

Laboratorio 1: La maldición de la dimensionalidad

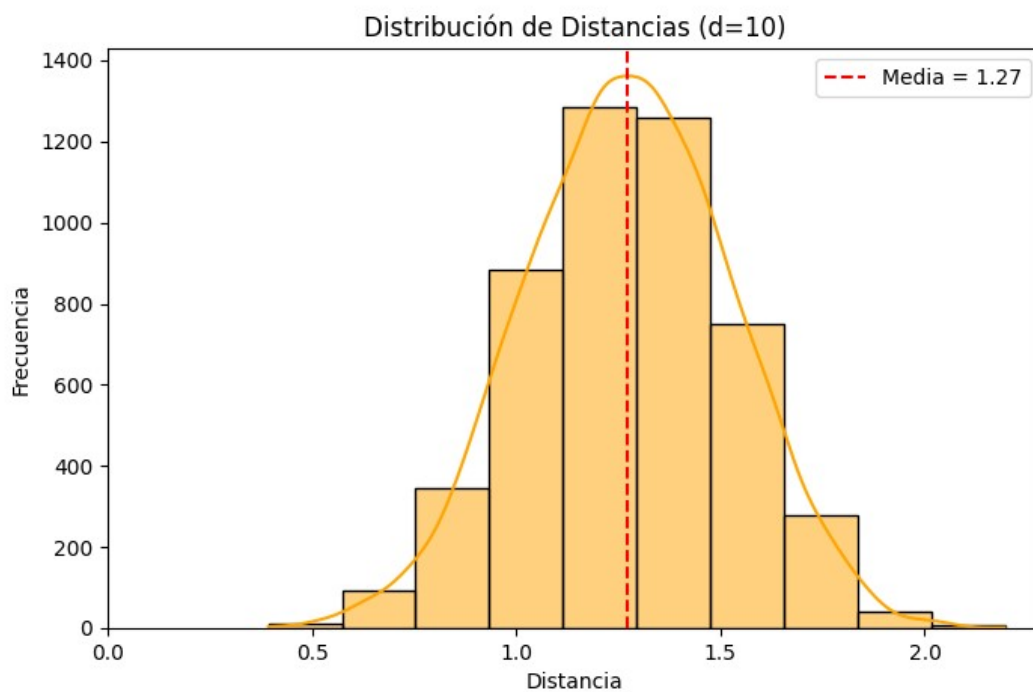
Alumno: Yanqui Vera Henry Aron

hyanquiv@unsa.edu.pe

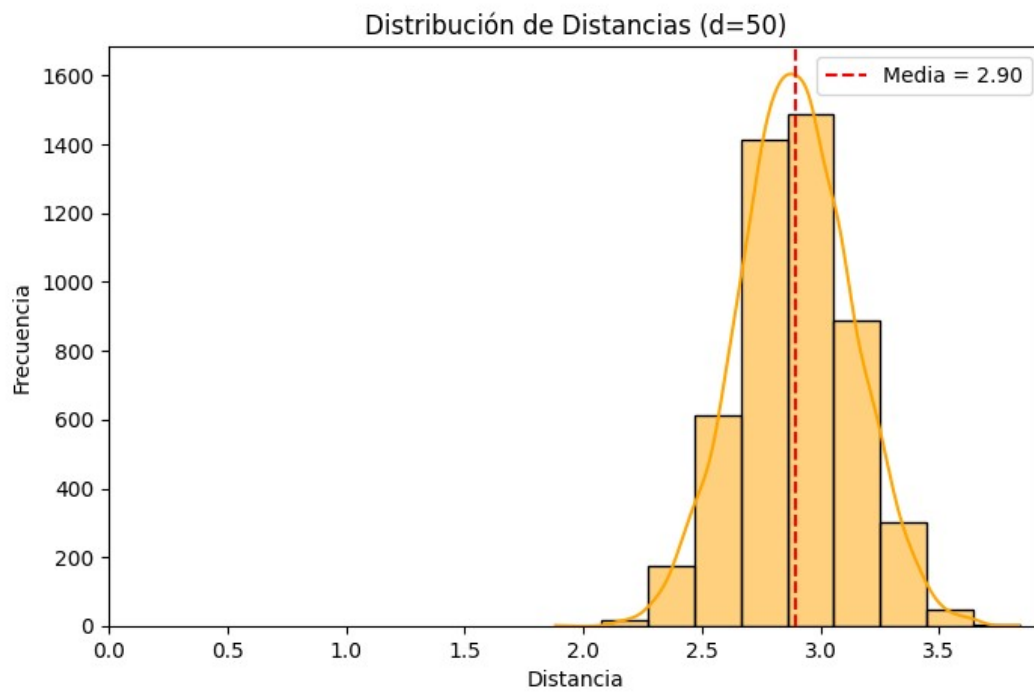
Repositorio: [hyanquiv/EDA](#)

