



# Reducción de Dimensionalidad: Aplicaciones de SVD, PCA y Regresión Lineal en Imágenes y Datos Complejos

Wasinger, Pablo Tiago<sup>1</sup>, Sidiropulos, Felipe<sup>2</sup>

Departamento de Ingeniería, Universidad de San Andrés, Victoria, Buenos Aires, Argentina.

E-mails: <sup>1</sup>pwasinger@udesa.edu.ar, <sup>2</sup>fsidiropuloslamouret@udesa.edu.ar

**Abstract:** Este informe explora la aplicación de técnicas de reducción de dimensionalidad, específicamente la Descomposición en Valores Singulares (SVD) y el Análisis de Componentes Principales (PCA), en conjuntos de datos de alta dimensionalidad para identificar patrones internos y mejorar modelos predictivos. Primero, se utiliza SVD para comprimir y analizar imágenes en escala de grises, reduciendo la dimensionalidad mientras se preserva la información visual esencial; se encontró que con un número reducido de componentes principales ( $d$ ) se capturan efectivamente las características principales de las imágenes. Segundo, se analiza un conjunto de datos numéricos  $X \in \mathbb{R}^{2000 \times 206}$  y un vector de respuestas  $y \in \mathbb{R}^{2000}$  aplicando PCA para proyectar  $X$  en espacios de menor dimensión, preservando las relaciones entre muestras. Se aplican técnicas de agrupamiento y regresión lineal en los espacios original y reducidos, demostrando que los modelos en dimensiones reducidas logran una precisión comparable a los de dimensión completa, pero con un costo computacional significativamente menor. Los resultados demuestran que SVD y PCA son eficaces para reducir la dimensionalidad manteniendo información crítica, facilitando la extracción de patrones significativos y mejorando la eficiencia en el procesamiento de datos de alta dimensionalidad.

**Keywords:** Reducción de dimensionalidad, SVD, PCA, Compresión de imágenes, Regresión lineal, Clustering, K-means, Datos de alta dimensionalidad

## 1 Introducción

En la era de la información, el manejo y análisis de grandes volúmenes de datos es fundamental en diversas áreas científicas y tecnológicas. La alta dimensionalidad de los datos plantea desafíos significativos en términos de almacenamiento, procesamiento y extracción de información relevante. En este contexto, las técnicas de reducción de dimensionalidad como la Descomposición en Valores Singulares (SVD) y el Análisis de Componentes Principales (PCA) se han convertido en herramientas esenciales para simplificar datos complejos sin perder información crucial.

El presente informe tiene como objetivo explorar y aplicar estas técnicas en dos ámbitos específicos: la compresión y análisis de imágenes, y el estudio de conjuntos de datos numéricos de alta dimensionalidad. En la primera parte, se utiliza SVD para comprimir un conjunto de imágenes en escala de grises, reduciendo su dimensionalidad mientras se preserva la calidad visual. Esto no solo permite un almacenamiento más eficiente sino que también facilita la identificación de patrones y características comunes entre las imágenes. En la segunda parte, se aplica PCA para analizar un conjunto de datos numéricos, identificando patrones internos y agrupamientos naturales. Además, se emplea la regresión lineal mediante mínimos cuadrados para evaluar la capacidad predictiva de los modelos en espacios de dimensión reducida.

La importancia de este trabajo radica en demostrar cómo las técnicas de reducción de dimensionalidad pueden abordar eficazmente los desafíos asociados con datos de alta dimensión. Al reducir la complejidad de los datos, es posible mejorar el rendimiento computacional y facilitar el análisis sin comprometer significativamente la precisión o la interpretabilidad. Los resultados obtenidos evidencian la utilidad práctica de SVD y PCA en la extracción de información relevante y en la mejora de modelos predictivos.

Las técnicas utilizadas en este estudio son fundamentales en campos como el aprendizaje automático, el procesamiento de señales y la visión por computadora. Al resaltar su aplicación en casos concretos, este informe contribuye a una comprensión más profunda de estas herramientas y su potencial para resolver problemas

complejos en ciencia e ingeniería. La estructura del informe abarca el desarrollo experimental, donde se detallan los procedimientos y métodos aplicados; los resultados, que presentan los hallazgos y visualizaciones clave; y las conclusiones, que sintetizan las implicaciones y recomendaciones derivadas del estudio.

