



Universidad Católica
San Pablo

Ciencia de la Computación

Estructuras de Datos Avanzadas

Docente Rosa Yuliana Gabriela Paccotacya Yanque

La maldición de la dimensionalidad

Entregado el 31/08/2023

Camila Andrea Luque Juarez

Semestre VI

2023-2

"La alumna declara haber realizado el presente trabajo de acuerdo a las normas de la Universidad Católica San Pablo"

LA MALDICIÓN DE LA DIMENSIONALIDAD

1. Objetivo

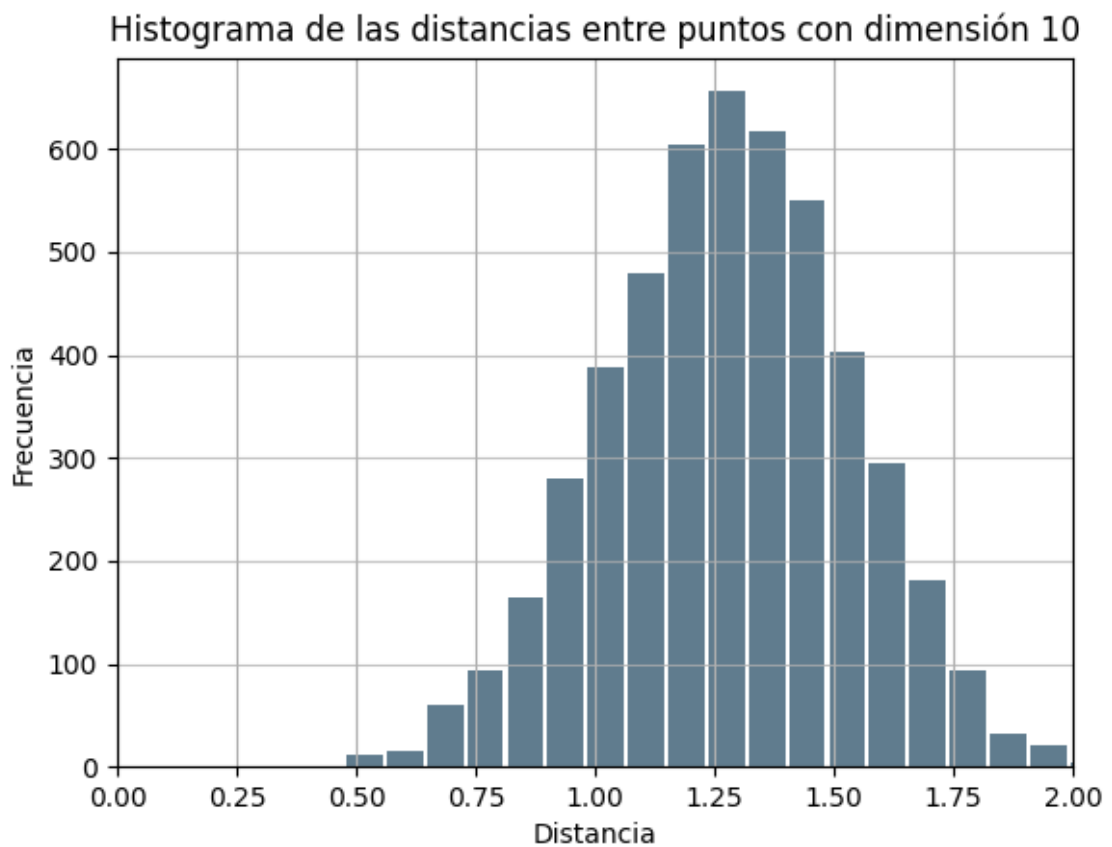
En el presente informe se analiza cómo el espacio cambia a medida que la dimensionalidad de los datos aumenta.

Para este caso en específico, las dimensiones a evaluar serán 10, 50, 100, 500, 1000, 2000 y 5000.

