

La maldición de la dimensionalidad

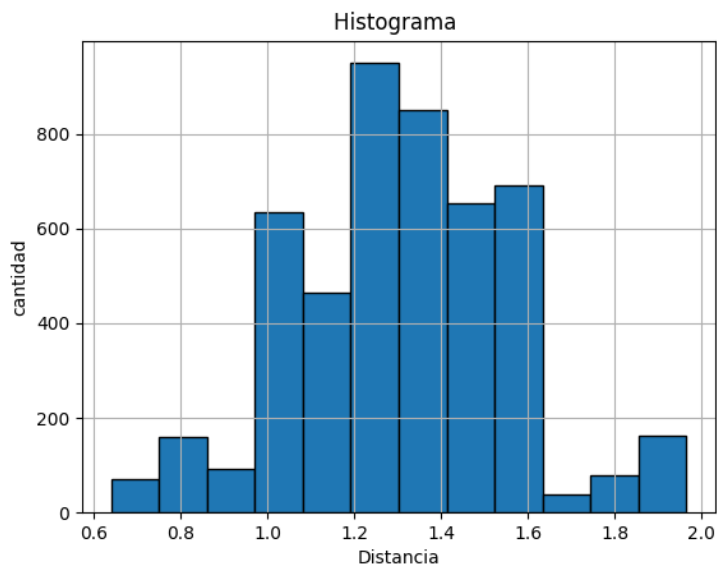
Es un problema cuando en un conjunto de datos se van aumentando las dimensiones o tienen una alta dimensionalidad.

Este experimento se hizo con la misma cantidad de puntos que son cien exactamente y con dimensiones (10,50,100,500,1000,2000,5000).

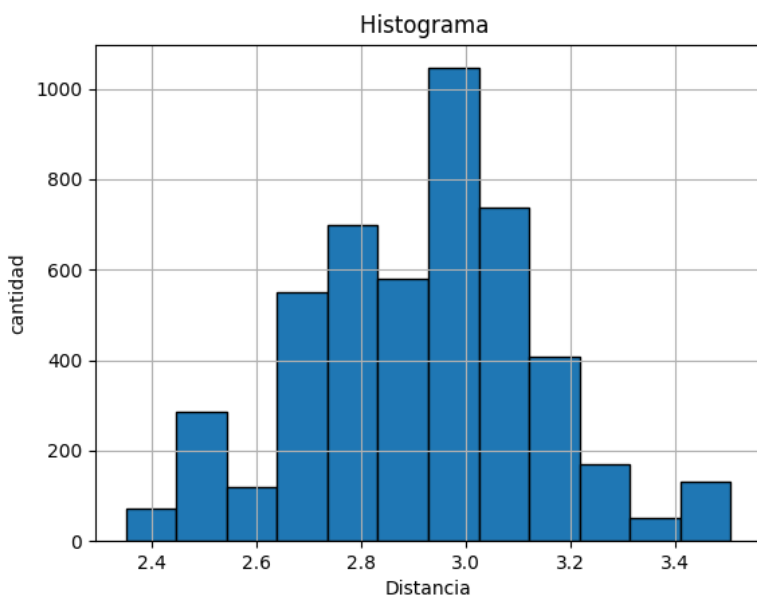
Como podemos observar las siete gráficas, cada vez las distancias de los puntos de distintas dimensiones van creciendo, esto se pasa a medida que vamos aumentando las dimensiones el volumen va creciendo, haciendo que la distancia de un punto a otro sea mayor.

Para demostrar que las distancias de puntos cada vez son mayores se tiene que usar para todas las dimensiones la misma cantidad de puntos, ya que no tendría sentido aumentar o disminuir los puntos ya que ocuparían un mayor o menor campo en el volumen.

