

Machine Learning para Inteligencia Artificial

Resumir una distribución

Universidad ORT Uruguay

26 de Marzo, 2025

Media/esperanza/promedio

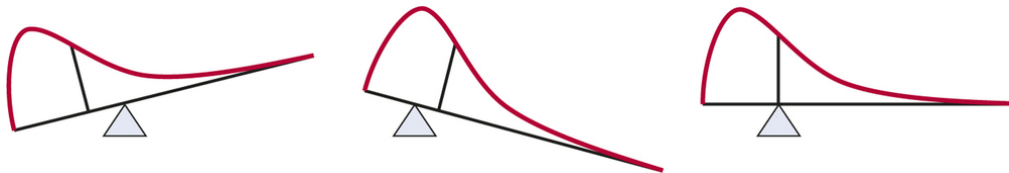
Empírico (datos $T = \{z_1, \dots, z_N\}$)

$$\hat{\mu} = \text{Promedio}_{z \sim T}[z] = \frac{1}{N} \sum_{i=1}^N z_i$$

Teórico (distribución \mathcal{D})

$$\mu = \text{Promedio}_{z \sim \mathcal{D}}[z] = \lim_{|T| \rightarrow \infty} \text{Promedio}_{z \sim T}[z]$$

Interpretación: centro de masa o punto de equilibrio



Varianza y desvío estándar

Empírico (datos $T = \{z_1, \dots, z_N\}$)

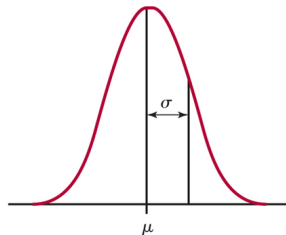
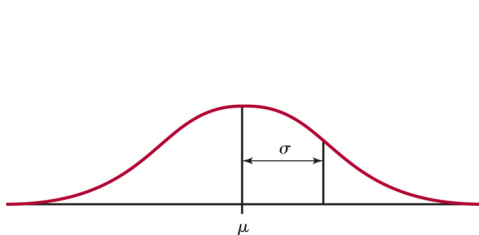
$$\hat{\sigma}^2 = V_{z \sim T}[z] = \frac{1}{N} \sum_{i=1}^N (z_i - \hat{\mu})^2$$

Teórico (distribución \mathcal{D})

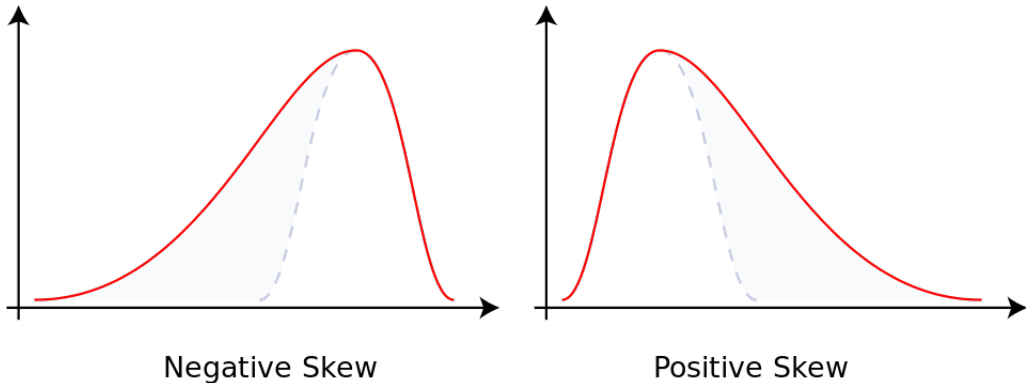
$$\sigma^2 = V_{z \sim \mathcal{D}}[z]$$

Interpretación: dispersión entorno a la media

Desvío estándar: es la raíz de la varianza



Característica a observar: simetría



Fuente: Wikipedia

Mediana

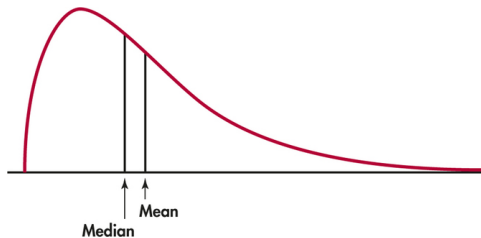
Empírico (datos $T = \{z_1, \dots, z_N\}$)

