

Universidad Nacional de San Agustín

Facultad de Producción y Servicios

Escuela Profesional de Ciencia de la Computación



CURSO ESTRUCTURA DE DATOS AVANZADOS

LABORATORIO 01 - LA MALDICIÓN DE LA DIMENSIONALIDAD

Alumno: Kevin Joaquin Chambi Tapia

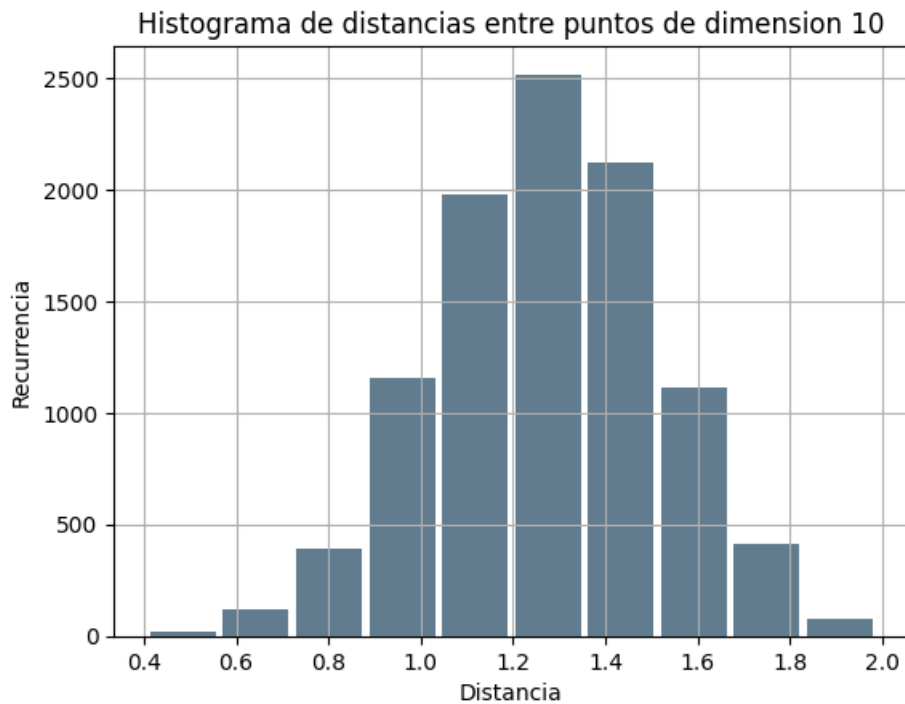
Arequipa – Perú

2023

En este informe, se analiza cómo el espacio cambia a medida que la dimensionalidad de los datos aumenta.

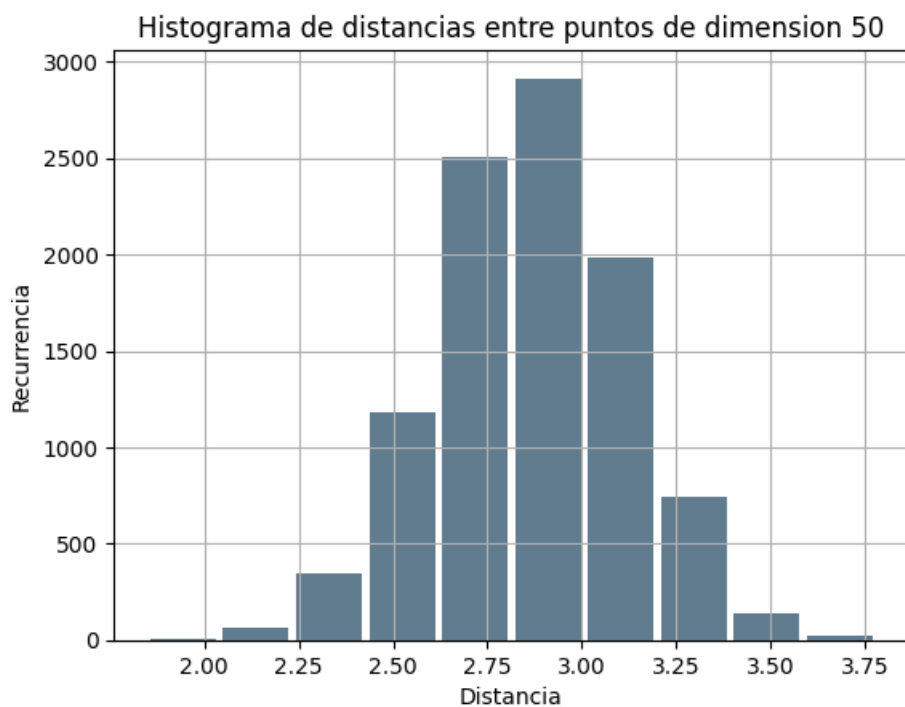
Dimensión 10

El valor mínimo es 0.406232, el valor máximo es 1.9869, y el valor promedio es 1.274298. La desviación estándar es 0.241397.



