



# Regresión Lineal

## Evaluación



# Regresión (lineal): Métricas

Mean Absolute Error (MAE). Errores en unidades del target

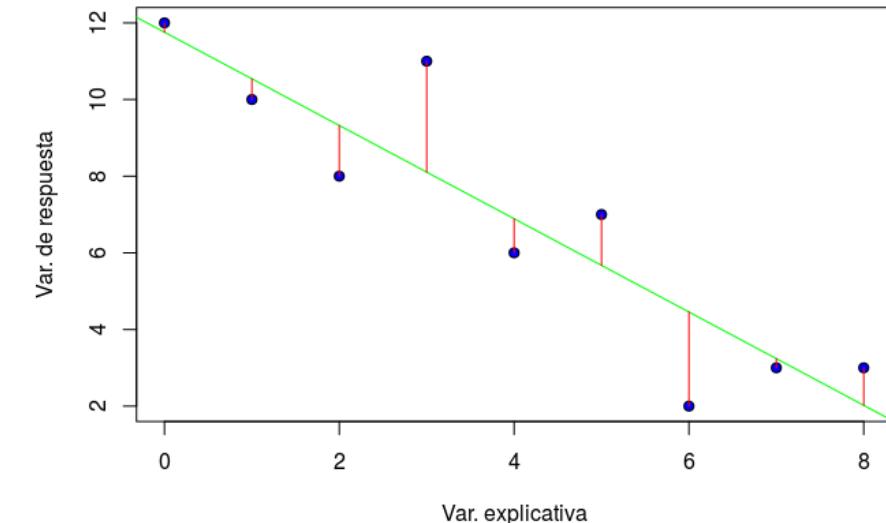
$$MAE = \frac{1}{n} \sum_{n=1}^N |y_i - \hat{y}_i|$$

Mean Squared Error (MSE). No se interpreta bien al ir al cuadrado, pero enfatiza mucho más los errores altos

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2.$$

Root Mean Squared Error (RMSE)