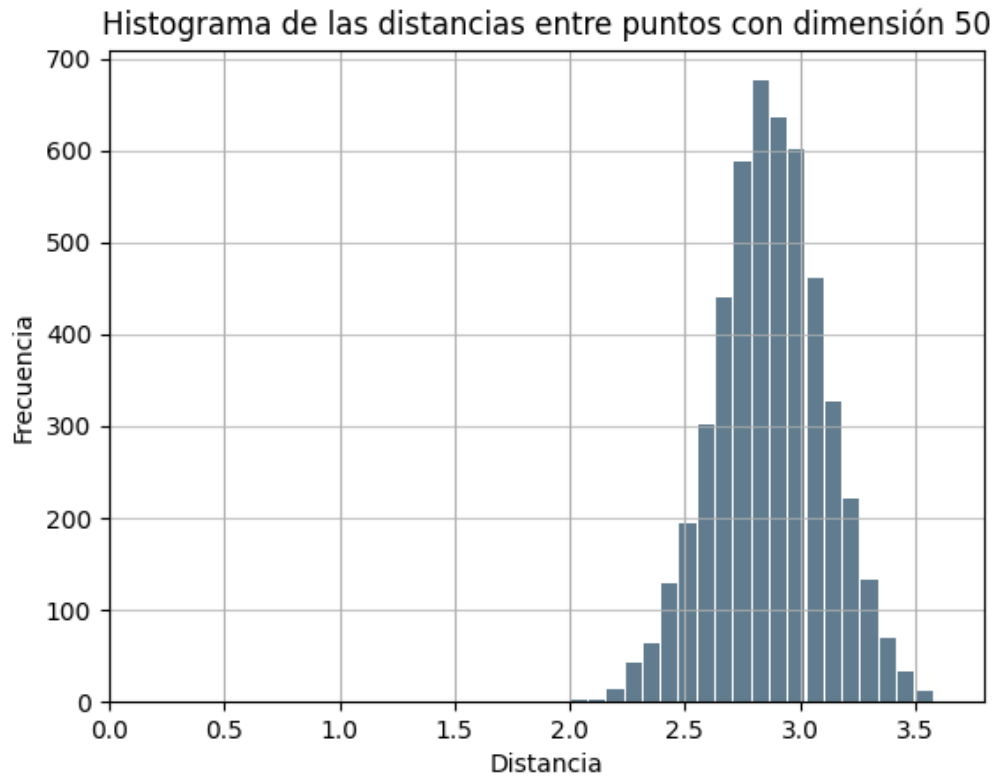
2. Gráfica de resultados en histogramas

a. Histograma con 10 dimensiones

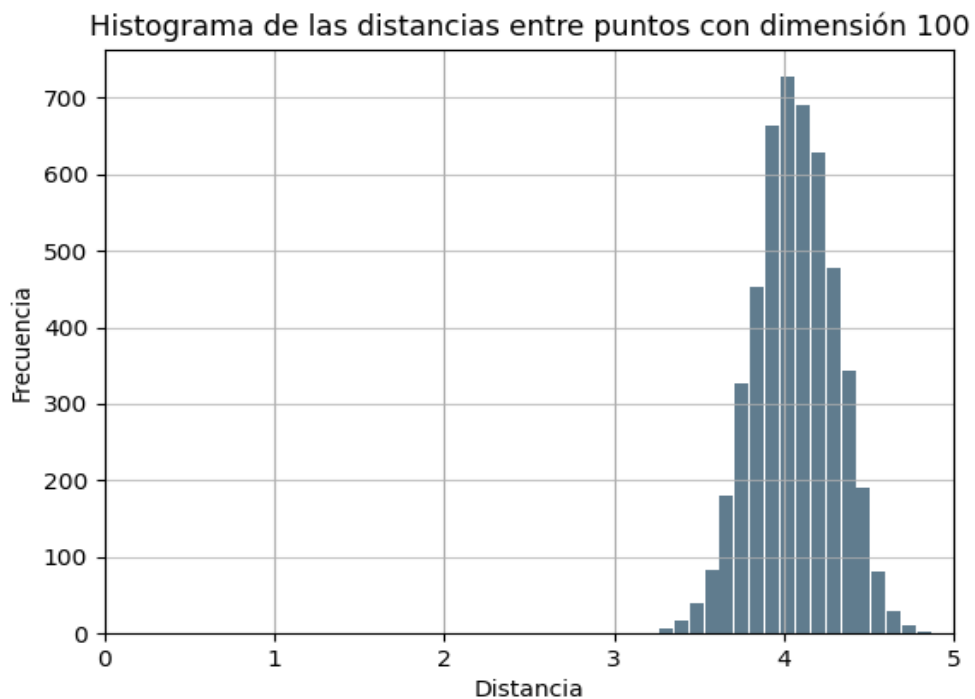
Rango (0 - 2)



