

“AÑO DE LA UNIDAD, LA PAZ Y EL  
DESARROLLO”.



UNSA  
UNIVERSIDAD NACIONAL DE SAN AGUSTÍN DE AREQUIPA

ESCUELA PROFESIONAL DE CIENCIA DE LA  
COMPUTACIÓN  
ESTRUCTURA DE DATOS AVANZADOS

---

## Laboratorio 1: La maldición de la dimensionalidad

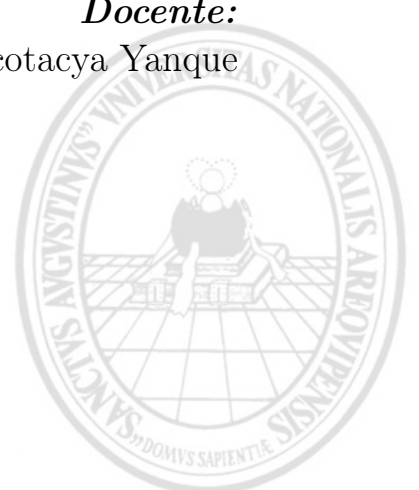
---

***Estudiante:***

Katherine Nikole Béjar Román

***Docente:***

Rosa Yuliana Gabriela Paccotacya Yanque



# 1. Análisis de distancias en espacios de alta dimensión

Este informe tiene como objetivo analizar las distancias entre puntos en espacios de alta dimensión, específicamente en dimensiones de 10, 50, 100, 500, 1000, 2000 y 5000. Para ello, se generaron datos aleatorios en cada espacio dimensional y se calcularon las distancias euclidianas entre todos los pares de puntos. A continuación, se presenta un análisis detallado de las distancias en cada dimensión:

## 1.1. Histograma de las distancias entre puntos con dimensión 10

En esta sección, analizaremos las distancias entre puntos en un espacio de 10 dimensiones.

