



Universidad Católica
San Pablo

PROGRAMA PROFESIONAL

Ciencias de la computación

Tarea: La maldición de la dimensionalidad

Curso:

Estructura de datos

Integrantes:

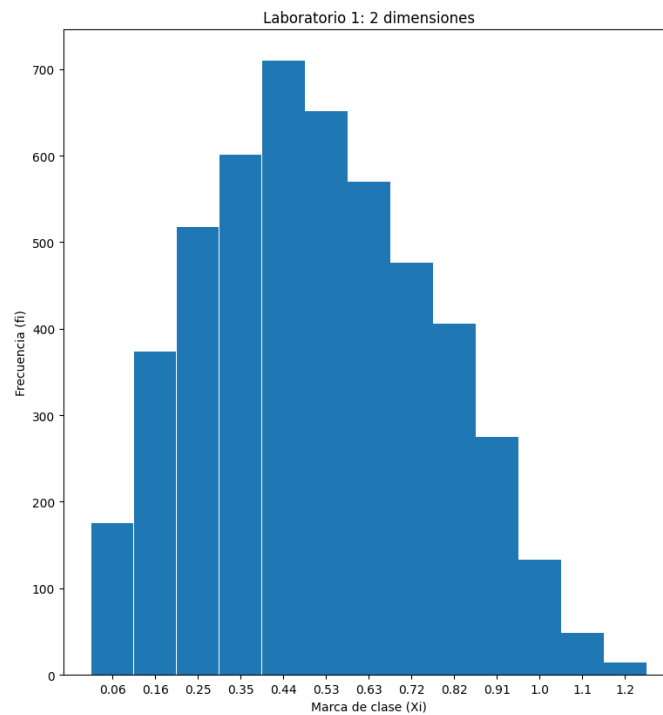
- Angel Josue Loayza Huarachi

CCOMP6-1.2

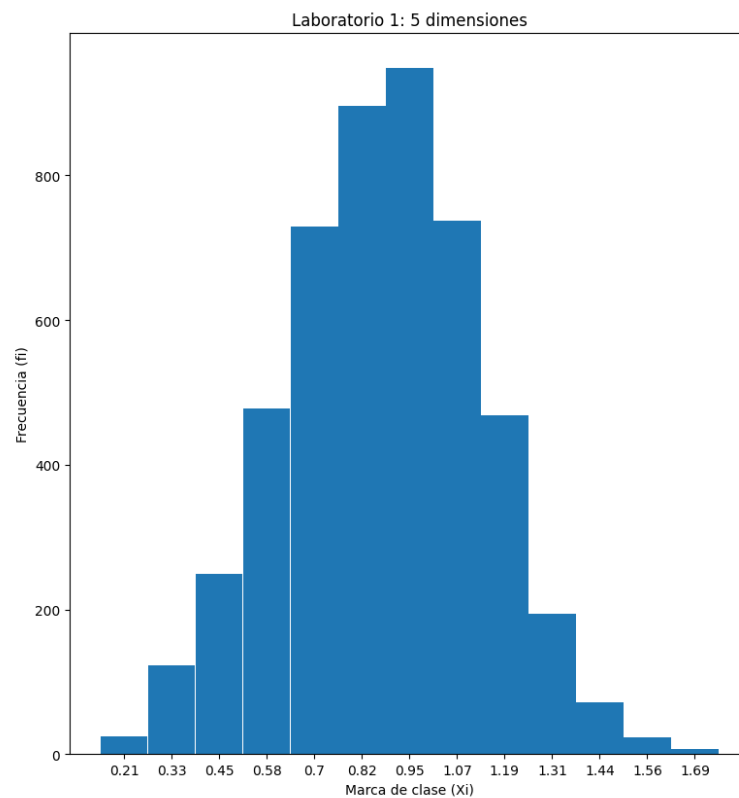
“Los alumnos declara haber realizado el presente trabajo de acuerdo a las normas de la Universidad Católica San Pablo”

2023

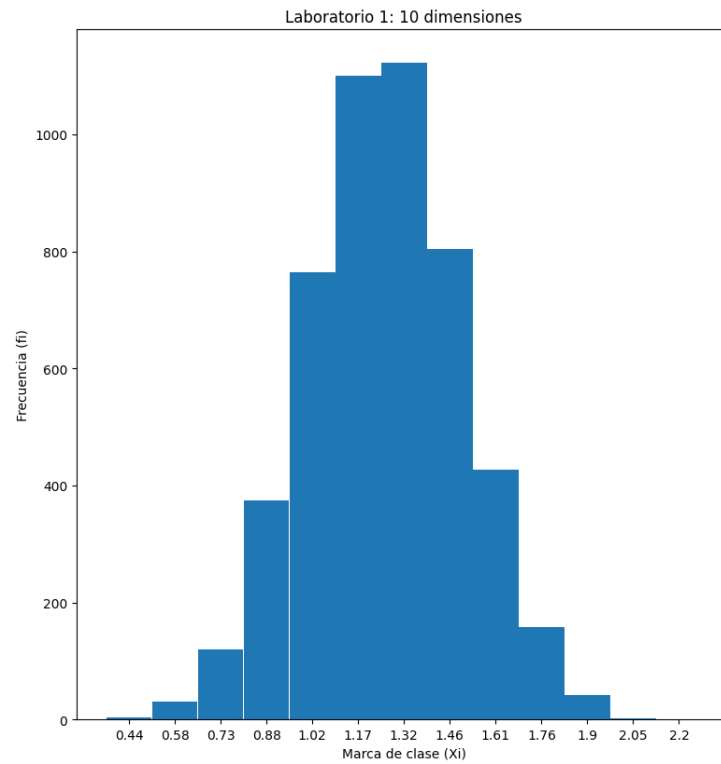
1. Gráficas resultantes:
a. 100 puntos con 2 dimensiones



