

Curso Aprendizaje Automático (ML) con Python

Nelson López Centeno

Nociones de Estadística

- ☐ Media
- ☐ Mediana
- ☐ Moda
- ☐ Varianza
- ☐ Desviación estándar
- ☐ Distribución gaussiana (normal)
- ☐ Correlación
- ☐ Covarianza



Demos / Ejercicios

- ☐ Estadísticas con Python
 - ☐ NumPy
 - ☐ Pandas
- ☐ Visualizaciones estadísticas con Python
 - ☐ Histograma
 - ☐ Cajas y bigotes (boxplot)
 - ☐ Dispersión
 - ☐ Mapa de calor

Media aritmética

Valor medio de una distribución.

$$media(x) = \frac{1}{n} \sum_{i=1}^n x_i$$

Suma de las observaciones dividida por el número total de observaciones.

Es sensible a los valores atípicos y a valores faltantes.

- Salario medio anual de España es 21.519 euros
- Una familia en mi ciudad tiene de media 0,6 hijos

Mediana

Es el valor central, donde hay una cantidad igual de valores menores y mayores que él.

1, 2, 3, 10, 30, 40, 50

1, 2, 3, 10, 20, 30, 40, 50

15 (media de 10 y 20)

No es sensible a los valores atípicos ni a los valores faltantes.
Describe bien los valores típicos, pero falla al detectar los atípicos.

Moda

Es el valor que más se repite.

Suele estar cerca del centro de una distribución.

- Unimodal
- Bimodal
- Multimodal

No es sensible a valores atípicos ni a valores faltantes.

Varianza

Indica cuanto difieren en promedio las observaciones del valor medio.

$$\text{varianza}(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \text{media}(x))^2$$

Diferencia media al cuadrado de cada observación con respecto a la media

Varianza baja: Valores están agrupados en torno a la media.

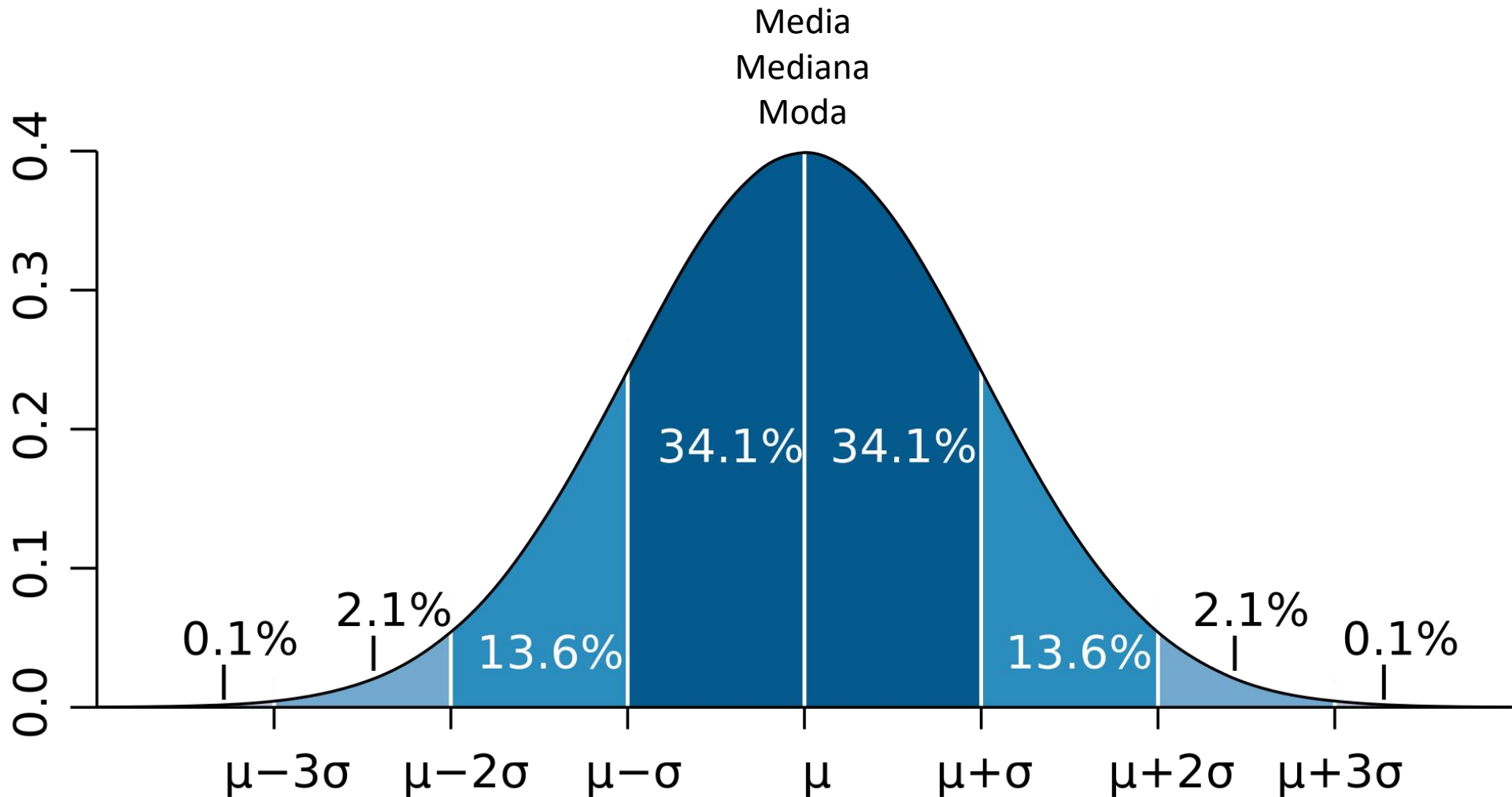
Varianza alta: Valores dispersos con relación a la media.