En resumen, este trabajo enfatiza la relevancia de las técnicas de reducción de dimensionalidad en el análisis de datos modernos y proporciona una guía práctica sobre su implementación y beneficios. Al abordar tanto la compresión de imágenes como el análisis de datos numéricos, se ofrece una perspectiva amplia sobre cómo estas metodologías pueden ser adaptadas a diferentes tipos de datos y objetivos analíticos.

## 2 Marco Teórico

### 2.1 Descomposición en Valores Singulares (SVD)

La Descomposición en Valores Singulares (SVD, por sus siglas en inglés) es una técnica fundamental en álgebra lineal que descompone una matriz en tres componentes esenciales. Dada una matriz  $A \in \mathbb{R}^{m \times n}$ , su descomposición SVD se expresa como:

$$A = U \Sigma V^T, \quad (1)$$

donde:

- $U \in \mathbb{R}^{m \times m}$  es una matriz ortogonal que contiene los vectores singulares izquierdos de  $A$ ,
- $\Sigma \in \mathbb{R}^{m \times n}$  es una matriz diagonal rectangular cuyos elementos no negativos en la diagonal son los valores singulares de  $A$ ,
- $V \in \mathbb{R}^{n \times n}$  es una matriz ortogonal que contiene los vectores singulares derechos de  $A$ .

#### 2.1.1 Proceso de Descomposición

El proceso de SVD se desarrolla de la siguiente manera:

1. **Cálculo de  $A^T A$  y  $AA^T$ :** Se calculan las matrices  $A^T A \in \mathbb{R}^{n \times n}$  y  $AA^T \in \mathbb{R}^{m \times m}$ , ambas simétricas y semidefinidas positivas. Estas matrices comparten los mismos valores propios no nulos, lo que es clave para la construcción de las matrices  $V$  y  $U$ .

2. **Obtención de vectores y valores propios:** Se calculan los valores propios  $\lambda_i$  y vectores propios  $v_i$  de  $A^T A$ , y los valores propios  $\lambda_i$  y vectores propios  $u_i$  de  $AA^T$ . Los valores propios satisfacen  $\lambda_i = \sigma_i^2$ , donde  $\sigma_i$  son los valores singulares de  $A$ .

3. **Construcción de las matrices  $U$ ,  $\Sigma$  y  $V$ :** Los vectores propios normalizados  $v_i$  forman las columnas de  $V$ , y los vectores propios normalizados  $u_i$  forman las columnas de  $U$ . La matriz  $\Sigma$  es diagonal, con los valores singulares  $\sigma_i = \sqrt{\lambda_i}$  ordenados de mayor a menor en la diagonal.

4. **Reducción de la dimensión:** Para comprimir los datos, se puede truncar  $\Sigma$  eliminando los valores singulares más pequeños y sus vectores asociados en  $U$  y  $V$ . Esto retiene solo los componentes principales de mayor magnitud, minimizando la pérdida de información y proporcionando una aproximación de menor rango de  $A$ .

### 2.1.2 Interpretación Geométrica

Geométricamente, la SVD descompone la transformación lineal representada por  $A$  en una secuencia de operaciones:

- Una rotación o reflexión inicial definida por  $V^T$ , que reorienta los vectores de entrada al espacio de los vectores singulares derechos.
- Un escalamiento por los valores singulares en  $\Sigma$ , donde cada valor singular  $\sigma_i$  define la magnitud del escalamiento en su dirección correspondiente.
- Una rotación o reflexión final dada por  $U$ , que reorienta los vectores escalados en el espacio de salida.

Este proceso revela cómo la matriz  $A$  actúa sobre los vectores de entrada y resalta las direcciones principales en las que  $A$  tiene mayor influencia.

## 2.2 Análisis de Componentes Principales (PCA)

El Análisis de Componentes Principales (PCA) es una técnica estadística de reducción de dimensionalidad que busca representar los datos en un espacio de menor dimensión, maximizando la varianza retenida. Dado un conjunto de datos  $X \in \mathbb{R}^{n \times p}$  con  $n$  observaciones y  $p$  variables, el PCA transforma los datos en un conjunto de componentes principales que capturan la mayor cantidad posible de varianza.

El PCA encuentra un nuevo sistema de coordenadas, o *componentes principales*, a través de los vectores propios de la matriz de covarianza de los datos. Estos vectores se ordenan en función de la magnitud de sus valores propios, que indican la proporción de la varianza explicada por cada componente.