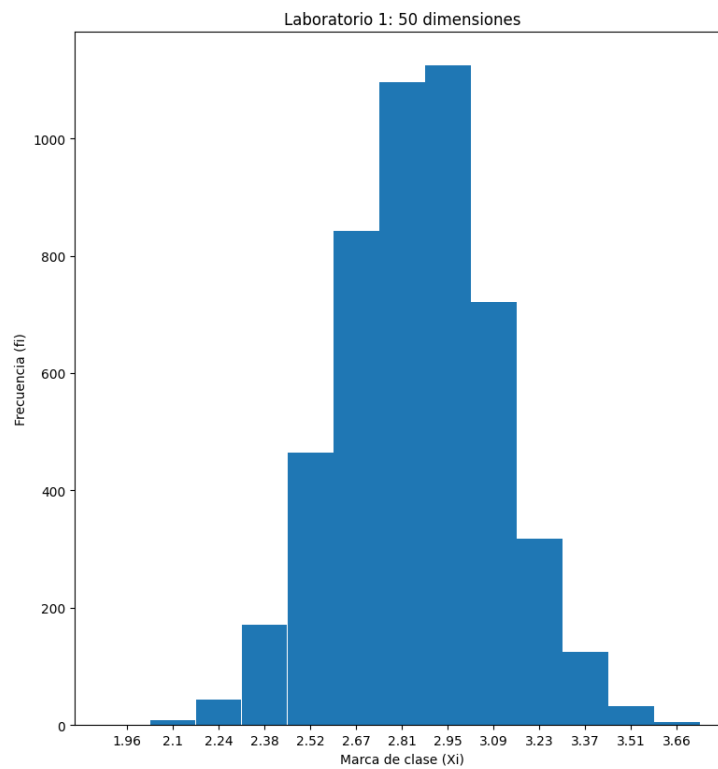
b. 100 puntos con 5 dimensiones



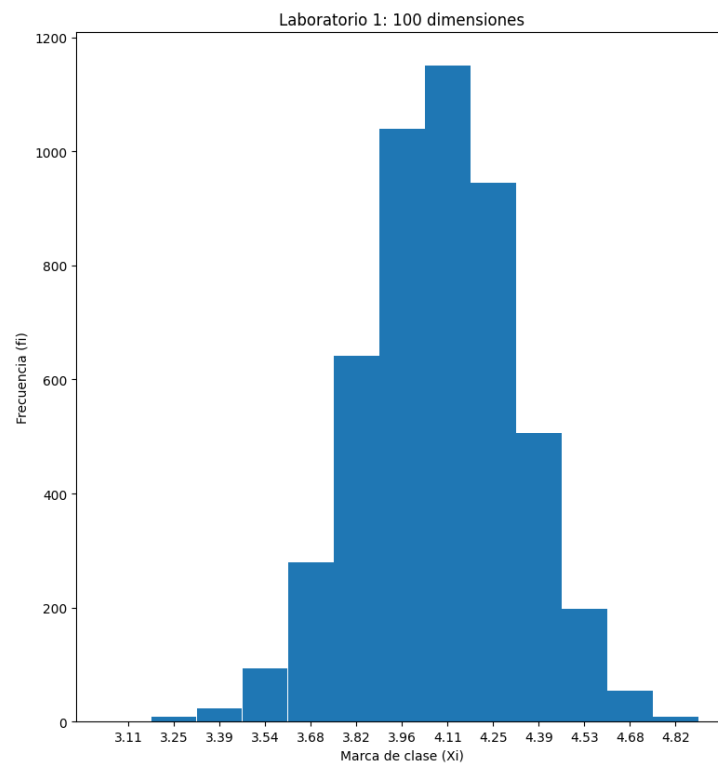
c. 100 puntos con 10 dimensiones



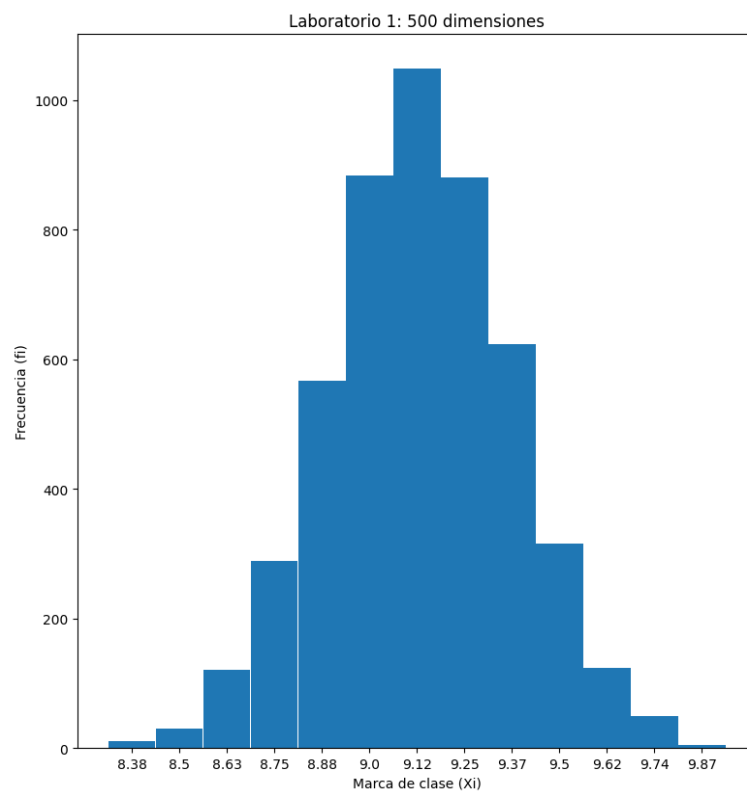
d. 100 puntos con 50 dimensiones



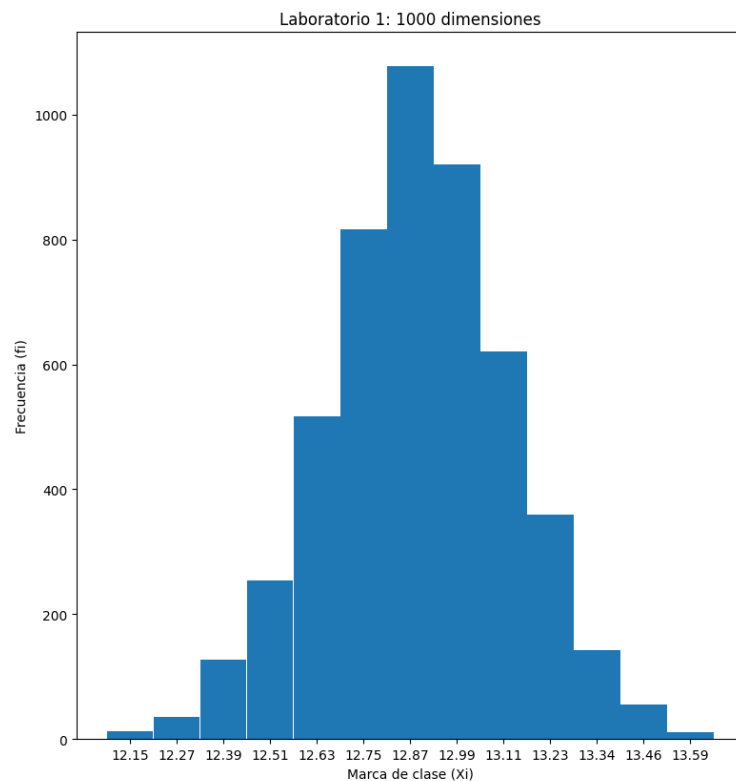
e. 100 puntos con 100 dimensiones



f. 100 puntos con 500 dimensiones



g. 100 puntos con 1000 dimensiones



2. Conclusiones:

- Efectivamente, gracias a los gráficos mostrados, podemos ver la campana de gauss formarse, a pesar del repetido cambio de dimensiones de los puntos.
- La maldición de la dimensionalidad también se ve muy presente al momento de aumentar las dimensiones de los puntos. Pues, si comparamos los resultados de la gráfica: “Laboratorio 1: 2 dimensiones” con la gráfica “Laboratorio 1: 1000 dimensiones”, hay una diferencia muy notoria en el eje x (resultado de la distancia entre 2 puntos). Eso quiere decir que, mientras más dimensiones sigamos aumentando, los puntos creados se dispersaran o alejarán mucho más.
- Respecto a la frecuencia de datos de cada gráfico, en este tipo de experimentos, es inevitable que los datos se acumulen más en la parte central. Si lo viéramos en un espacio tridimensional, y colocamos estos puntos en nuestro “espacio”, notaríamos que muchos de estos puntos se concentrarán en cierta área; y unos cuantos estarían más alejados de lo normal, que en este caso serían los que están en los extremos derecho e izquierdo de cada gráfico. Ello no quiere decir que mientras varios puntos se centren en un área sigue siendo bueno; a pesar de estar relativamente acumulados, siguen estando separados por una distancia muy notable. Y si el experimento fuera con 1 o 2 millones de datos, causaría un costo computacional abismal.

- El hecho de que los puntos se alejen más y más, conforme más dimensiones sigan aumentando, efectivamente será más complicado o costoso hacer operaciones como el “vecino más cercano”, o una búsqueda por similitud

3. Link de GitHub:

<https://github.com/angel452/Maldicion-de-la-Dimensionalidad>