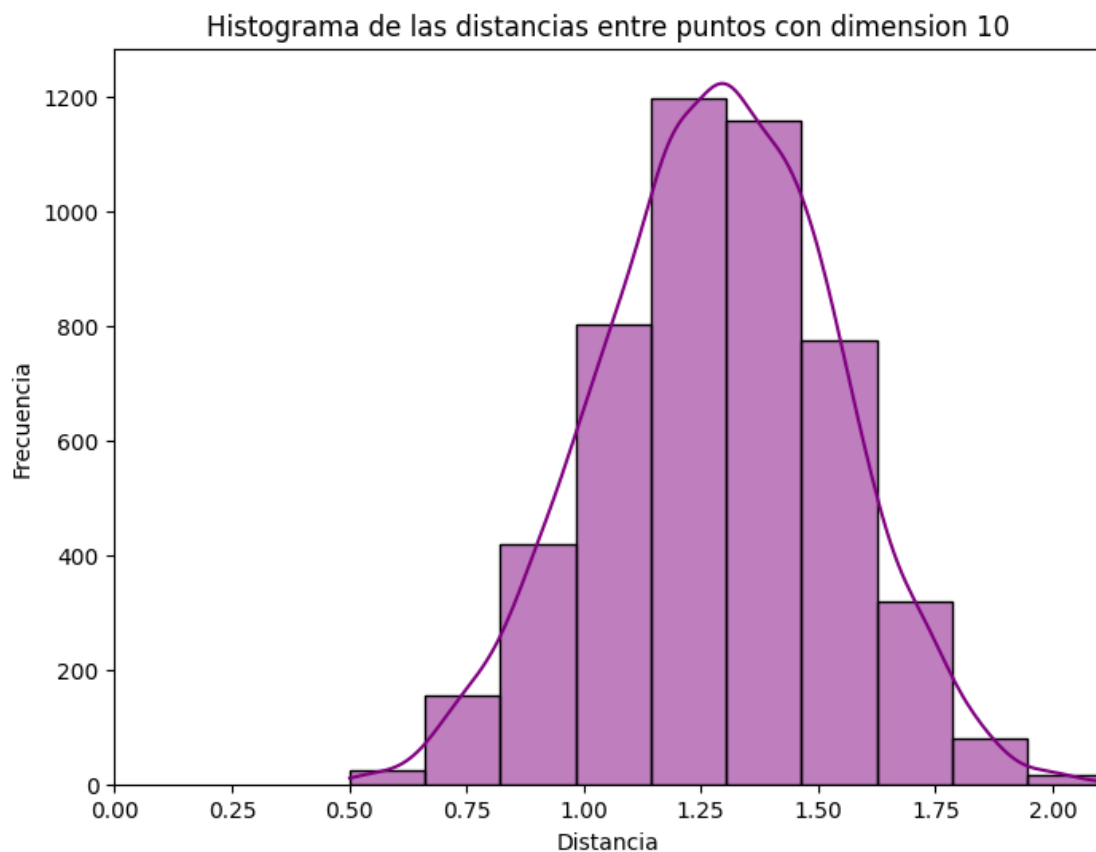


Figura 1: Histograma de las distancias entre puntos con dimensión 10

- **Distribución de las Distancias:** En dimensiones de 10, las distancias tienden a distribuirse en un rango de 0.5 a 2.1, con una apariencia bastante simétrica.
- **Valores en los Ejes X e Y:** El eje x representa las distancias entre puntos, mientras que el eje y muestra la frecuencia con la que ocurre cada distancia.
- **Cambios en la Distribución:** A medida que aumenta la dimensión del espacio, la distribución de las distancias tiende a volverse más dispersa, con valores más altos en el eje x y mayor variabilidad en el eje y. Esto se debe a la "maldición de la dimensionalidad".

```

ANALISIS DE LAS DISTANCIAS ENTRE PUNTOS CON DIMENSION 10
-----
Distancia Promedio: 1.2842420901326947
Distancia Mínima: 0.5013659777091073
Distancia Máxima: 2.106521851935357
Desviación Estándar: 0.2514504513583585

```

Figura 2: Resultado de las distancias entre puntos con dimensión 10

- **Promedio de las Distancias:** El promedio de las distancias es de aproximadamente 1.28, lo que indica que, en promedio, los puntos están a una distancia moderada entre sí.
- **Distancia Mínima:** La distancia mínima es de aproximadamente 0.50, lo que sugiere que algunos puntos están bastante cerca entre sí.
- **Distancia Máxima:** La distancia máxima es de aproximadamente 2.11, lo que indica que en este espacio de 10 dimensiones, algunos puntos están relativamente alejados.
- **Desviación Estándar:** La desviación estándar es de aproximadamente 0.25, reflejando una dispersión moderada en las distancias.

Este análisis proporciona una comprensión inicial de cómo se distribuyen las distancias en un espacio de 10 dimensiones y cómo cambian estas medidas estadísticas a medida que aumenta la dimensión.

## 1.2. Histograma de las distancias entre puntos con dimensión 50

En esta sección, se repite el análisis anterior, pero ahora en un espacio de 50 dimensiones.

- **Distribución de las Distancias:** En dimensiones más altas, como 50, las distancias tienden a distribuirse de manera más dispersa en comparación con el espacio de 10 dimensiones.
- **Valores en los Ejes X e Y:** Los valores en los ejes x e y siguen representando las distancias y la frecuencia respectivamente.
- **Cambios en la Distribución:** Se observa que la dispersión de las distancias es mucho mayor en comparación con la dimensión 10. La mayoría de las distancias tienden a ser más grandes.