Informe:

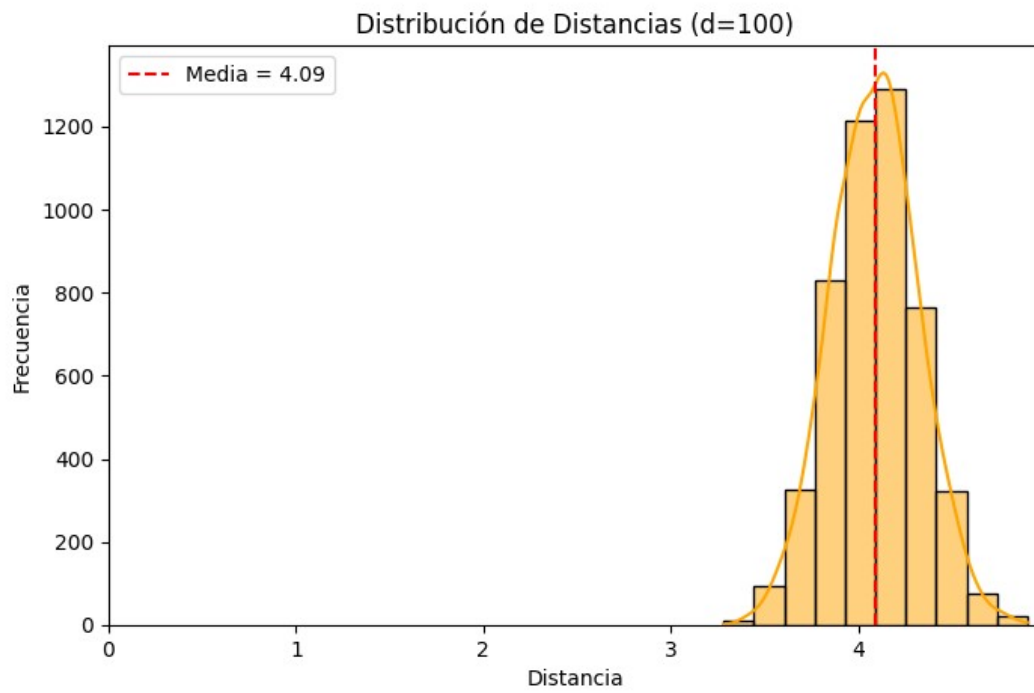
- Dimensión 10:



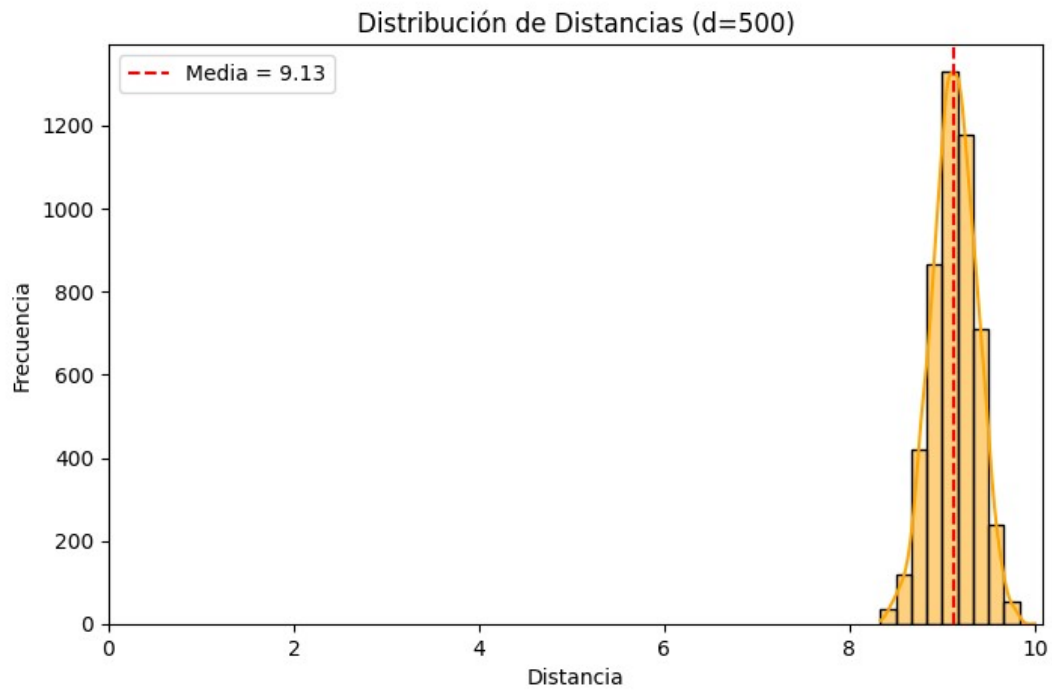
- Dimensión 50:



