Notebook_Clase2_y_3_Regresiones

Diego Figueroa

2025-05-17

Regresión Lineal

La regresión lineal es una técnica estadística fundamental utilizada para modelar la relación lineal entre una variable dependiente (o de respuesta) y una o más variables independientes (o predictoras). El objetivo es encontrar la mejor línea recta (en el caso univariado) o hiperplano (en el caso multivariado) que describa cómo la variable dependiente cambia en función de las variables independientes.

Regresión Lineal Univariada

La regresión lineal univariada involucra una única variable predictora (X) para modelar una variable de respuesta (Y). El modelo se expresa de la siguiente manera:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

Donde:

- Y_i Representa el valor de la variable dependiente para la i-ésima observación.
- X_i Representa el valor de la variable independiente para la i-ésima observación.
- β_0 Es la intersección (el valor de Y cuando X es 0).
- β_1 Es la pendiente (el cambio en Y por cada unidad de cambio en X).
- ϵ_i Es el error aleatorio o residuo para la i-ésima observación, que representa la diferencia entre el valor observado y el valor predicho por el modelo. Se asume que estos errores tienen una media de cero y una varianza constante.

El objetivo es estimar los coeficientes β_0 y β_1 que minimizan la suma de los cuadrados de los residuos (método de mínimos cuadrados ordinarios - OLS).

Regresión Lineal Múltiple

La regresión lineal múltiple extiende el concepto a más de una variable predictora. El modelo general con p variables predictoras se escribe como:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \epsilon_i$$

En el modelo de regresión lineal múltiple:

- Y_i representa el valor de la variable dependiente para la i-ésima observación.
- X_{ij} representa el valor de la j-ésima variable independiente para la i-ésima observación (donde $j = 1, 2, \dots, p$).
- β_0 es la intersección.
- β_j es el coeficiente asociado con la j-ésima variable predictora, representando el cambio en Y por cada unidad de cambio en X_j, manteniendo constantes las demás variables predictoras.
- ϵ_i es el error aleatorio para la i-ésima observación.

Al igual que en la regresión univariada, el objetivo es estimar los coeficientes $\beta_0, \beta_1, \dots, \beta_p$ utilizando el método de mínimos cuadrados ordinarios, minimizando la suma de los cuadrados de los residuos.

Consideraciones Importantes: * **Linealidad**: Se asume una relación lineal entre las variables predictoras y la variable de respuesta. * **Independencia de los errores**: Los errores deben ser independientes entre sí. * **Homocedasticidad**: La varianza de los errores debe ser constante para todos los niveles de las variables predictoras. * **Normalidad de los errores**: Los errores deben seguir una distribución normal (esta asunción es más importante para pruebas de hipótesis e intervalos de confianza). * **Multicolinealidad (en regresión múltiple)**: Las variables predictoras no deben estar altamente correlacionadas entre sí, ya que esto puede dificultar la interpretación de los coeficientes individuales.