

# Machine learning y redes neuronales

## 1 MCP

Con el fin de diseñar una IA, se definió a una célula nerviosa como una puerta lógica con salidas binarias. A esta célula nerviosa entran múltiples datos por las dendritas, estos datos se agrupan y si la señal acumulada supera cierto umbral, se genera una salida transmitida por el axón. A la abstracción matemática de este modelo de neurona se le llama MCP.

Una neurona artificial es usada para tareas de clasificación binaria, donde existen dos clases, la primera clase, clase positiva(1) y la segunda, clase negativa(-1).

Una función de decisión  $\phi(Z)$  es aquella que determina la clase a la que pertenece cada muestra.  $Z$  está dada por una combinación lineal de ciertos valores de entrada( $X$ ) y sus pesos correspondientes( $W$ ).

$$Z = \sum_{k=1}^N X_k W_k$$

Es decir que  $Z$  es el producto punto entre los vectores  $X$  y  $W$ .

$$Z = X \cdot W$$

Si  $\phi(Z)$  es mayor que un umbral definido  $\theta$  en una muestra determinada, se dice que esta muestra pertenece a la clase 1. Para fines prácticos podemos igualar a  $X_0 = 1$  y  $W_0 = -\theta$ . Un umbral negativo también se le puede llamar cesgo.

Si  $\phi(Z) > 0$  entonces la muestra pertenece a la clase positiva. Si  $\phi(Z) \leq 0$  entonces la muestra pertenece a la clase negativa.

## 2 Aprendizaje del Perceptrón

El perceptrón es el tipo más simple de red neuronal. El perceptrón umbralizado y el MCP imitan el funcionamiento de una neurona biológica. Estos modelos son capaces de clasificar muestras dentro de dos clases siguiendo la **Regla del perceptrón inicial de Rosenblatt**, la cual se reduce a dos pasos:

1. Iniciar los pesos a números aleatorios mayores pero cercanos a 0
2. Por cada muestra en el entrenamiento  $X^i$  se calcula el valor  $\hat{y}$  y se actualizan los pesos  $W$

Para actualizar los pesos usamos la formula:

$$W_j := W_j + \Delta W_j$$

Para obtener  $\Delta W_j$  usamos la siguiente regla:

$$\Delta W_j = \eta(y^i - \hat{y}^i)X_j^i$$

Donde  $\eta$  = rango de aprendizaje,  $y^i$  = etiqueta de clase de la muestra y  $\hat{y}$  = etiqueta de clase predicha.

Algo importante a destacar es que el perceptron tiene mayor precision si las dos clases son linealmente separables y si el rango de aprendizaje es pequeno. Si las dos clases no son linealmente separables podemos ajustar un limite de correcciones a los pesos y un umbral para el numero de clasificaciones erroneas. Si no se hace esto, el perceptron seguira ajustando los pesos infinitamente. Se dice que el modelo convergio cuando logra predecir las etiquetas de las muestras de manera correcta y sin errores en una epoca.