

Regresión Logística

La regresión logística es una herramienta estadística y de minería de datos que predice la probabilidad de un resultado que sólo puede tener dos valores (es decir, una variable Dicotómica). La predicción se basa en el uso de uno o varios predictores (numéricos y categóricos). El análisis de regresión logística se enmarca en el conjunto de Modelos Lineales Generalizados (GLM por sus siglas en inglés) que usa como función de enlace la función Logit, la cual se define a continuación:

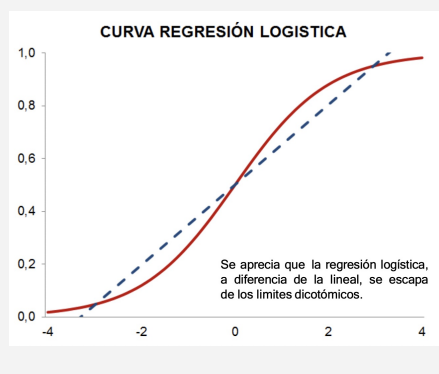
Regresión Logística

$$F(x^t \beta) = \frac{\exp(x^t \beta)}{1 + \exp(x^t \beta)} \quad p = \frac{1}{1 + e^{-(b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p)}}$$

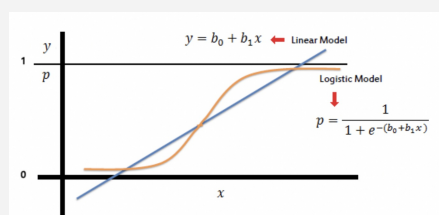
Por lo tanto, se deben mantener las siguientes consideraciones:

- La Variable Dependiente debe ser dicotómica (poseer valores 0 y 1).
- Las Variables Independientes pueden ser cuantitativas o cualitativas, sin embargo, estas últimas se deben transformar a variables cuantitativas.
- La ecuación oscila entre 0 y 1.

Regresión Logística



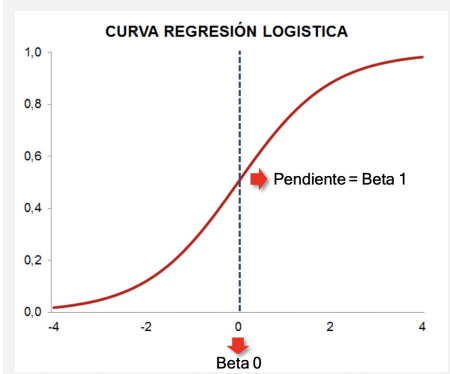
Regresión Logística



Propiedades de la Función

Por otro lado, una regresión logística produce una curva logística, que se limita a valores entre 0 y 1. La regresión logística es similar a una regresión lineal, pero la curva se construye utilizando el logaritmo natural de las "probabilidades" de la variable de destino, en lugar de la probabilidad. Por otra parte, los predictores no tienen que tener una distribución normal o tienen la misma diferencia en cada grupo.

Regresión Logística



Propiedades de la Función

En la regresión logística simple la constante (Beta 0) mueve la curva de la izquierda a la derecha y la pendiente (Beta 1) define la pendiente de la curva.