$$RMSD = \sqrt{\frac{\sum_{t=1}^T (\hat{y}_t - y_t)^2}{T}}.$$

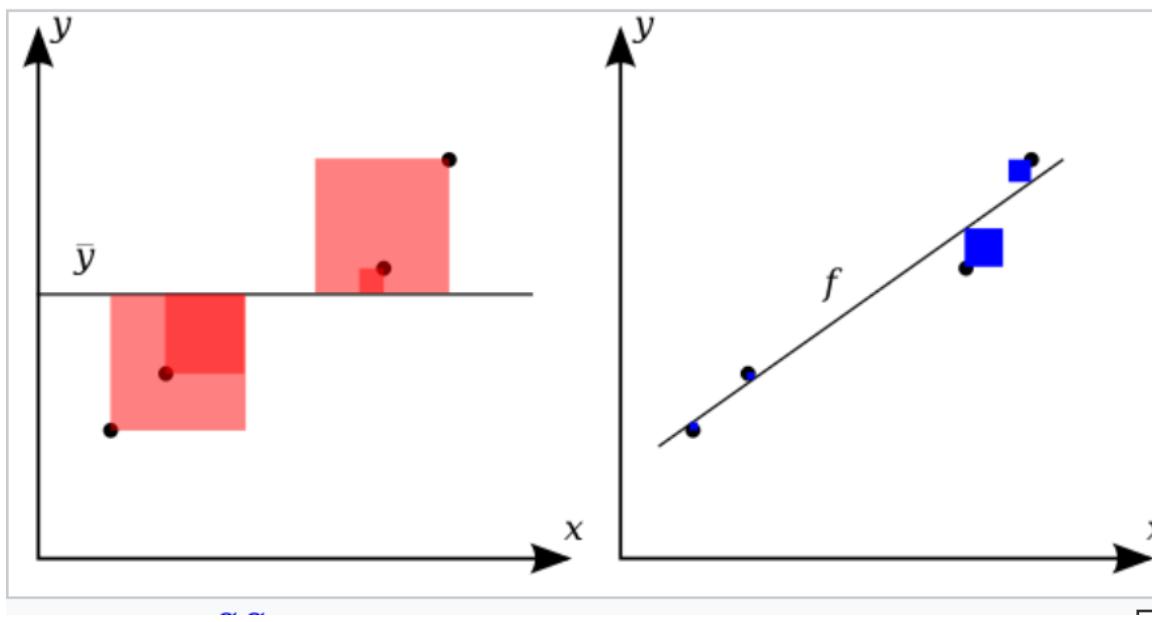


**Mean Absolute Percentage Error (MAPE)**

$$MAPE = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

## R-Squared

Coeficiente de determinación. Mide cuánto de bien una regresión se ajusta a los datos. También se define como la porción de variación de la variable dependiente ( $y$ ) predecible mediante la independiente ( $x$ ). Va de  $[0, 1]$ . Cuanto mejor se ajuste, más se acercará a 1. Cuanto más cercano a 0, menos fiable será.



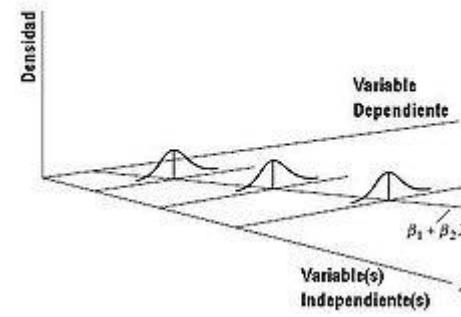
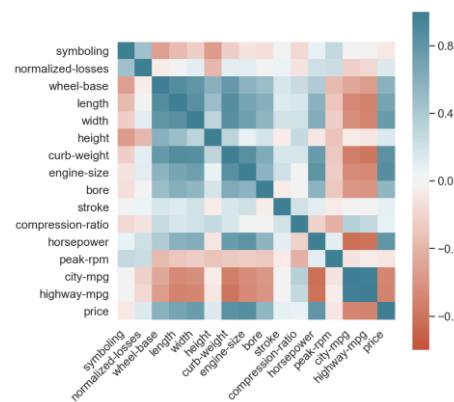
$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$

$$SS_{\text{res}} = \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

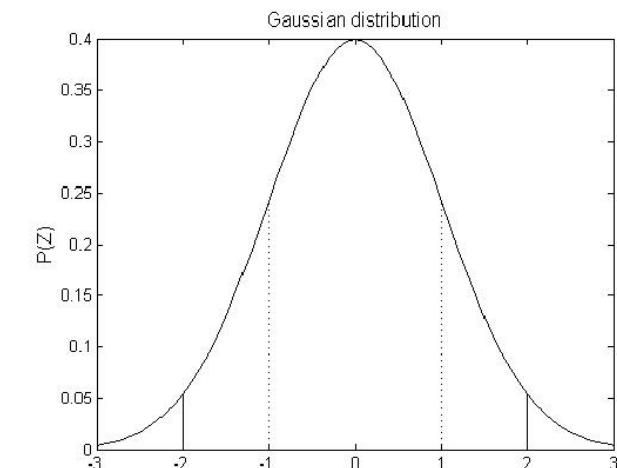
$$SS_{\text{tot}} = \sum_i (y_i - \bar{y})^2$$

# Condiciones

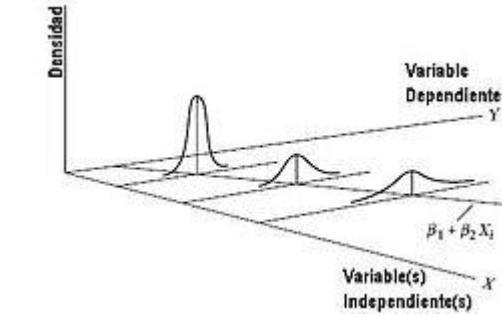
1. Distribución normal del target
2. No colinealidad o multicolinealidad. Correlación entre los predictores. Lo solucionamos eliminando uno de ellos. Hay que garantizar la independencia entre todos ellos.
3. Relación lineal entre target y predictores. Matriz de correlación
4. Homocedasticidad. Varianza constante de los errores a lo largo de las observaciones.



Distribución Homocedástica



Gaussian distribution



Distribución Heterocedástica



# Regresión lineal: resumen

Objetivo: encontrar la relación lineal entre todas las variables del problema. Encontrar 'a' y 'b'.

El valor añadido es poder predecir valores inexistentes.

Tiene ciertas limitaciones. Un ejemplo, datos no lineales.

Se genera un error global que es la distancia entre todos los datos y nuestro modelo (línea, plano, hiperplano).