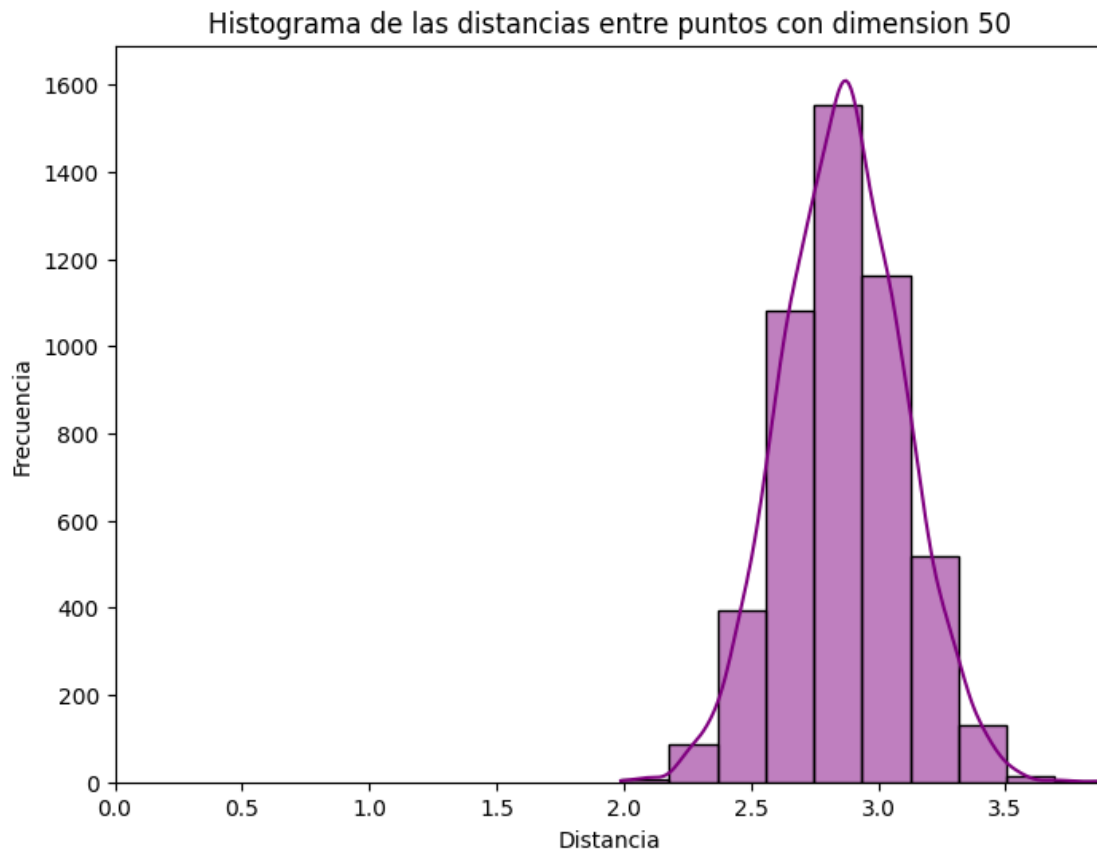


Figura 3: Histograma de las distancias entre puntos con dimensión 50

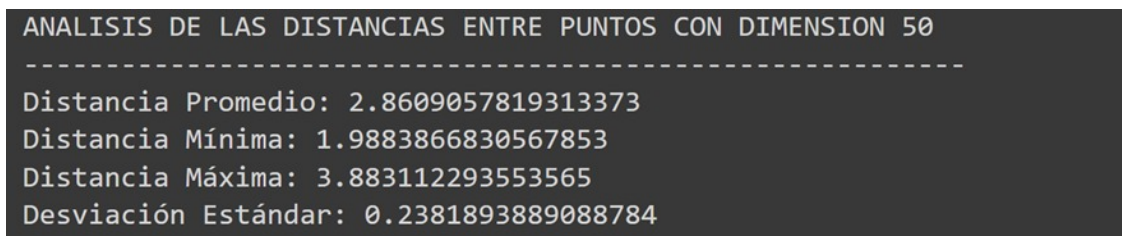


Figura 4: Resultado de las distancias entre puntos con dimensión 50

- **Promedio de las Distancias:** El promedio de las distancias es de aproximadamente 2.86 en comparación con la dimensión 10, donde era de 1.28. Esto indica que, en promedio, los puntos están más separados en este espacio de mayor dimensión.
- **Distancia Mínima:** La distancia mínima es de aproximadamente 1.99 en comparación con la dimensión 10, donde era de 0.50. Esto sugiere que, en dimensiones más altas, es más probable que algunos puntos estén más cerca entre sí.
- **Distancia Máxima:** La distancia máxima es de aproximadamente 3.88 en comparación con la dimensión 10, donde era de 2.11. Esto indica que en dimensiones más altas, es más probable encontrar puntos que están muy alejados entre sí.
- **Desviación Estándar:** La desviación estándar es de aproximadamente 0.24 en

comparación con la dimensión 10, donde era de 0.25. A pesar de la mayor dispersión, la variabilidad en las distancias sigue siendo moderada.