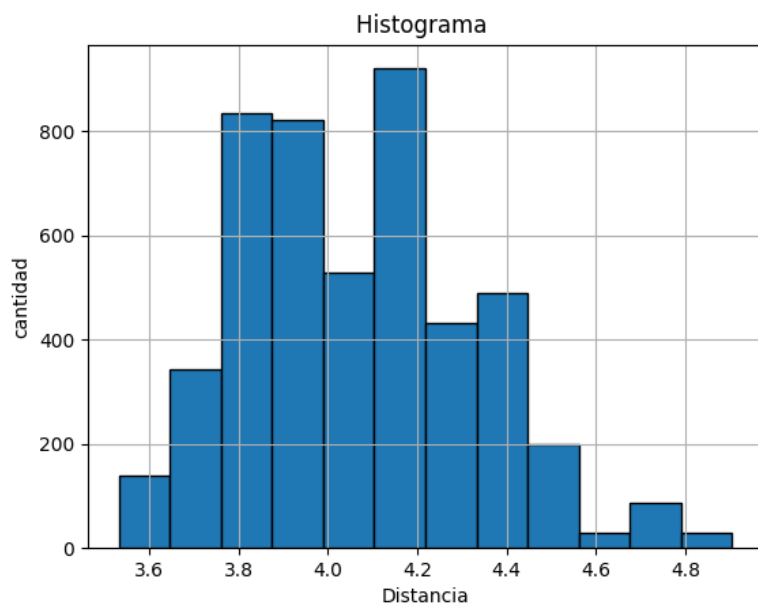
Dimensión 10



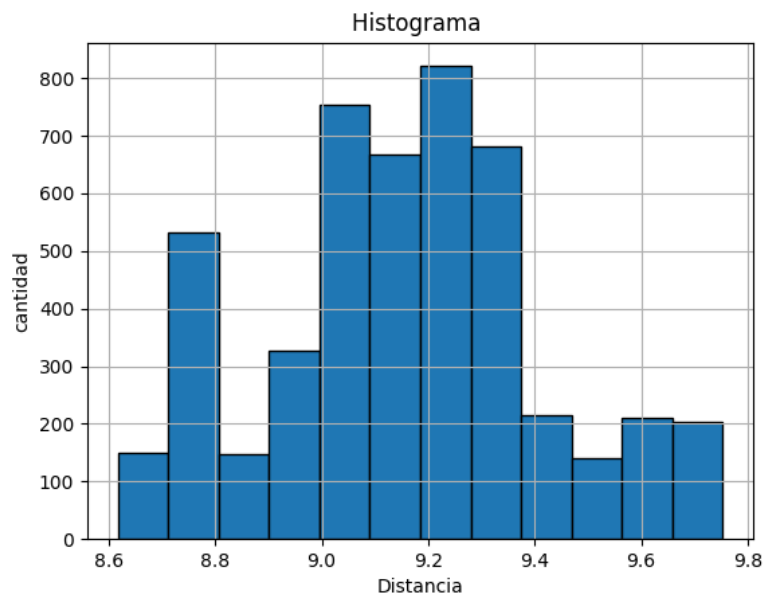
Dimensión 50



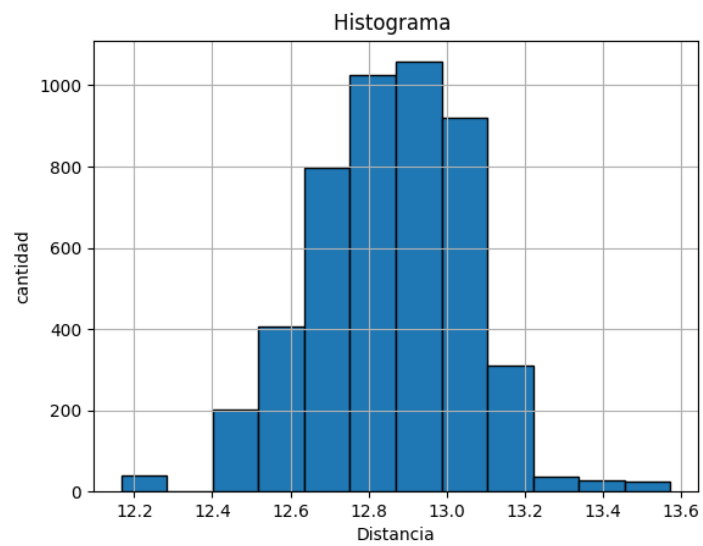
Dimensión 100



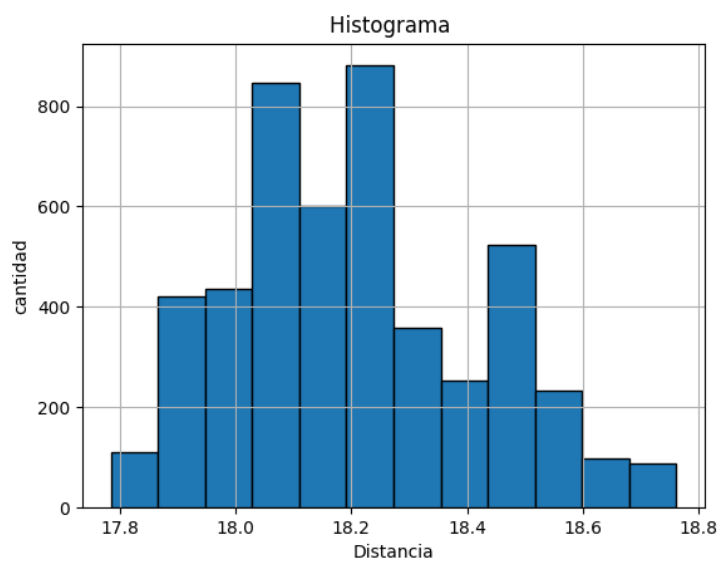
Dimensión 500



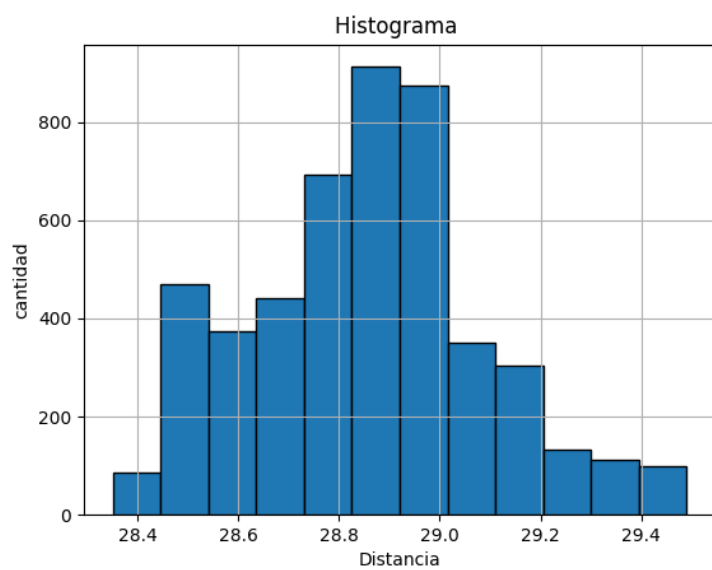
Dimensión 1000



Dimensión 2000



Dimensión 5000



- Uno de los problemas de tener una alta cantidad de dimensiones es el almacenamiento de los datos ya que algunos son más difíciles de procesar para la computadora.
- Cuando hay mas dimensiones la variabilidad de las distancias de los datos disminuye exponencialmente, no para todos los casos es un problema, pero para aprendizaje automático que se basa en distancias como vecinos mas cercanos o clustering si lo es.ⁱ

<https://github.com/nicolas2031/EDA.git>

ⁱ Heras, J. M. (2020, 19 septiembre). La maldición de la dimensión en machine learning - IArtificial.net. *IArtificial.net*. Recuperado 29 de agosto de 2023, de <https://www.iartificial.net/la-maldicion-de-la-dimension-en-machine-learning/>