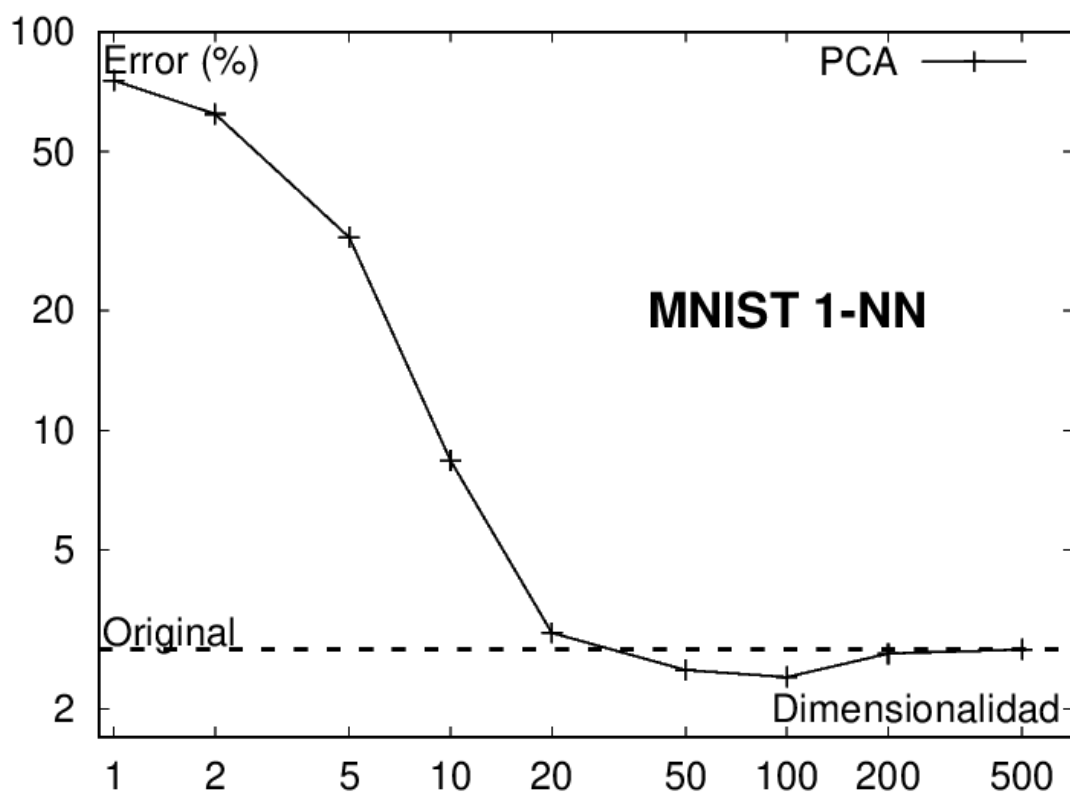


# Análisis del PCA y del clasificador knn

## Presentación y justificación

A continuación se presentarán las conclusiones obtenidas de la implementación y ejecución de la técnica de reducción de dimensionalidad PCA y el clasificador knn.

## Análisis de resultados experimentales



Esta gráfica corresponde a los datos obtenidos tras la ejecución del archivo implementado `pca+knn-exp.m`.

La representación muestra la relación entre la tasa de error del clasificador y la dimensionalidad empleada.

De esta prueba se puede concluir que la dimensión de los datos afecta en gran medida a la eficacia del clasificador. Uno de los hechos observables más importantes es que no siempre a menor dimensionalidad mayor tasa de acierto, si no que para la obtención de valores próximos es importante no reducir a dimensiones que hagan perder información sobre los objetos a clasificar.

En nuestro caso concreto, en la gráfica se puede observar que la menor tasa de error se consigue con una dimensionalidad igual a 100.

### Análisis de la influencia del PCA

Clasificador	Tasa de error
knn	3.09
knn+PCA(100)	2.84

Esta tabla muestra los resultados obtenidos de la ejecución tras el aprendizaje del clasificador.

Se puede observar que la aplicación del PCA sobre los datos permite una mejora notable a la hora de clasificarlos.

Esta mejora es explicable gracias a que tras la aplicación de la técnica empleada solo se emplean las 100 dimensiones más distintivas.

### Comparación con la base de datos MNIST

Con la ejecución de nuestro clasificador hemos obtenido resultados similares a los obtenidos por Kenneth Wilder y U. Chicago mediante un clasificador knn y el empleo de la distancia euclídea (L2).

Al aplicar la técnica PCA hemos mejorado los resultados hasta los obtenidos por Kenneth Wilder y U. Chicago utilizando una distancia L3.

### Conclusiones

Para el caso concreto de los dígitos manuscritos y la utilización del clasificador de los k vecinos más cercanos se observa que los mejores resultados se pueden obtener reduciendo la dimensionalidad de los datos a 100 dimensiones mediante la técnica PCA.

Además, con la comparación de los datos obtenidos y los presentes en la base MNIST se demuestra que la implementación llevada a cabo para este trabajo es correcta ya que se obtienen resultados similares.