### 1.3. Histograma de las distancias entre puntos con dimensión 100

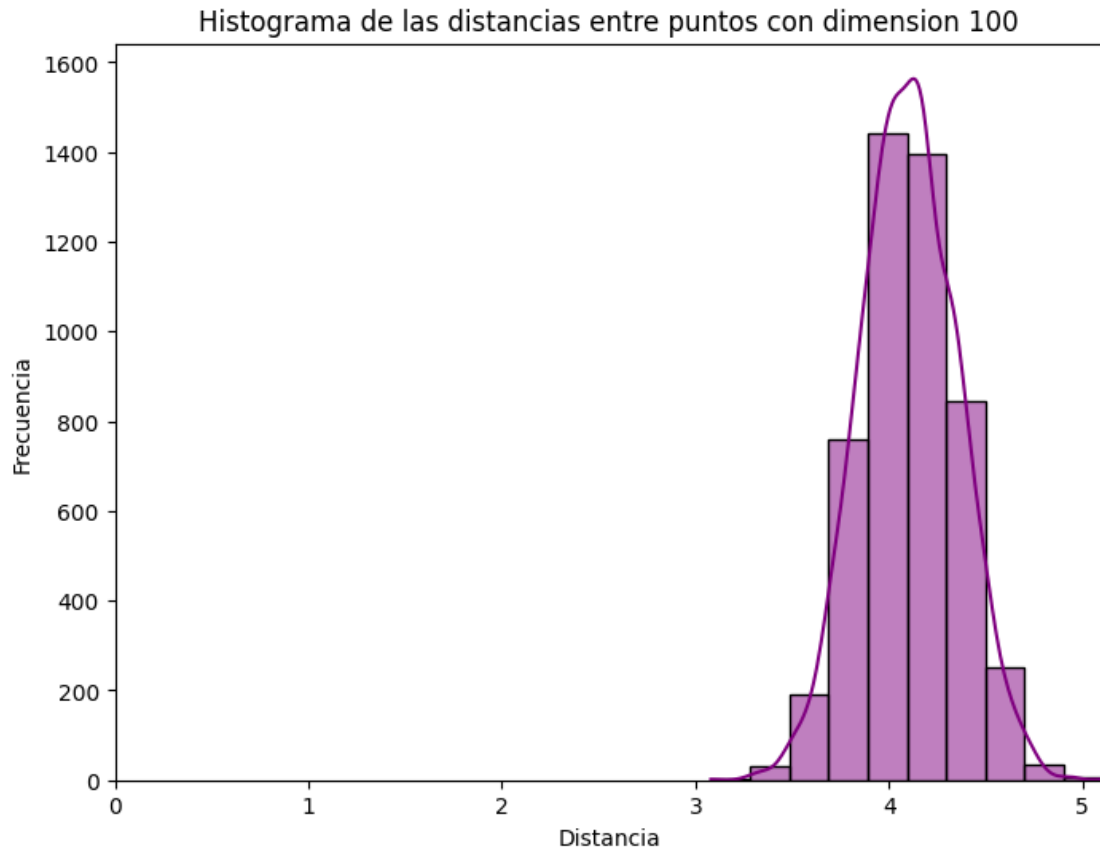


Figura 5: Histograma de las distancias entre puntos con dimensión 100

#### ANALISIS DE LAS DISTANCIAS ENTRE PUNTOS CON DIMENSION 100

```
-----
Distancia Promedio: 4.102731901596953
Distancia Mínima: 3.08025738524367
Distancia Máxima: 5.105836473329327
Desviación Estándar: 0.2502403777299548
```