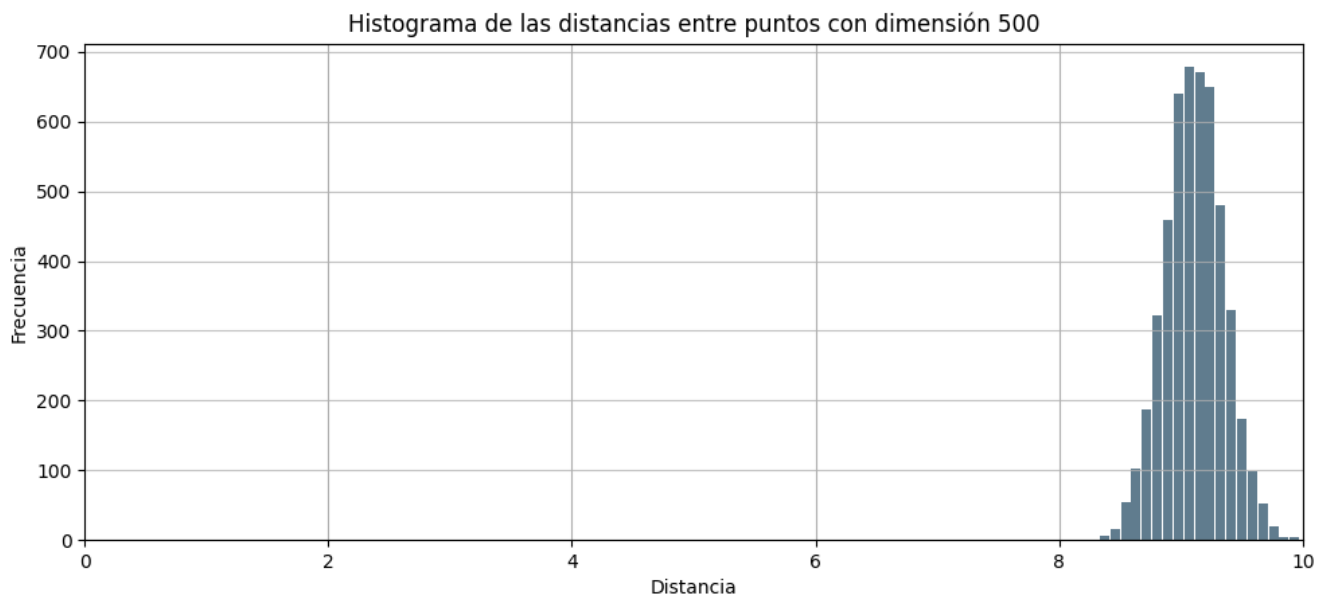
b. Histograma con 50 dimensiones
Rango (0 - 3.8)



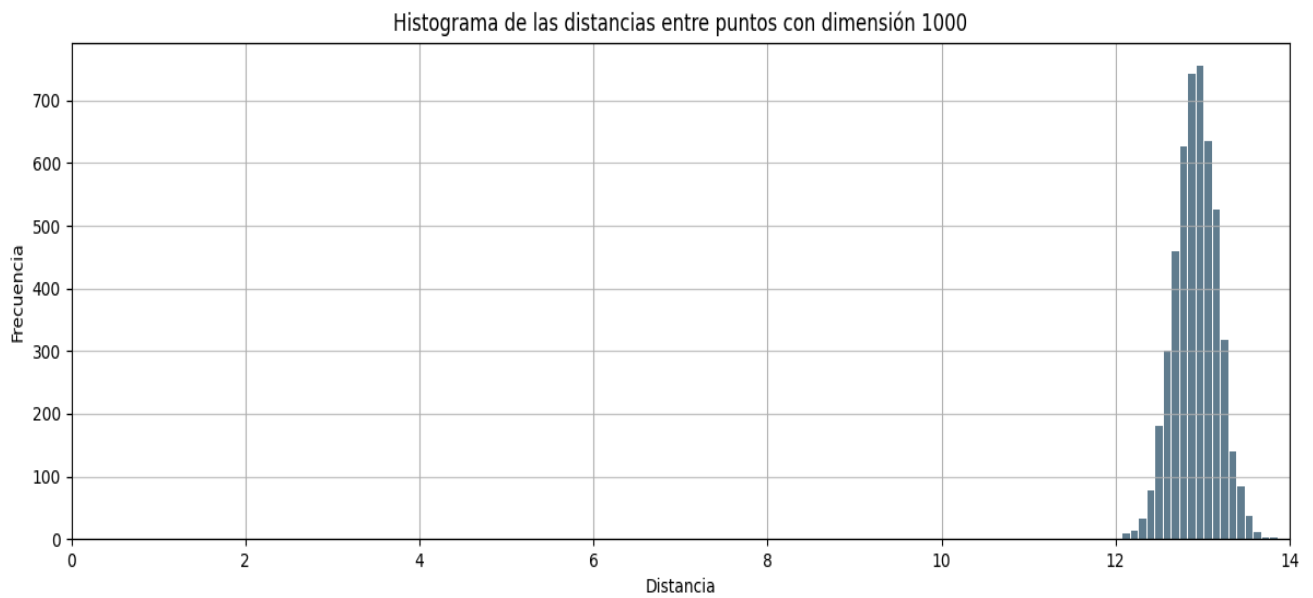
c. Histograma con 100 dimensiones
Rango (0 - 5)



