Modelos Lineales para Clasificación

Aprendizaje Automático

Dr. Luis Eduardo Falcón Morales
ITESM
Campus Guadalajara

Hiperplanos y Vectores Normales

En el caso de un modelo lineal con respecto a las variables de entrada x_k y los pesos a determinar ω_k, la función de activación f puede expresarse como:

$$f(\omega_0 + \omega_1 x_1 + \omega_2 x_2 + \cdots + \omega_m x_m) = y$$

donde f es en general una función no lineal.

Usando notación vectorial,

$$\vec{\omega} = (\omega_1, \omega_2, \dots, \omega_m)^T, \qquad \vec{x} = (x_1, x_2, \dots, x_m)^T$$

la expresión

$$\omega_0 + \omega_1 x_1 + \omega_2 x_2 + \cdots + \omega_m x_m$$

puede escribirse como

$$\omega_0 + \vec{\omega}^T \vec{x}$$

O bien, usando coordenadas homogéneas

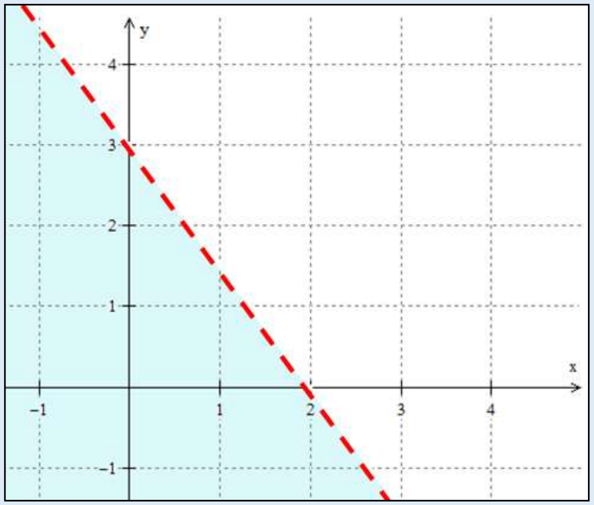
$$\vec{\omega} = (\omega_0, \omega_1, \omega_2, \dots, \omega_m)^T, \qquad \vec{x} = (1, x_1, x_2, \dots, x_m)^T$$

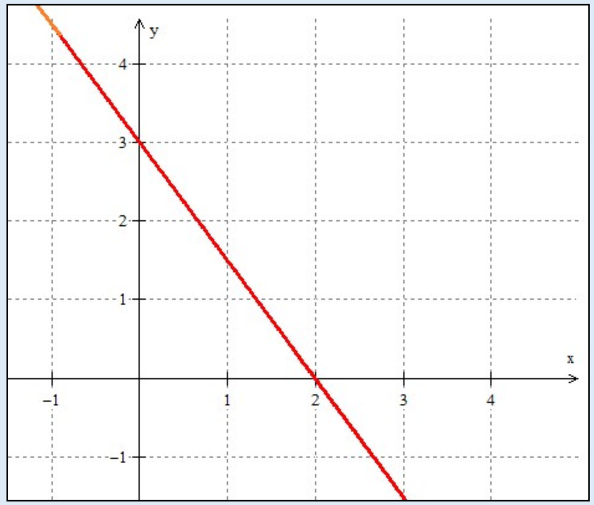
la expresión indicada y por lo tanto la función de activación quedarían ahora como