Figura 6: Resultado de las distancias entre puntos con dimensión 100

## 1.4. Histograma de las distancias entre puntos con dimensión 500

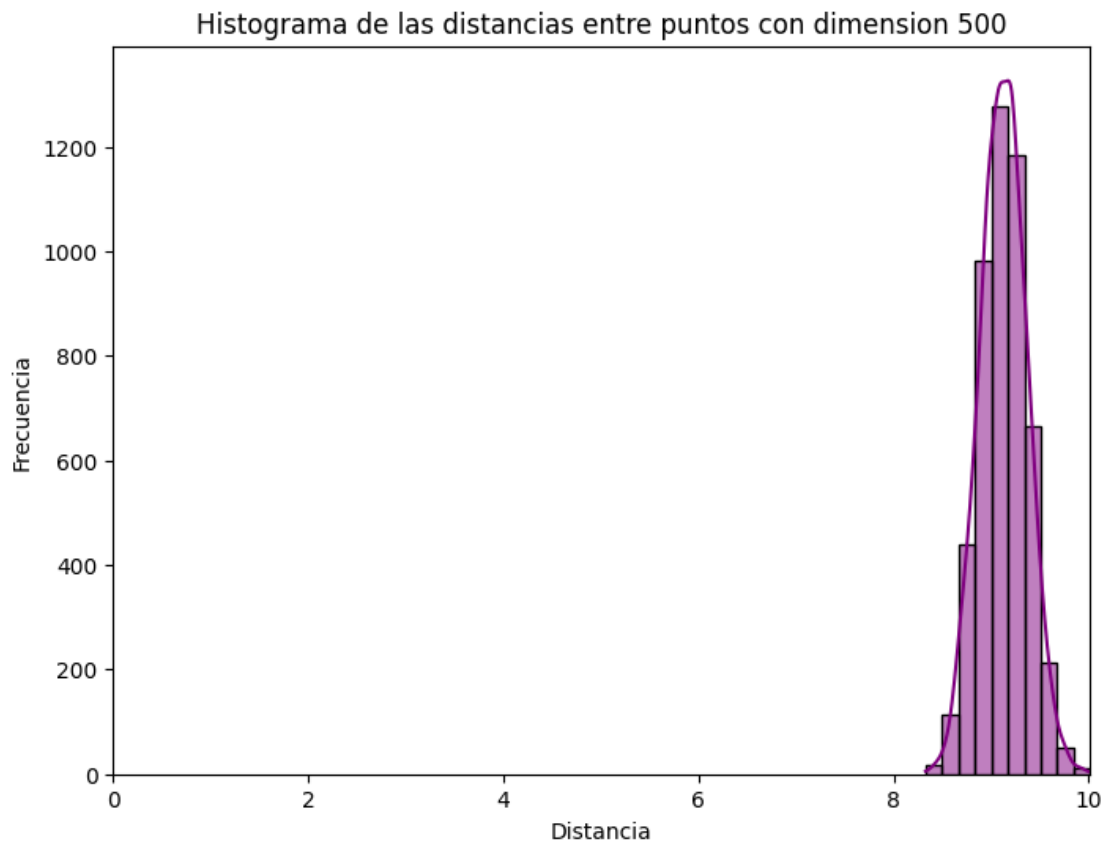


Figura 7: Histograma de las distancias entre puntos con dimensión 500

ANALISIS DE LAS DISTANCIAS ENTRE PUNTOS CON DIMENSION 500

-----  
 Distancia Promedio: 9.12592123554184  
 Distancia Mínima: 8.327042343334089  
 Distancia Máxima: 10.016707863771977  
 Desviación Estándar: 0.24160371614773757