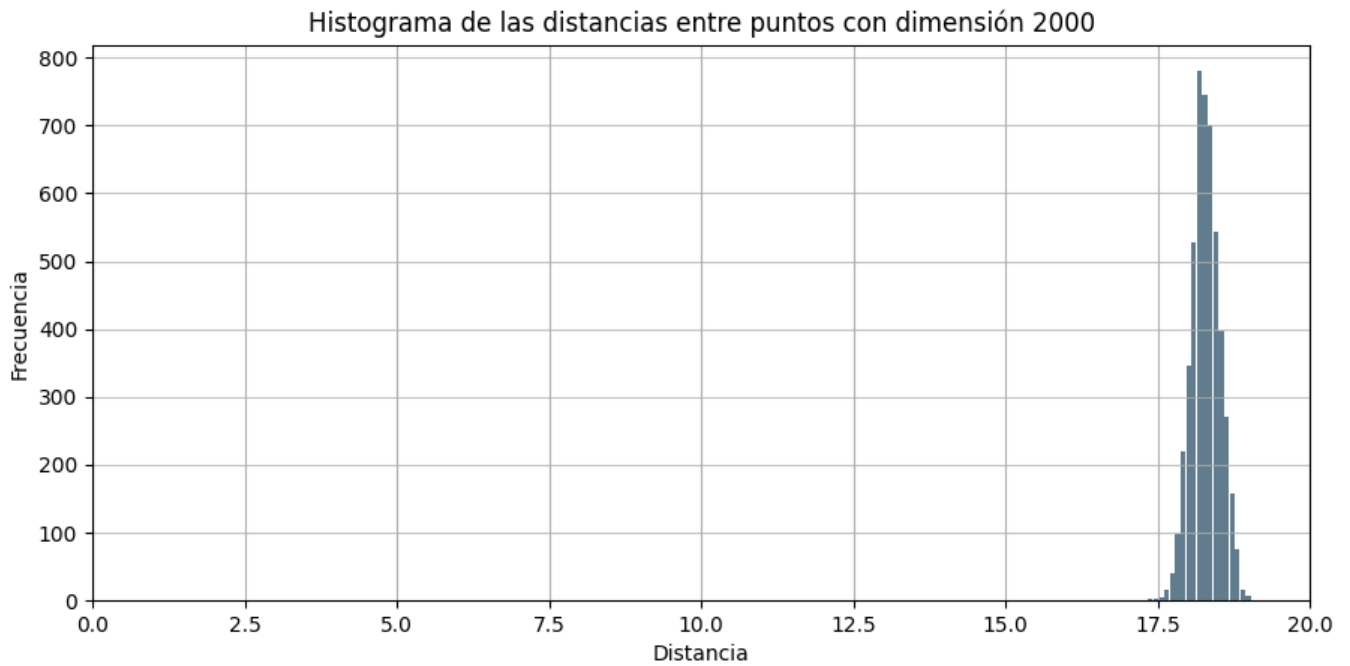
d. Histograma con 500 dimensiones
Rango (0 - 10)