Desviación estándar

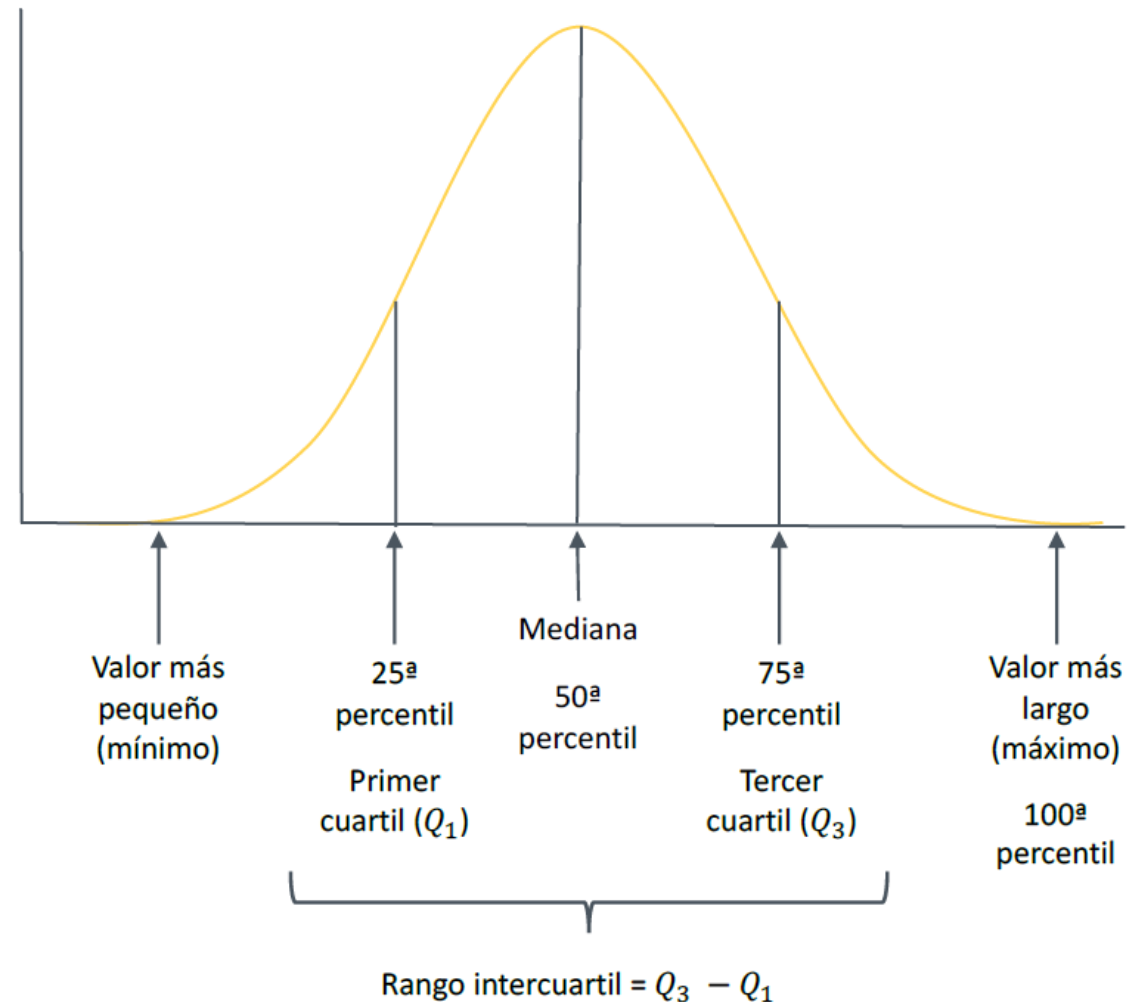
Raíz cuadrada de la varianza.

