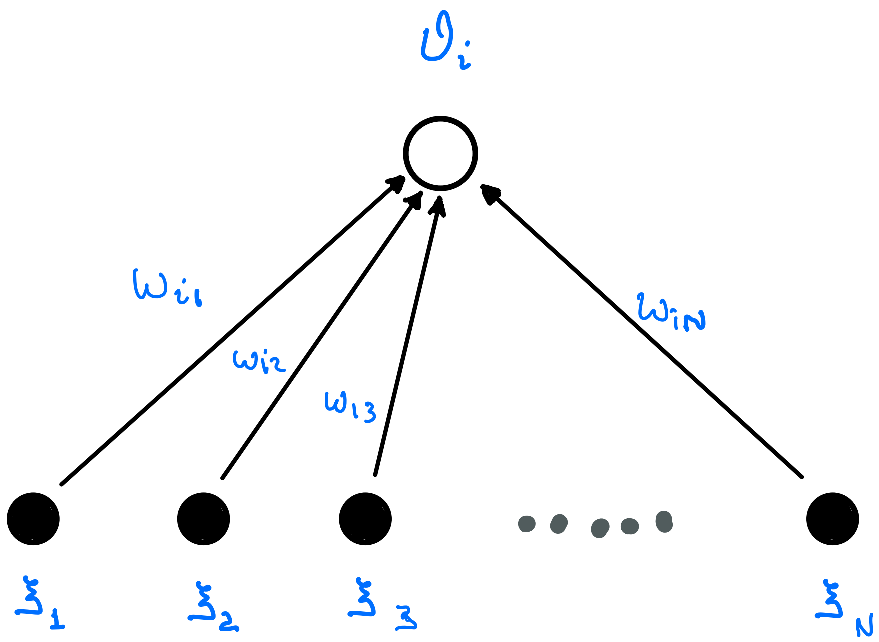



Consideremos una arquitectura feed-forward como la de la figura de arriba. Observamos que en realidad hay ***tres*** subredes desacopladas con una única salida cada una, pues cada subred tiene su propia neurona de salida y sus propias sinapsis, que no se comparten con las otras subredes. Hemos pintado las sinapsis de cada subred con un color distintivo, para que sea más fácil identificarlas. Esto nos dice que podemos tratar a cada subred en forma independiente de las otras. O sea, podemos entrenar cada subred por separado.

**Nota 1:** aquí ponemos tres neuronas de salida pero es solo un ejemplo. Pueden ser tantas como queramos, o mejor, como necesitemos.

**Nota 2:** veremos más adelante que si la red tiene más de una capa, no podremos usar este desacople. No obstante será muy instructivo considerar primero estas redes de una única capa.

# EL PERCEPTRÓN SIMPLE




Una sola neurona de salida $\qquad O_i = g(\bar{w} \cdot \bar{\xi} - \mu_i)$

*N* entradas (no neuronas) $\qquad \xi_k \qquad$ *(k=1, 2, ... ,N)*

El conjunto de entradas $\{\xi_1, \xi_2, \ldots, \xi_N\}$ se piensa como un vector arbitrario en $\boldsymbol{R}^N$

$$\bar{\xi} = (\xi_1, \xi_2, \xi_3, \ldots, \xi_N) \in \mathbb{R}^N$$

Las variables de entrada pueden tomar valores discretos o continuos, acotados o no.

El rango de valores y el tipo de valores que puede tomar la neurona de salida $O_i$ dependerá de la función de activación *g(x)* .

$$O_i = g\left(h_i(t) - \mu_i\right)$$
$$= g\left(\sum_{k=1}^{N} w_{ik}\,\xi_k - \mu_i\right)$$
$$= g\left(\overline{w}_i \cdot \overline{\xi} - \mu_i\right)$$

Buscamos el vector $\overline{\boldsymbol{W}}_i$

$W_{ik}$ : Eficacia sináptica entre la entrada *k* y la neurona de salida *i*

$\mu_i$ : Umbral de activación de la neurona a *i*.

## Aprendizaje supervisado en un perceptron simple

Supongamos que deseamos modelar cierto proceso que asigna a vectores en $\mathbb{R}^N$ un único valor real $O_i$. No tenemos ninguna información sobre la expresión matemática de dicha asignación, pero podemos imaginarla como en función real que come una $N$-upla de números reales y devuelve un número real. Llamemos $f$ a esa relación. Entonces:

$$f(\bar{\xi}) = O \qquad\qquad f : \mathbb{R}^N \longrightarrow \mathbb{R}$$

Supongamos que la única información de la cual disponemos es el valor que está función asigna a cierto conjunto de valores de entrada. Supongamos que son $p$ los datos de entrada cuya solución conozco y los enumeramos con la letra griega $\alpha$ .

$$\bar{\xi}^\alpha = (\xi_1^\alpha, \xi_2^\alpha, \ldots, \xi_N^\alpha) \longrightarrow \zeta_i^\alpha \in \mathbb{R}$$

Experto

$$\alpha = 1, 2, 3, \ldots, p$$

$$\left\{ \bar{\xi}^\alpha, \zeta_i^\alpha \right\}_{\alpha=1}^{p} \longleftarrow$$ Conjunto de datos etiquetados

$\zeta_i^\alpha$ : representa el valor que debería tomar $O_i$ cuando le presentamos $\bar{\xi}^\alpha$. Pero la neurona de salida tomará el valor

$$
\begin{aligned}
O_i^\alpha &= g\left(h_i^\alpha(t) - \mu_i\right) \\
&= g\left(\sum_{k=1}^{N} w_{ik}\,\xi_k^\alpha - \mu_i\right) \\
&= g\left(\bar{w}_i \cdot \bar{\xi}^\alpha - \mu_i\right)
\end{aligned}
$$

Nuestro desafío ahora es encontrar un conjunto posible de eficacias sinápticas $w_{ik}$ y un valor del umbral $\mu_i$ que sean capaces de hacer que nuestra red asigne correctamente las entradas y las respectivas salidas para el conjunto de datos de entrenamiento.

Si tenemos suerte y si el conjunto de entrenamiento es suficientemente grande y variado, esta red, que aprendió el conjunto de entrenamiento, será capaz de resolver bien incluso los casos que nunca vió. O sea, podrá asignar una salida correcta para una entrada arbitraria. Diremos que **generaliza**.

Observemos que podemos pensar a las eficacias sinápticas, que recordamos pueden tomar cualquier valor real, como formando un vector en $R$ .

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

$$\bar{w}_i = (w_{i1}, w_{i2}, \dots, w_{iN}) \in \mathbb{R}^N$$

Ambos vectores, $\bar{w}_i$ y $\bar{\xi}$ viven en el mismo espacio.

**¡Son vecinos!**

Si tenemos $N$ entradas o features y $M$ neuronas de salida, podemos aprender una función arbitraria

$$f : \mathbb{R}^N \longrightarrow \mathbb{R}^M$$