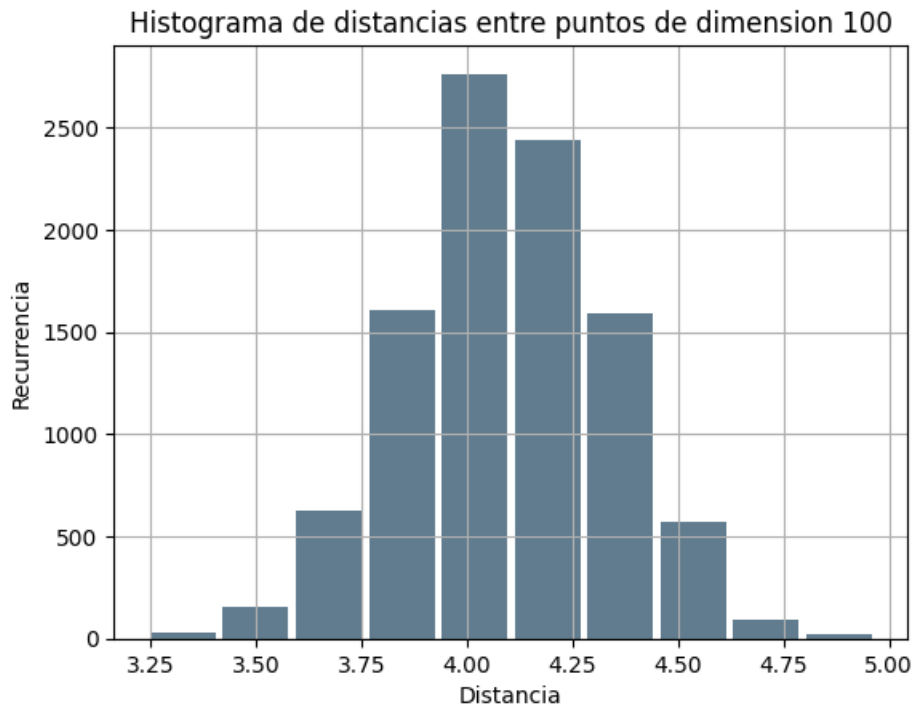
Dimensión 50

El valor mínimo es 1.8473, el valor máximo es 3.7793, y el valor promedio es 2.8666. La desviación estándar es 0.2503.