Es más fácil de comparar con los valores originales.

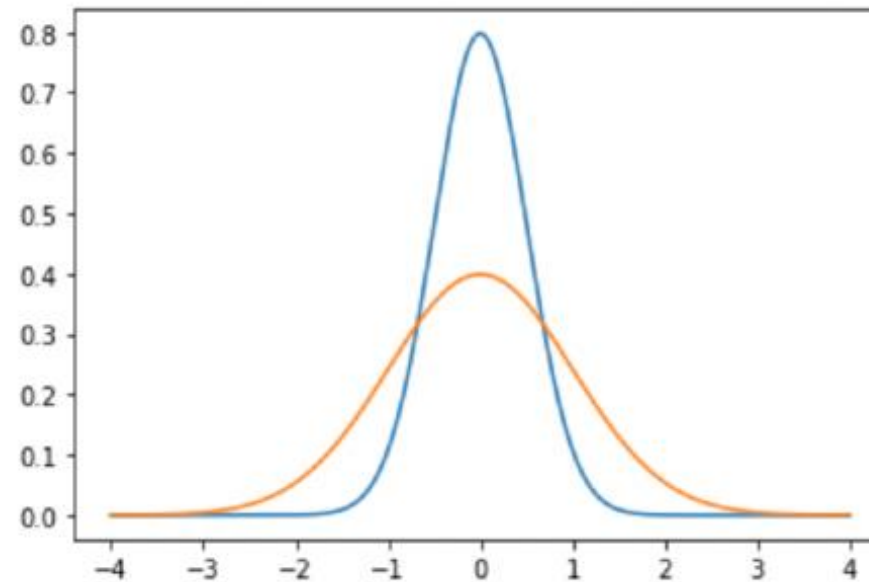
Distribución gaussiana (normal)