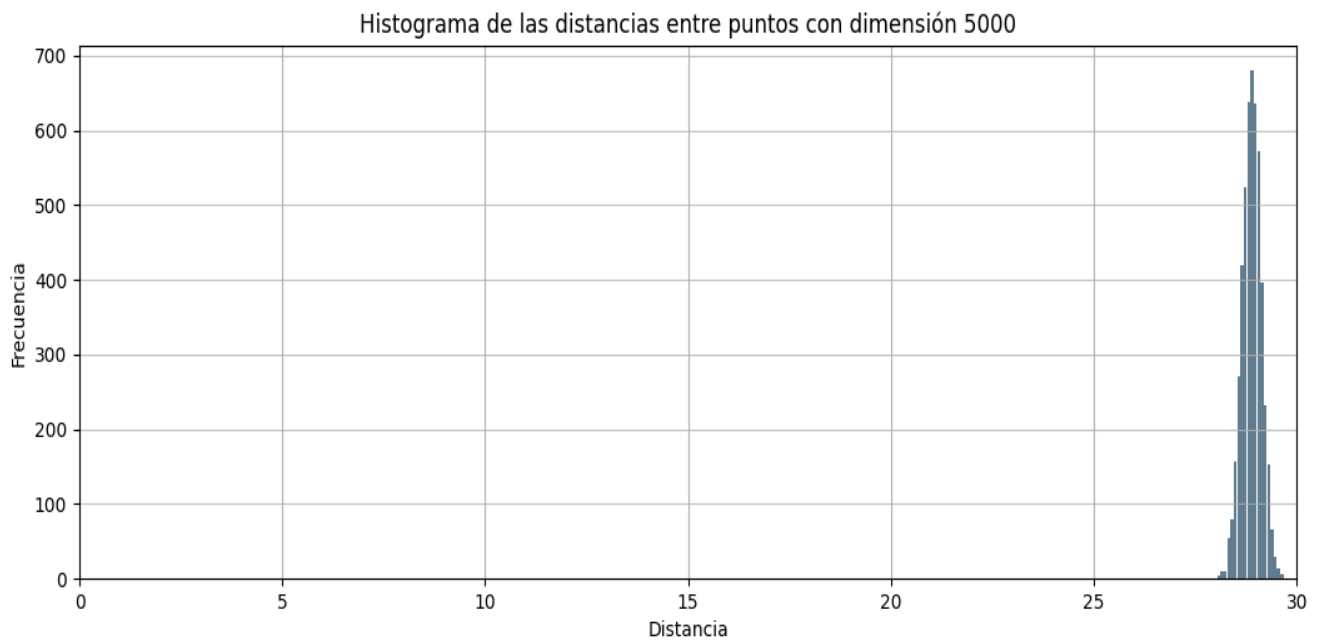
e. Histograma con 1000 dimensiones
Rango (0 - 14)



f. Histograma con 2000 dimensiones
Rango (0 - 20)



g. Histograma con 5000 dimensiones
Rango (0 - 30)



3. Análisis de los resultados

Luego de analizar las gráficas resultantes podemos observar que conforme las dimensiones aumentan, el rango de distancia también lo hace, en el gráfico de dimensión 10, se observa que las distancias se encuentran en un rango de 0 a 2.5, mientras que en el gráfico de dimensión 5000, las distancias van hasta un rango de 30, mostrando así que conforme la dimensión aumenta los datos denotan un crecimiento exponencial hacia la derecha en cuanto a distancia. Al enfocarnos en la frecuencia de los datos con respecto a su dimensión, podemos observar que al aumentar la dimensión, la frecuencia de los datos también lo hace, si observamos la gráfica de la dimensión 10, podemos notar que los datos muestran mayor frecuencia en el rango central y menor frecuencia en el inicio y final del rango, sin embargo, en la dimensión 5000, notamos que si bien los datos muestran mayor frecuencia en el rango central, la frecuencia en el inicio y final del rango es mucho menor a comparación de las gráficas de menor dimensión.

4. Conclusiones

Podemos concluir que conforme el número de dimensiones aumenta, la distancia también y habrá mayor cantidad de datos similares que aumentarán la frecuencia de ciertos datos y disminuirán otros considerablemente a comparación de las gráficas de menor dimensión.

Link de github:

<https://github.com/CamilaLuque/EDA/tree/main>