Modelado Matemático de la Regresión Lineal Múltiple

La **Regresión Lineal Múltiple** es un método estadístico que modela la relación entre una variable dependiente y múltiples variables independientes. La ecuación general del modelo es:

$$\hat{Y} = W_1X_1 + W_2X_2 + \cdots + W_nX_n + b \tag{1}$$

Donde:

- $\hat{Y}$ es la predicción de la variable dependiente (Calificación en IMDb).
- X_i son las variables independientes (Número de votos, Duración, Puntuación en Metacritic, Recaudación bruta).
- W_i son los coeficientes asociados a cada variable independiente.
- b es el término de sesgo o intercepto.

Función de Costo

La función de costo utilizada es el **Error Cuadrático Medio (MSE)**, que se define como:

$$J(W, b) = \frac{1}{2m} \sum_{i=1}^{m} (\hat{Y}^{(i)} - Y^{(i)})^2 \tag{2}$$

Donde:

- m es el número de observaciones.
- $\hat{Y}^{(i)}$ es la predicción para la observación i.
- $Y^{(i)}$ es el valor real de la observación i.

Descenso de Gradiente

Para minimizar la función de costo, se utiliza el algoritmo de **Descenso de Gradiente**. Las actualizaciones de los parámetros se realizan según las siguientes ecuaciones:

$$W_j := W_j - \alpha \frac{\partial J(W, b)}{\partial W_j} \tag{3}$$

$$b := b - \alpha \frac{\partial J(W, b)}{\partial b} \tag{4}$$

Donde:

- α es la tasa de aprendizaje.

- $\frac{\partial J(W,b)}{\partial W_j}$ es la derivada parcial de la función de costo con respecto al peso W_j.
- $\frac{\partial J(W,b)}{\partial b}$ es la derivada parcial de la función de costo con respecto al sesgo b.

Derivadas Parciales

Las derivadas parciales de la función de costo se calculan como sigue:

Para el peso W_j:

$$\frac{\partial J(W,b)}{\partial W_j} = \frac{1}{m} \sum_{i=1}^{m} (\hat{Y}^{(i)} - Y^{(i)}) X_j^{(i)} \tag{5}$$

Para el sesgo b:

$$\frac{\partial J(W,b)}{\partial b} = \frac{1}{m} \sum_{i=1}^{m} (\hat{Y}^{(i)} - Y^{(i)}) \tag{6}$$