$$\hat{m} = z_{\lfloor \frac{N}{2} \rfloor}^*$$

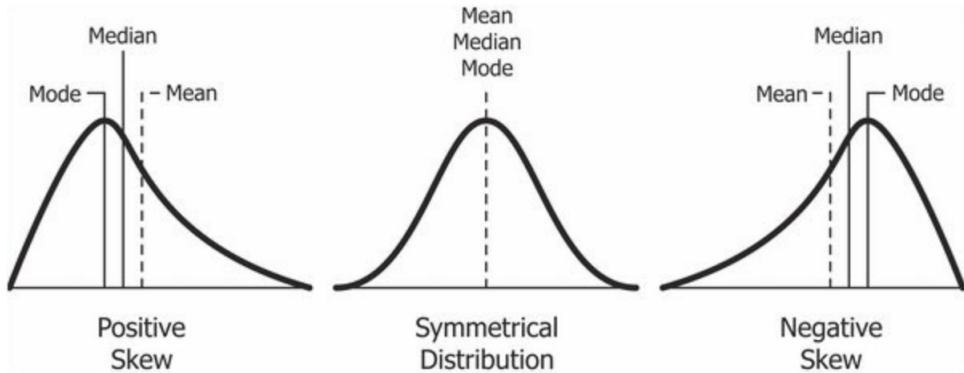
Teórico (distribución \mathcal{D})

$$m \text{ es tal que } \text{Prob}_{\mathcal{D}}[z \leq m] = \frac{1}{2}$$

Interpretación: es el valor que divide en dos partes iguales la distribución

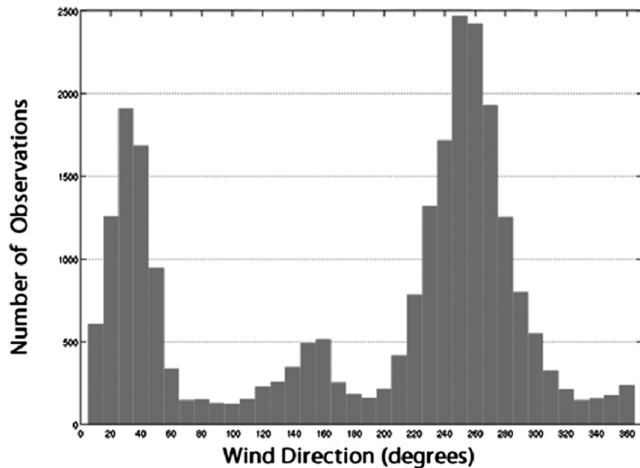


Simetría: media vs mediana



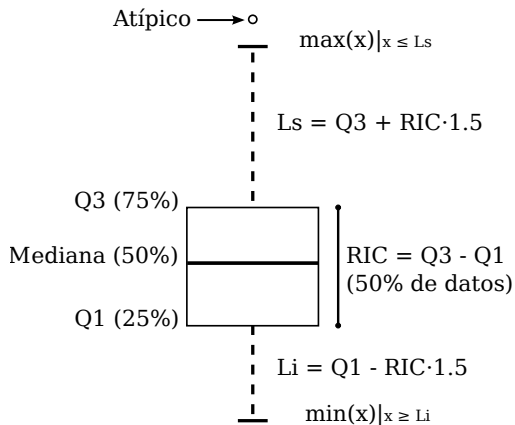
Fuente: Wikipedia

Características a observar: multimodalidad

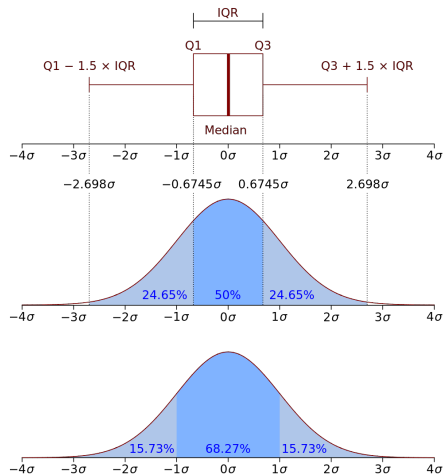


Fuente: Climatology of High Wind Events in the Owens Valley, California

Resumen visual-numérico: el boxplot



Boxplot: comparación con la normal



- En la distribución normal $N(\mu, \sigma^2)$, el 99.73 % del área debajo la campana está a menos de 3σ de distancia de μ .
- El IQR es en este caso 1.349σ .
- Los brazos del boxplot contienen el 99.3 % del área.
- El valor 0.6745σ se conoce históricamente como *error probable*.

Fuente: Wikipedia