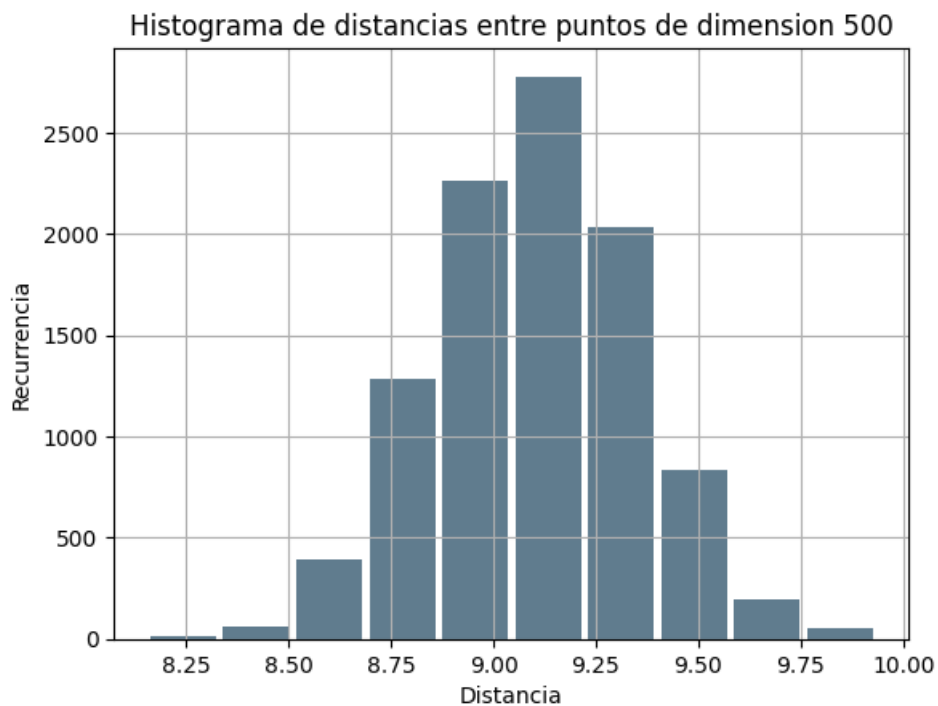
Dimensión 100

El valor mínimo es 3.2394, el valor máximo es 4.9680, y el valor promedio es 4.0932. La desviación estándar es 0.2412.



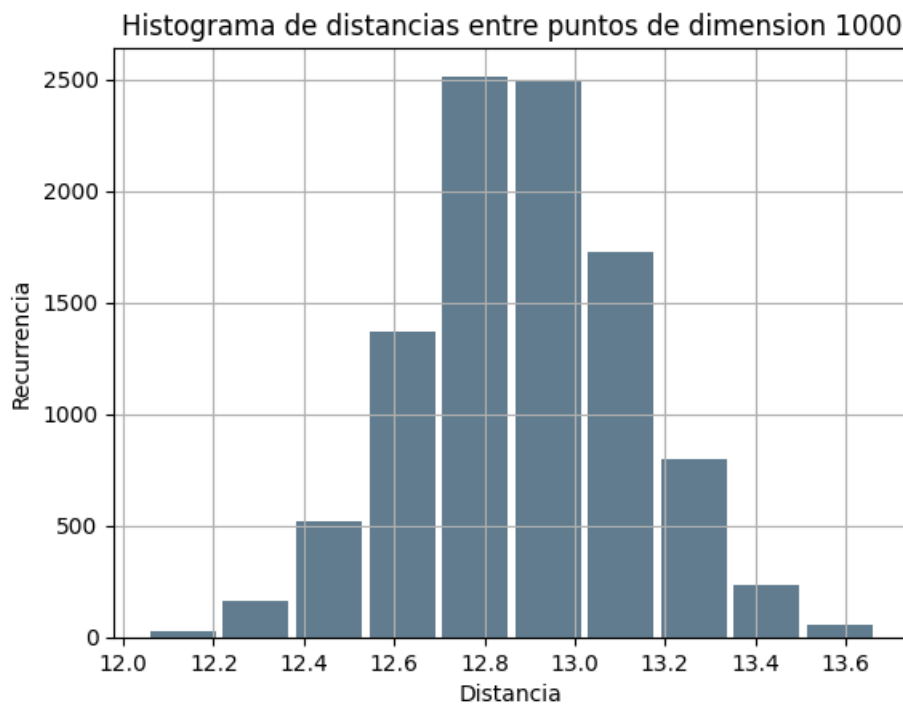
Dimensión 500

El valor mínimo es 8.1547, el valor máximo es 9.9343, y el valor promedio es 9.1003. La desviación estándar es 0.2459.