Figura 8: Resultado de las distancias entre puntos con dimensión 500

## 1.5. Histograma de las distancias entre puntos con dimensión 1000

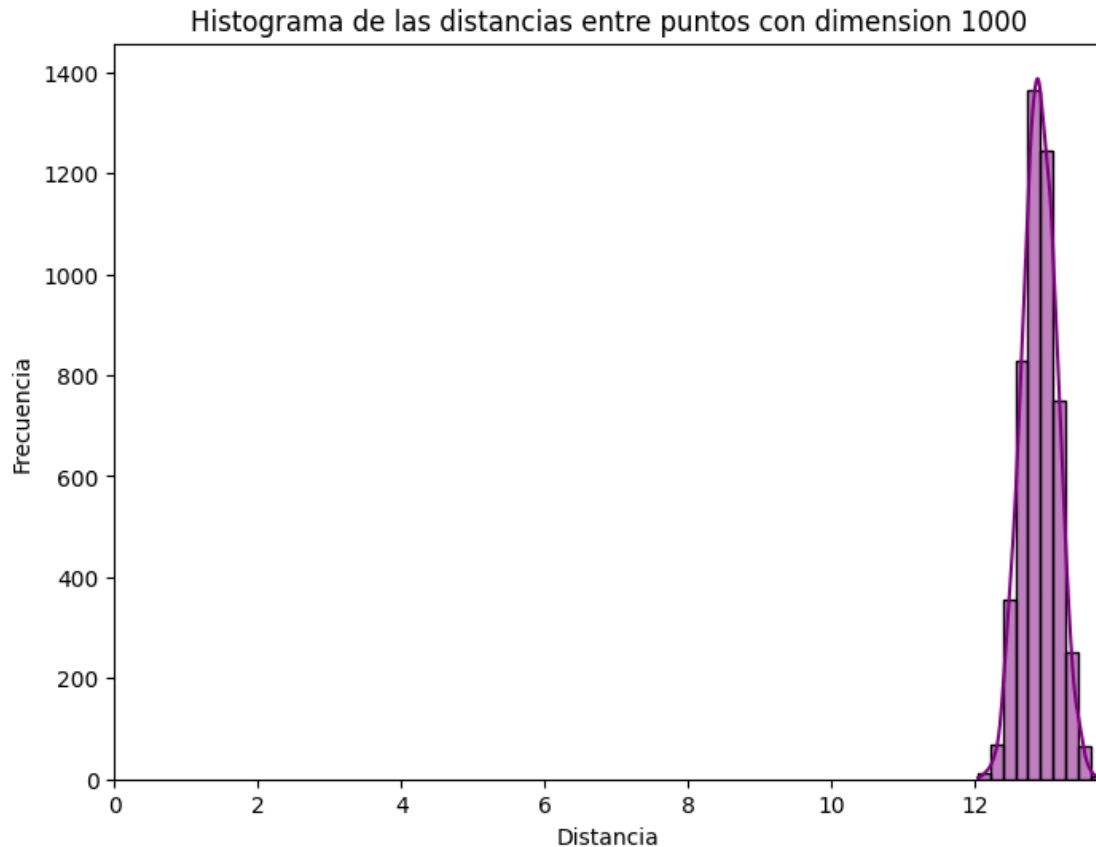


Figura 9: Histograma de las distancias entre puntos con dimensión 1000

### ANALISIS DE LAS DISTANCIAS ENTRE PUNTOS CON DIMENSION 1000

```
-----
Distancia Promedio: 12.892445430582764
Distancia Mínima: 12.038571439663443
Distancia Máxima: 13.777185827397464
Desviación Estándar: 0.24229453344194618
```