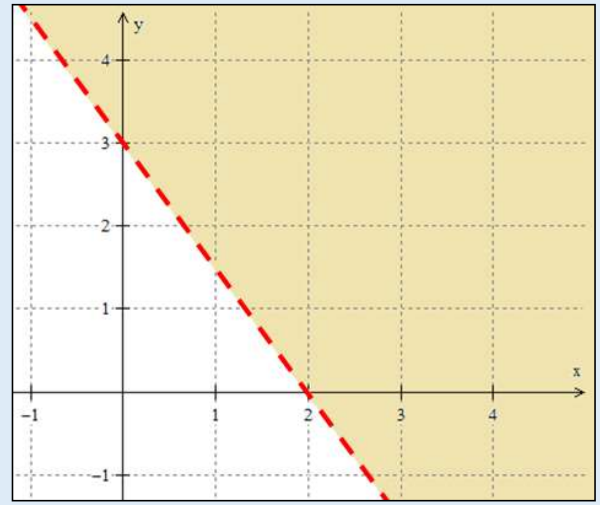
$$f(\vec{\omega}^T \vec{x}) = y$$

Igualdades y Desigualdades

$$3x + 2y - 6 < 0 \qquad 3x + 2y - 6 = 0 \qquad 3x + 2y - 6 > 0$$

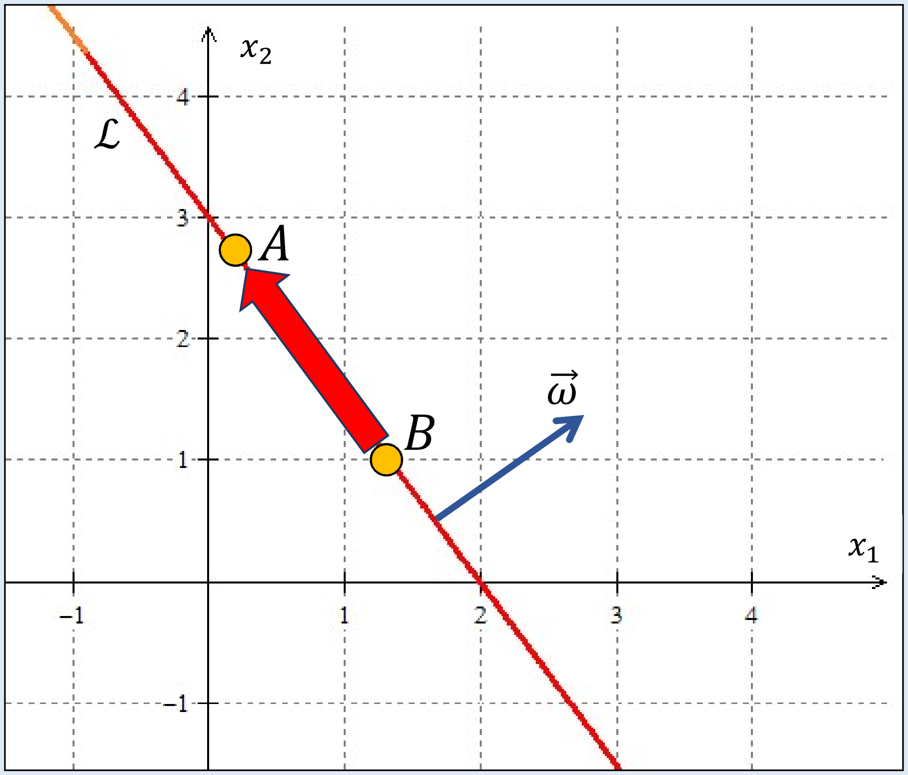






Dado un punto $(p, q) \in \mathbb{R}^2$,

¿cómo determinar algebraicamente en cuál de estas tres regiones se encuentra dicho punto?



Dada la ecuación de una recta $\mathcal{L}$ en $\mathbb{R}^2$:

$$\mathcal{L}:\ \omega_0 + \omega_1 x_1 + \omega_2 x_2 = 0$$

¿cuál es el significado de los pesos ω_k?

Si los puntos $A = (a_1, a_2)$, $B = (b_1, b_2)$ están sobre la recta $\mathcal{L}$, entonces:

$$\omega_0 + \omega_1 a_1 + \omega_2 a_2 = 0$$

$$\omega_0 + \omega_1 b_1 + \omega_2 b_2 = 0$$

Y restando ambas ecuaciones, obtenemos que:

