

QUE ES LA REGRESIÓN LOGÍSTICA

La historia de la regresión logística tiene sus raíces en el desarrollo de métodos estadísticos y la necesidad de modelar relaciones complejas entre variables en el contexto de eventos binarios.

Desarrollo temprano:

- Aunque la regresión logística como tal se desarrolló en el siglo XX, los conceptos fundamentales de la regresión estaban siendo explorados en el siglo XIX, cuando científicos como Francis Galton y Karl Pearson trabajaron en el desarrollo de métodos estadísticos.

Siglo XX:

- La regresión logística moderna comenzó a tomar forma a principios del siglo XX. En la década de **1940**, el estadístico y genetista británico Ronald A. Fisher propuso un enfoque para ajustar una curva logística a datos en el campo de la genética. Fisher también es conocido por sus contribuciones al análisis de varianza y al diseño experimental.

Década de 1950:

- El estadístico Daniel G. Cox desarrolló el modelo de regresión logística en la década de **1950** mientras trabajaba en la investigación biomédica en el Instituto de Estadística Biomédica de la Universidad de Stanford. Cox publicó un artículo en **1958** donde describió cómo aplicar la regresión logística para analizar la relación entre la presencia de una enfermedad y múltiples factores de riesgo.

Ampliación a otras disciplinas:

- A medida que avanzaba la estadística y la capacidad de cómputo, la regresión logística se extendió a una variedad de campos, como la epidemiología, la psicología, la economía y la ciencia social. Se convirtió en una herramienta valiosa para analizar y predecir resultados binarios en una variedad de aplicaciones.

Avances en la informática:

- Con la creciente disponibilidad de computadoras y software estadístico en las últimas décadas del siglo **XX**, la regresión logística se volvió más accesible y aplicable. Los investigadores y profesionales podían realizar análisis más complejos y precisos utilizando esta técnica.

Aplicación en el aprendizaje automático:

- A medida que el campo del aprendizaje automático creció en importancia, la regresión logística se convirtió en una parte esencial de los modelos de clasificación. Junto con otros algoritmos, la regresión logística se utiliza para realizar clasificaciones en una amplia gama de aplicaciones, como el procesamiento del lenguaje natural, la visión por computadora y la medicina.

LA REGRESIÓN LOGÍSTICA

La regresión logística es una técnica estadística utilizada para modelar y analizar relaciones entre variables, especialmente cuando la variable de resultado o dependiente es categórica. A diferencia de la regresión lineal, que se emplea para predecir valores numéricos continuos, **la regresión logística se enfoca en predecir la probabilidad de que una observación pertenezca a una categoría en particular.**

Importancia de la regresión logística:

Modelo para variables categóricas:

- La regresión logística es esencial cuando se trabaja con variables categóricas, como sí/no, éxito/fracaso, verdadero/falso, etc. Permite analizar cómo diferentes variables predictoras influyen en la probabilidad de que ocurra un evento.

Aplicaciones en diversas disciplinas:

- La regresión logística tiene aplicaciones en una amplia gama de campos, como la medicina (diagnóstico de enfermedades), la economía (análisis de factores que influyen en decisiones), el marketing (segmentación de clientes) y muchas otras áreas donde se necesite comprender y predecir eventos categóricos.

Interpretación de resultados:

- La regresión logística proporciona coeficientes que indican cómo las variables predictoras influyen en la probabilidad de éxito del evento. Estos coeficientes se pueden interpretar para entender cómo los cambios en las variables predictoras afectan las probabilidades de un resultado en particular.

Características de la regresión logística:

1. Variable dependiente binaria:

En la regresión logística, la variable de resultado debe ser binaria, es decir, tener dos categorías mutuamente excluyentes, como 0/1 o sí/no.

2. Función logit:

La regresión logística utiliza la función logit para transformar la relación lineal entre las variables predictoras y la probabilidad de éxito en una forma que abarca todo el rango de números reales.

3. Coeficientes de regresión:

Los coeficientes de regresión en la regresión logística indican cómo los cambios en las variables predictoras se asocian con el logaritmo de la razón de probabilidades. Estos coeficientes se pueden exponenciar para obtener las razones de probabilidades reales.

4. Modelo de probabilidad:

El modelo de regresión logística estima la probabilidad de éxito para cada observación. Esta probabilidad se relaciona con las variables predictoras a través de la función logit.

5. Hipótesis y pruebas:

Al igual que en otros modelos estadísticos, es posible realizar pruebas de hipótesis para determinar si los coeficientes son estadísticamente significativos y si las variables predictoras tienen un impacto significativo en la probabilidad de éxito.

6. Supuestos moderados:

Aunque la regresión logística no requiere los mismos supuestos estrictos que la regresión lineal (como la normalidad y la homocedasticidad), es importante considerar posibles problemas de multicolinealidad y selección de variables para obtener resultados precisos.