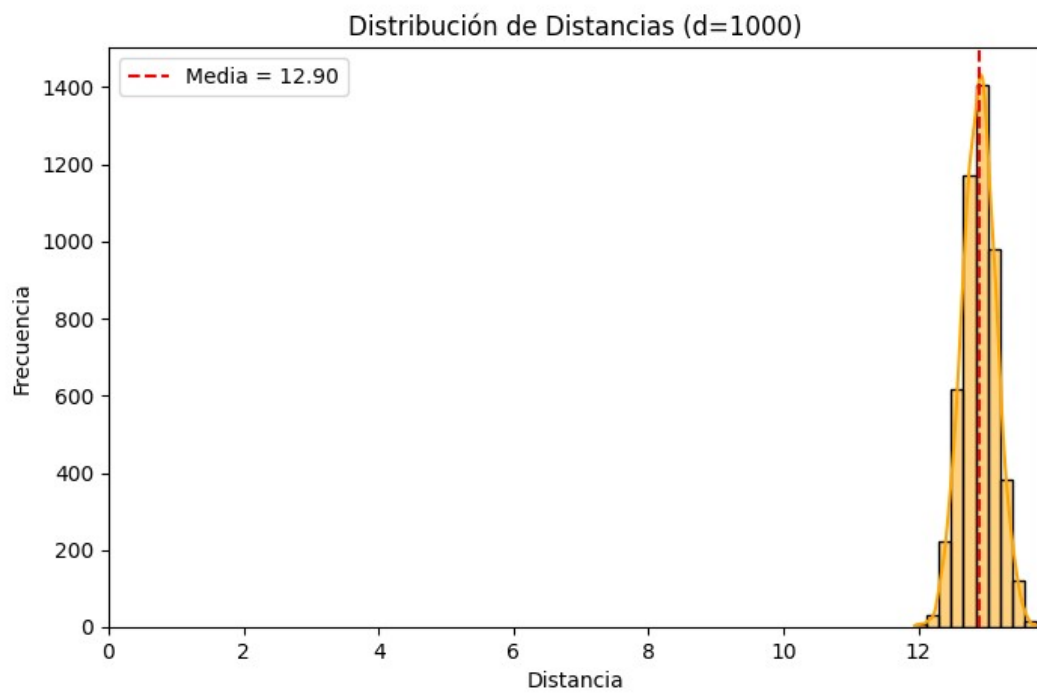
- Dimensión 100:



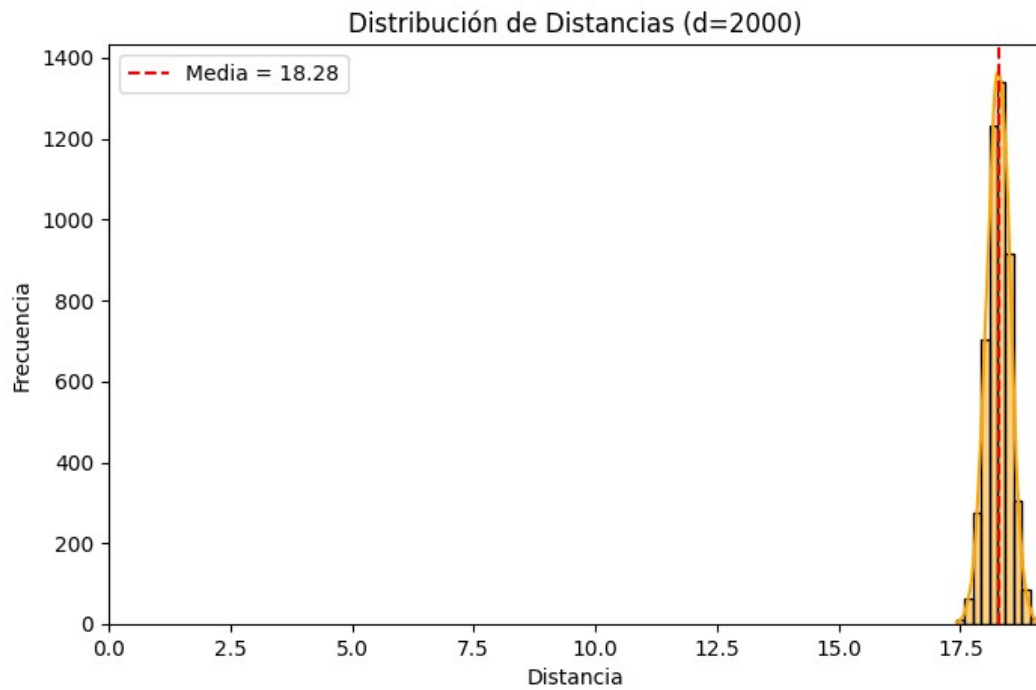
- Dimensión 500:



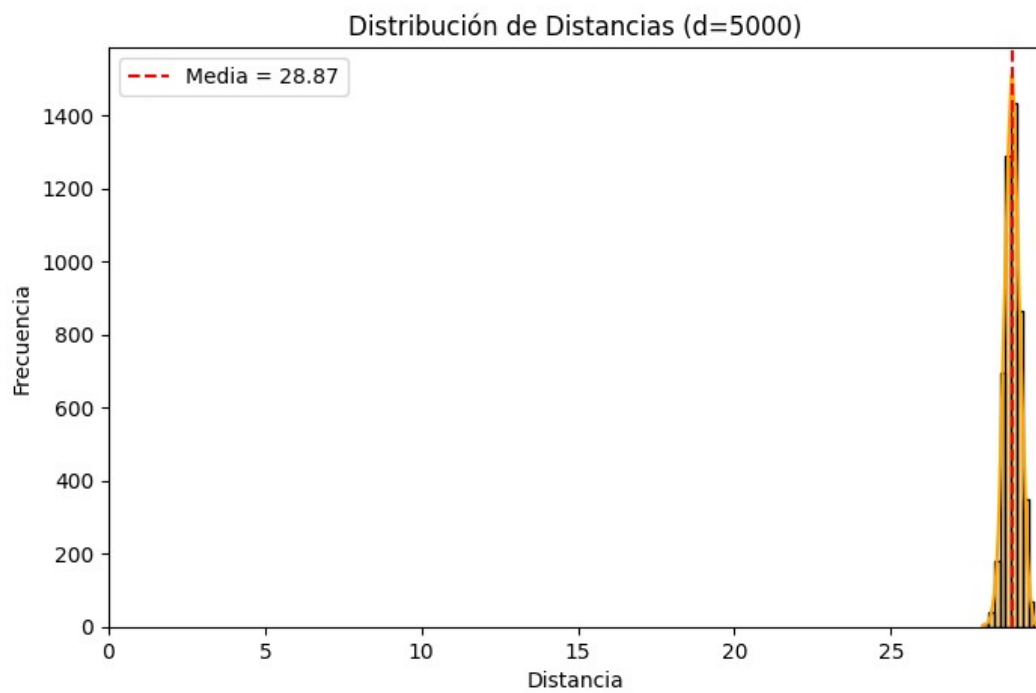
- Dimensión 1000:



- Dimensión 2000:



- Dimensión 5000:



Conclusiones:

Podemos observar que a medida que la dimensión de los conjuntos de datos aumenta, el número de cálculos necesarios para hallar las distancias crece significativamente, más no el total de distancias que siempre serán 4950 pues el conjunto de puntos que se utilizan para los experimentos es siempre 100. Esto puede tener un impacto en la complejidad computacional y el tiempo requerido para realizar los cálculos.

Es probable que las distribuciones de distancias en conjuntos de datos de mayor dimensión sean más dispersas. Esto se debe a que, en dimensiones más altas, los puntos tienden a estar más alejados entre sí basándonos en la distancia euclidiana, esto podemos apreciarlo a simple vista en los gráficos generados pues la media que se utiliza comúnmente como medida estadística va aumentando mientras se aumenta la dimensión, lo cual nos da a entender que la distancia en general aumenta.

Bibliografía:

- "Estadística Matemática con Aplicaciones" de Dennis Wackerly, William Mendenhall y Richard L. Scheaffer
- Wikipedia - Distancia Euclidiana :
https://es.wikipedia.org/wiki/Distancia_euclidiana