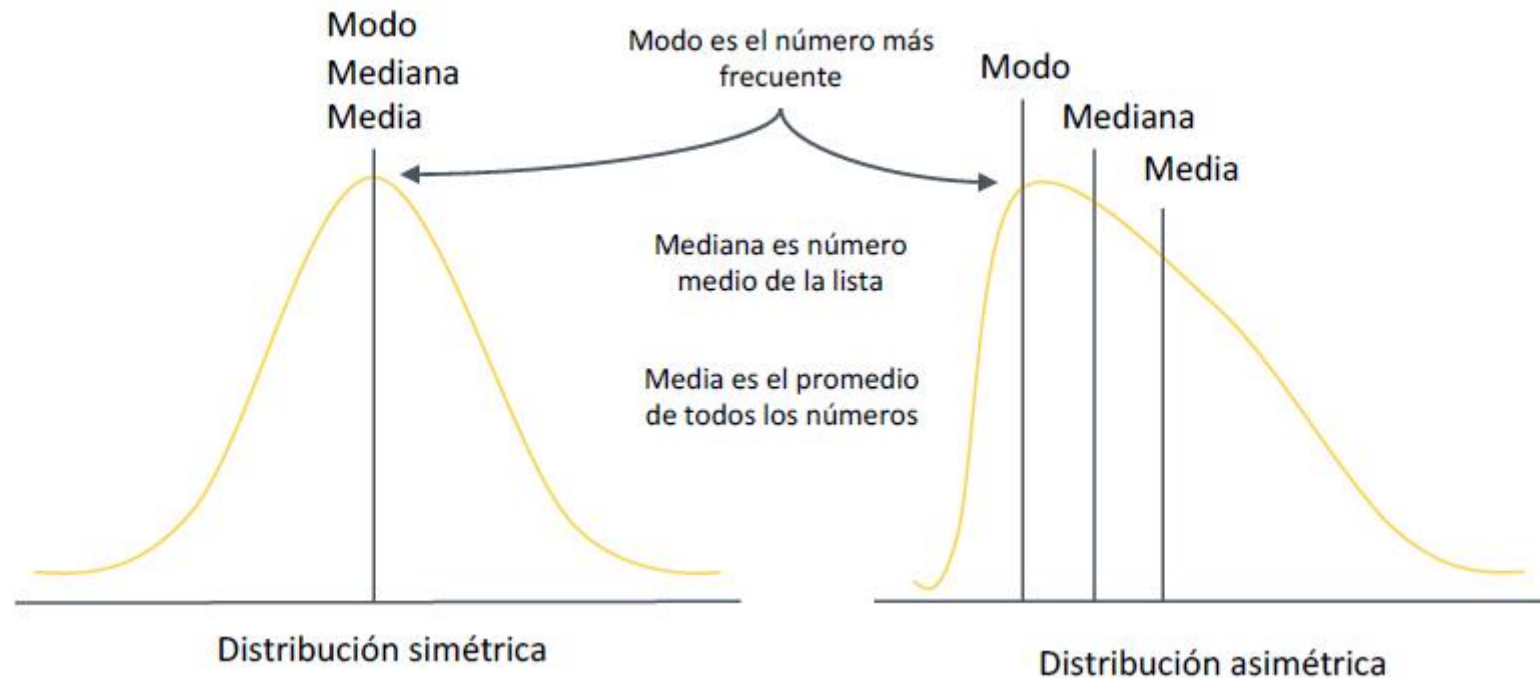
Distribución gaussiana (normal)



Distribución gaussiana (normal)



Distribución gaussiana (normal)

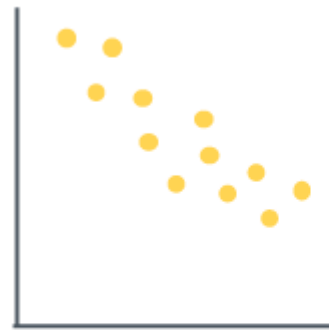


Correlación

Mide el grado en que dos valores se mueven en relación con el otro.

Coeficiente de correlación: -1.0 a 1.0

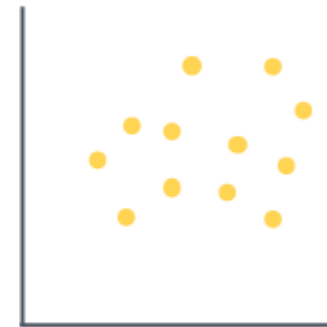
- 1 Si un valor se mueve el otro se mueve al mismo tiempo y en la misma dirección.
- -1 Los valores se mueven en direcciones opuestas.
- 0 No hay relación en el cambio de las variables



Correlación
negativa



Correlación
positiva



Correlación
neutra

Covarianza

Mide la fuerza de la correlación entre dos variables.

$$cov(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - media(x)) (y_i - media(y))$$

Covarianza positiva: Los dos valores se mueven juntos

Covarianza negativa: Los dos valores se mueven a la inversa



Ejercicios

- ☐ Cálculos estadísticos con NumPy
- ☐ Cálculos estadísticos con Pandas
- ☐ Gráficos estadísticos con Seaborn