$$\omega_1(a_1 - b_1) + \omega_2(a_2 - b_2) = 0$$

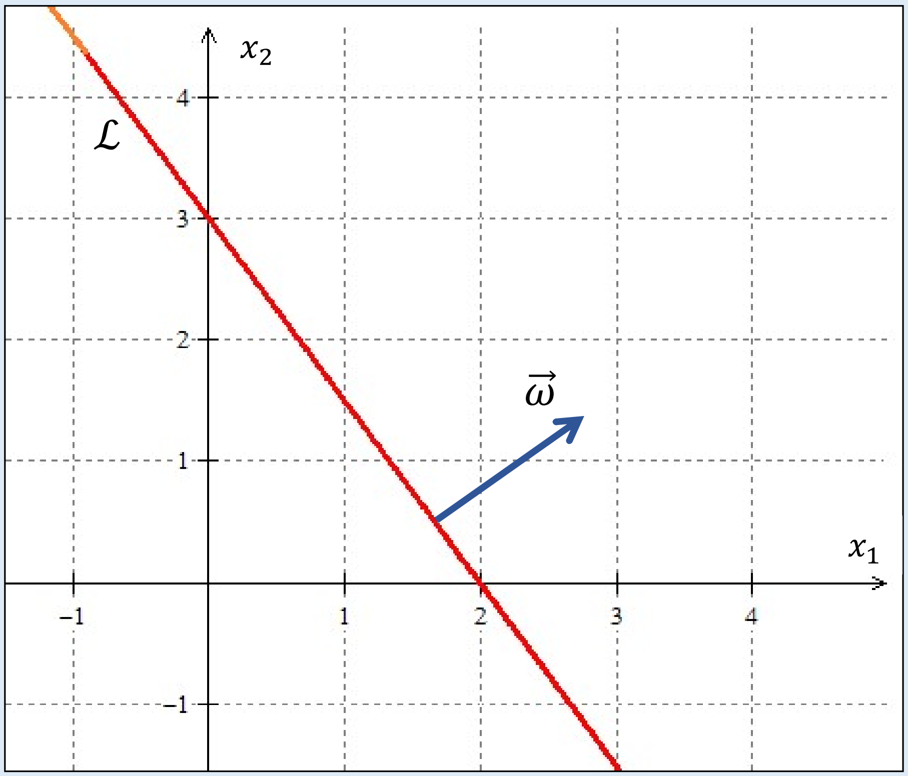
es decir, el vector

$$\overrightarrow{BA} = (a_1 - b_1)\hat{\imath} + (a_2 - b_2)\hat{\jmath}$$

es ortogonal al vector

$$\vec{\omega} = \omega_1 \hat{\imath} + \omega_2 \hat{\jmath}$$

En resumen, el vector $\vec{\omega} = \omega_1 \hat{\imath} + \omega_2 \hat{\jmath}$ siempre es ortogonal a la recta $\mathcal{L}:\ \omega_0 + \omega_1 x_1 + \omega_2 x_2 = 0$



Entonces, dada la ecuación de una recta $\mathcal{L}$ en $\mathbb{R}^2$:

$$\mathcal{L}: \; \omega_0 + \omega_1 x_1 + \omega_2 x_2 = 0$$

el vector:

$$\vec{\omega} = \omega_1 \hat{\imath} + \omega_2 \hat{\jmath}$$

es un vector ortogonal a la recta $\mathcal{L}$.

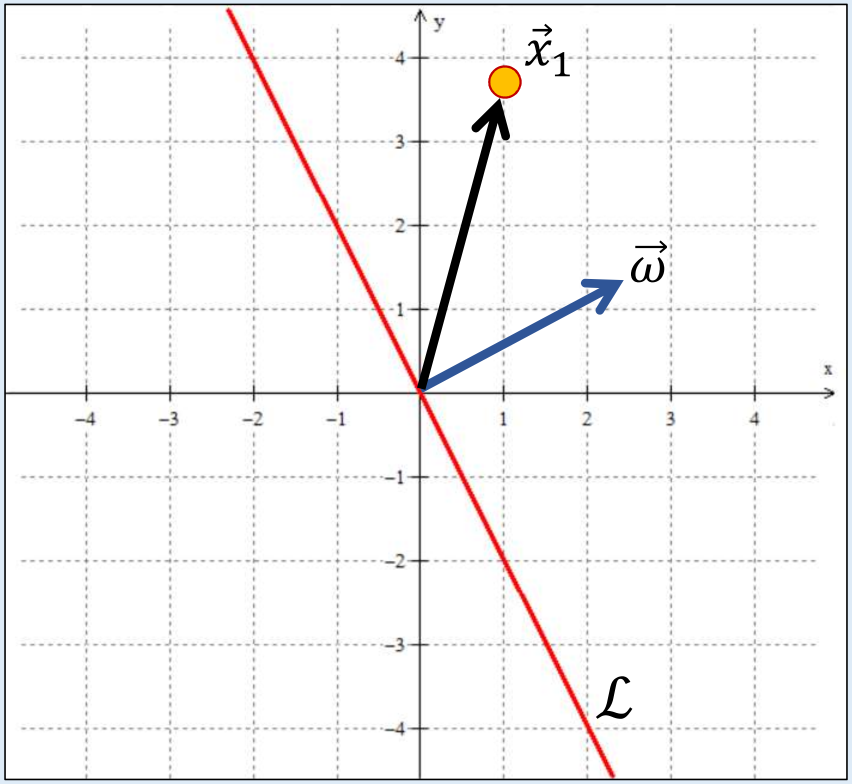
Es decir, este vector ortogonal $\vec{\omega}$ determina la orientación de la recta.

Y entonces, vectorialmente la ecuación de la recta puede escribirse como:

$$\vec{\omega}^T \vec{x} = -\omega_0$$

Al valor ω_0 lo llamaremos *bias*, y al valor $-\omega_0$ lo llamaremos **umbral**, ya que es el valor que determina la frontera entre la región superior e inferior de la recta.