Figura 10: Resultado de las distancias entre puntos con dimensión 1000

## 1.6. Histograma de las distancias entre puntos con dimensión 2000

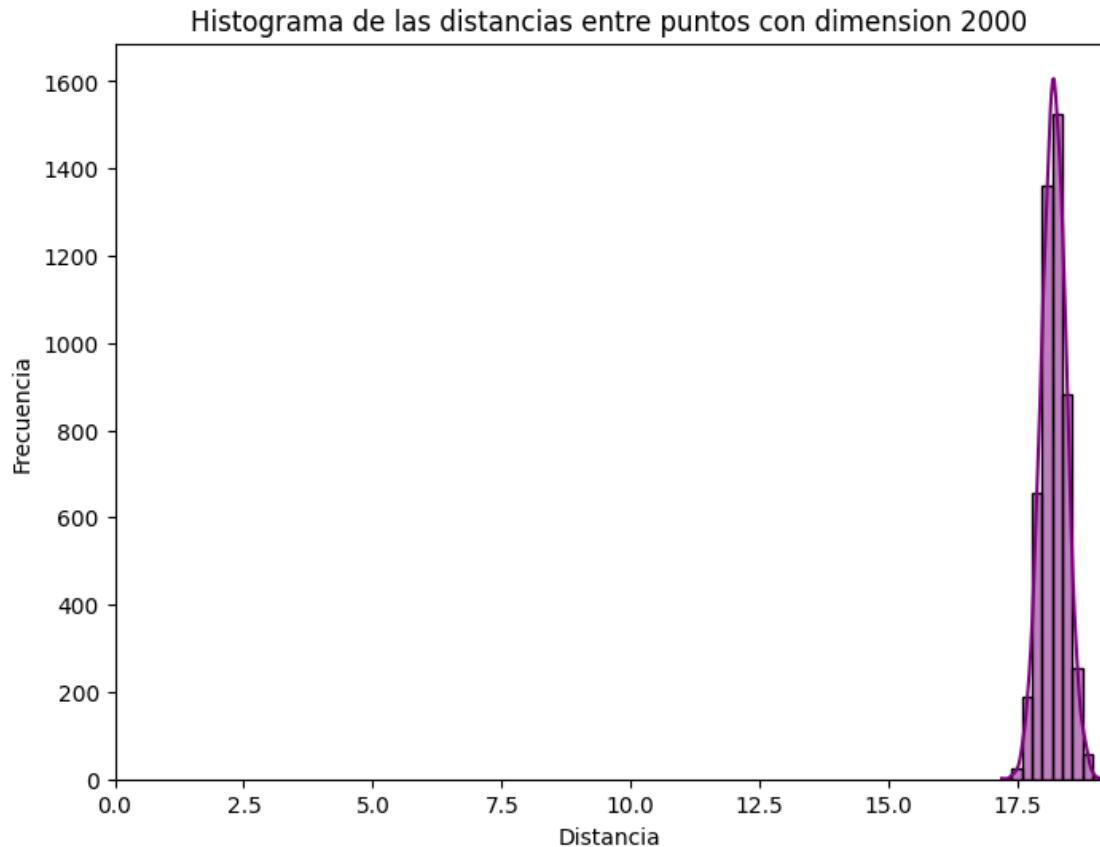


Figura 11: Histograma de las distancias entre puntos con dimensión 2000

### ANALISIS DE LAS DISTANCIAS ENTRE PUNTOS CON DIMENSION 2000

```
-----
Distancia Promedio: 18.197052158252955
Distancia Mínima: 17.189414085357754
Distancia Máxima: 19.149812414140282
Desviación Estándar: 0.2423798857117343
```