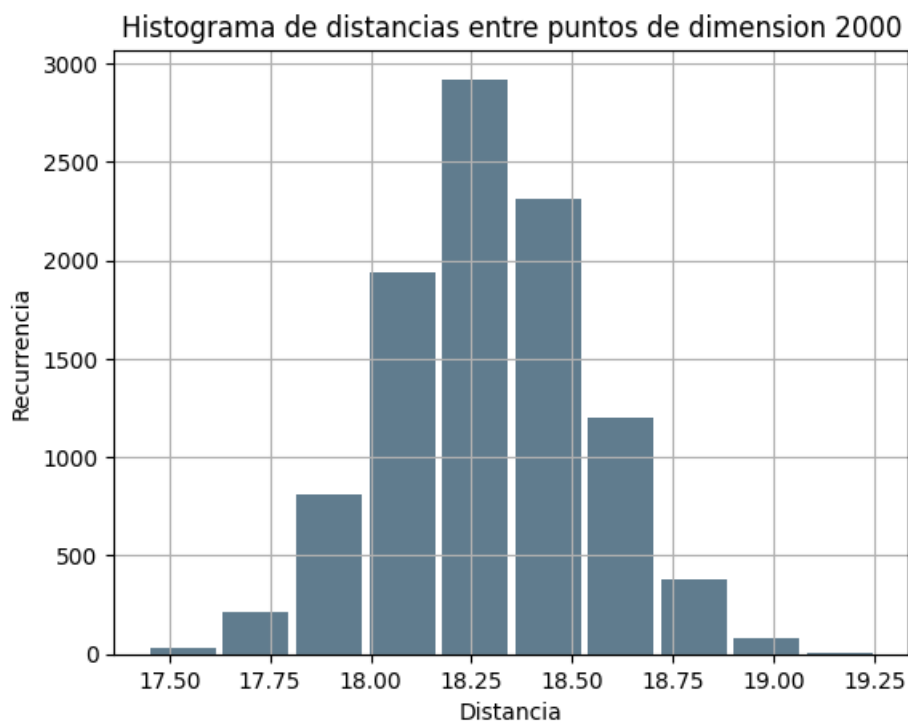
Dimensión 1000

El valor mínimo es 12.0522, el valor máximo es 13.6676, y el valor promedio es 12.8860. La desviación estándar es 0.2411.



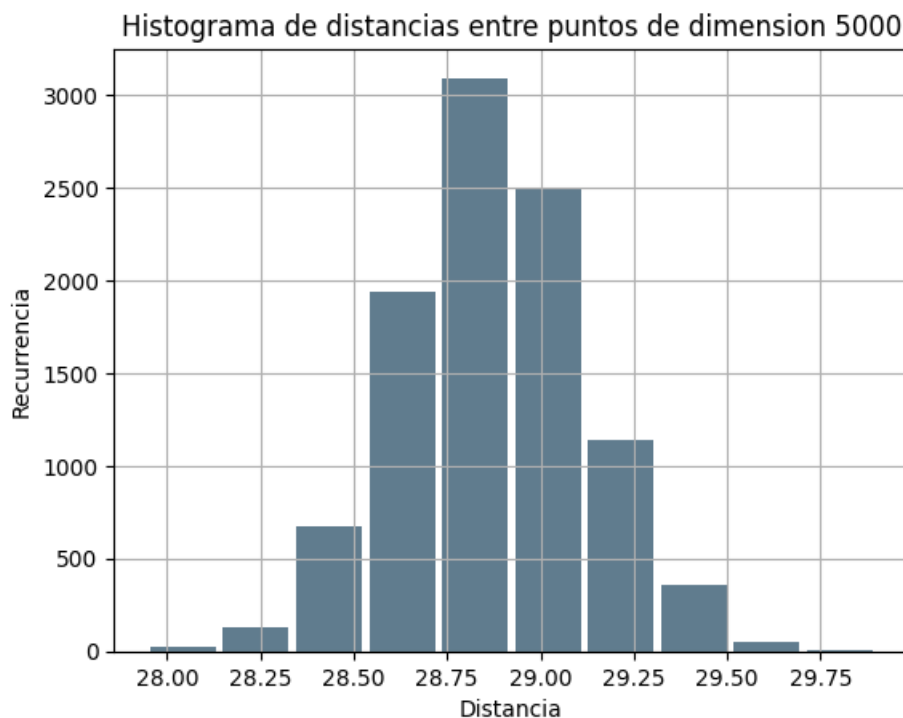
Dimensión 2000

El valor mínimo es 17.4441, el valor máximo es 19.2539, y el valor promedio es 18.2938. La desviación estándar es 0.2479.



Dimensión 5000

El valor mínimo es 27.946, el valor máximo es 29.8987, y el valor promedio es 28.8705. La desviación estándar es 0.2477.



Conclusión

La distancia euclidiana entre los puntos también aumenta a medida que la dimensionalidad aumenta. Esto se debe a que la distancia euclidiana es una función no lineal de la dimensionalidad. Como resultado, la distancia entre dos puntos puede aumentar significativamente a medida que se agregan más dimensiones.

La combinación de estos factores hace que sea más desafiante clasificar los datos de alta dimensionalidad. Esto se debe a que los algoritmos de clasificación se basan en la distancia entre los puntos para tomar decisiones. Como la distancia entre los puntos aumenta a medida que la dimensionalidad aumenta, los algoritmos de clasificación tienen más dificultades para distinguir entre los puntos.

Código:

<https://github.com/croko22/CS-UNSA/tree/main/4%20EDA/LAB%20-%20%201%20-%20Dimensionalidad>