Interpretación geométrica de un hiperplano/recta $\mathcal{L}$ y su vector ortogonal $\vec{\omega}$: **caso ángulo agudo**



Ecuación de la recta en $\mathbb{R}^2$: $\omega_1 x_1 + \omega_2 x_2 = 0$

Vector ortogonal a la recta: $\vec{\omega} = \omega_1 \hat{\imath} + \omega_2 \hat{\jmath}$

Por la definición de producto interior y suponiendo que los vectores $\vec{\omega}$, $\vec{x}_1$ están del mismo lado del hiperplano como se muestra en la figura, entonces:

$$signo(\vec{\omega}^T \vec{x}_1) = signo(|\vec{\omega}||\vec{x}_1| cos\theta_1)$$

$$= signo(cos\theta_1)$$

$$= +1$$

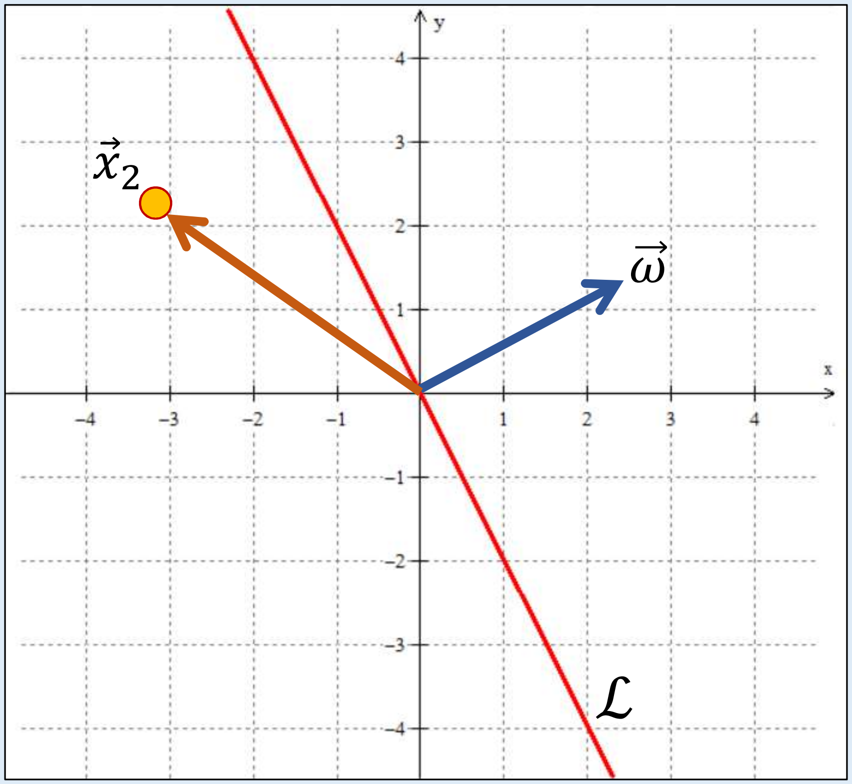
O bien, usando notación matricial:

$$\vec{x}_1 = (1, x_{11}, x_{12})^T$$

$$\vec{\omega} = (0, \omega_1, \omega_2)^T$$

entonces $\vec{\omega}^T \vec{x}_1 > 0$ cuando ambos vectores se encuentran del mismo lado del hiperplano $\mathcal{L}$, es decir, cuando el ángulo entre ellos es menor a $90°$.

Interpretación geométrica de un hiperplano/recta $\mathcal{L}$ y su vector ortogonal $\vec{\omega}$: **caso ángulo obtuso**



Ecuación de la recta en $\mathbb{R}^2$: $\omega_1 x_1 + \omega_2 x_2 = 0$

Vector ortogonal a la recta: $\vec{\omega} = \omega_1 \hat{\imath} + \omega_2 \hat{\jmath}$

Análogamente, si consideramos ahora los vectores $\vec{\omega}$ y $\vec{x}_2$ en lados opuestos del hiperplano como se muestran en la figura, entonces:

$$signo(\vec{\omega}^T \vec{x}_2) = signo(|\vec{\omega}||\vec{x}_2|cos\theta_2)$$

$$= signo(cos\theta_2)$$

$$= -1$$

Y nuevamente, si usamos coordenadas homogéneas:

$$\vec{x}_2 = (1, x_{21}, x_{22})^T$$

$$\vec{\omega} = (0, \omega_1, \omega_2)^T$$

entonces $\vec{\omega}^T \vec{x}_2 < 0$ cuando los vectores se encuentran en lados opuestos del hiperplano $\mathcal{L}$, es decir, cuando el ángulo entre ellos es mayor a 90°.