Figura 12: Resultado de las distancias entre puntos con dimensión 2000



## 1.7. Histograma de las distancias entre puntos con dimensión 5000

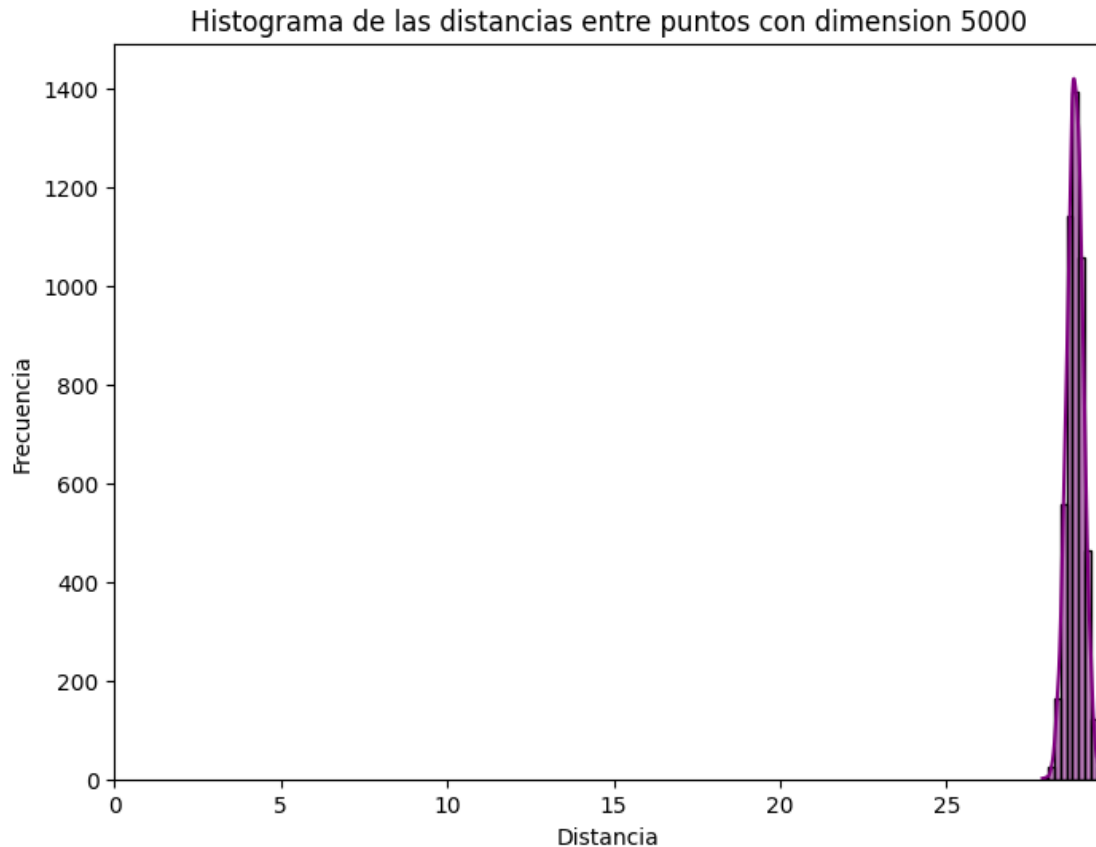


Figura 13: Histograma de las distancias entre puntos con dimensión 5000

```
ANALISIS DE LAS DISTANCIAS ENTRE PUNTOS CON DIMENSION 5000
-----
Distancia Promedio: 28.86353357950137
Distancia Mínima: 27.881281927546468
Distancia Máxima: 29.69557307517274
Desviación Estándar: 0.2481773578581373
```

Figura 14: Resultado de las distancias entre puntos con dimensión 5000

## 2. Conclusiones

1. A medida que aumenta la dimensión del espacio, el promedio de las distancias tiende a aumentar. Por ejemplo, en dimensiones de 10, el promedio es de aproximadamente 1.28, mientras que en dimensiones de 5000, es de aproximadamente 28.86. Esto indica que en dimensiones más altas, los puntos tienden a estar más separados en promedio.
2. La distancia mínima tiende a disminuir a medida que aumenta la dimensión. Por ejemplo, en dimensiones de 10, la distancia mínima es de aproximadamente 0.50, mientras que en dimensiones de 5000, es de aproximadamente 27.88. Esto sugiere que en dimensiones más altas, es más probable que algunos puntos estén más cerca entre sí.
3. La distancia máxima tiende a aumentar significativamente a medida que aumenta la dimensión. Por ejemplo, en dimensiones de 10, la distancia máxima es de aproximadamente 2.11, mientras que en dimensiones de 5000, es de aproximadamente 29.70. Esto indica que en dimensiones más altas, es más probable encontrar puntos que están muy alejados entre sí.
4. La desviación estándar aumenta de manera significativa a medida que aumenta la dimensión del espacio, lo que refleja una mayor variabilidad en las distancias entre puntos en dimensiones superiores.
5. Estos hallazgos subrayan la "maldición de la dimensionalidad", que tiene importantes implicaciones en la interpretación de datos y en el diseño de algoritmos de aprendizaje automático en espacios de alta dimensión.

Los códigos en GitHub lo puede encontrar aquí: <https://github.com/Katy-Bejar/EDA-LAB01>

## Referencias

- [1] Bregma (2011). `Std::UniformRealDistribution`, *cppreference.com*. Available at : [https://en.cppreference.com/w/cpp/numeric/random/uniform\\_real\\_distribution](https://en.cppreference.com/w/cpp/numeric/random/uniform_real_distribution)
- [2] Heras, J.M., C, F. and Reynaldo, J. (2020) La Maldición de la Dimensión en machine learning, IArtificial.net. Available at: <https://www.iartificial.net/la-maldicion-de-la-dimension-en-machine-learning/>
- [3] Histograma en seaborn con histplot (2022) PYTHON CHARTS | Visualización de datos con Python. Available at: <https://python-charts.com/es/distribucion/histograma-seaborn/>