### 2.2.1 Proceso Matemático de PCA

El proceso de PCA incluye los siguientes pasos:

1. **Centralización de los datos:** Se resta la media de cada variable en  $X$  para obtener un conjunto de datos centralizado  $X_{\text{centrado}}$  con media cero.

2. **Cálculo de la matriz de covarianza:** Se construye la matriz de covarianza  $C$  de los datos centralizados:

$$C = \frac{1}{n-1} X_{\text{centrado}}^T X_{\text{centrado}}. \quad (2)$$

3. **Obtención de vectores y valores propios:** Se calculan los valores propios  $\lambda_i$  y los vectores propios  $v_i$  de  $C$ . Los valores propios indican cuánta varianza explican las componentes asociadas.

4. **Selección de las  $d$  primeras componentes:** Se eligen los  $d$  vectores propios con mayores valores propios para formar la matriz de proyección  $V_d$ .

5. **Proyección de los datos:** Los datos se proyectan en el nuevo subespacio usando:

$$Z = X_{\text{centrado}} V_d, \quad (3)$$

donde  $Z \in \mathbb{R}^{n \times d}$  es la representación de los datos en el espacio de las componentes principales.

### 2.2.2 Relación entre PCA y SVD

El PCA se puede realizar utilizando la SVD de la matriz de datos centralizados  $X_{\text{centrado}}$ . Si aplicamos la descomposición SVD:

$$X_{\text{centrado}} = U \Sigma V^T, \quad (4)$$

entonces los vectores singulares derechos en  $V$  corresponden a los vectores propios de la matriz de covarianza  $C$ , y los valores propios  $\lambda_i$  están relacionados con los valores singulares  $\sigma_i$  por:

$$\lambda_i = \frac{\sigma_i^2}{n-1}. \quad (5)$$

De esta manera, el PCA puede ser implementado directamente a través de la SVD sin necesidad de calcular explícitamente la matriz de covarianza.

## 2.3 Norma de Frobenius

La norma de Frobenius es una medida utilizada para cuantificar el tamaño o la magnitud de una matriz. Dada una matriz  $A \in \mathbb{R}^{m \times n}$ , la norma de Frobenius de  $A$  se define como:

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2}, \quad (6)$$

donde  $a_{ij}$  representa el elemento de la matriz  $A$  en la fila  $i$  y la columna  $j$ . Esta norma puede interpretarse como la raíz cuadrada de la suma de los cuadrados de todos los elementos de la matriz, proporcionando una medida de la “energía” total contenida en  $A$ .

### 2.3.1 Propiedades de la Norma de Frobenius

La norma de Frobenius tiene varias propiedades útiles:

• **Subaditividad:** Para dos matrices  $A$  y  $B$  de las mismas dimensiones, se cumple:

$$\|A + B\|_F \leq \|A\|_F + \|B\|_F. \quad (7)$$

• **Invariancia bajo transformaciones ortogonales:** Si  $U$  y  $V$  son matrices ortogonales, entonces:

$$\|UAV\|_F = \|A\|_F. \quad (8)$$

• **Relación con los valores singulares:** La norma de Frobenius de una matriz también se puede expresar en términos de sus valores singulares  $\sigma_i$ :

$$\|A\|_F = \sqrt{\sum_{i=1}^r \sigma_i^2}, \quad (9)$$

donde  $r = \text{rango}(A)$  y  $\sigma_i$  son los valores singulares de  $A$ .

### 2.3.2 Aplicación de la Norma de Frobenius en la SVD

En el contexto de la SVD, la norma de Frobenius se utiliza para evaluar el error de aproximación de una matriz cuando se retienen solo los primeros  $d$  valores singulares. Si  $A_d$  es la matriz aproximada obtenida al truncar la SVD de  $A$  y conservar solo los primeros  $d$  valores singulares, el error de aproximación en la norma de Frobenius es:

$$\|A - A_d\|_F = \sqrt{\sum_{i=d+1}^r \sigma_i^2}. \quad (10)$$

Este resultado indica que la magnitud del error de la aproximación está determinada por la suma de los cuadrados de los valores singulares descartados. Por lo tanto, la norma de Frobenius es fundamental para entender cuánta “energía” o información se pierde al reducir la dimensión de una matriz utilizando la SVD. Un objetivo común es seleccionar  $d$  de manera que el error sea lo suficientemente pequeño como para que la aproximación  $A_d$  sea aceptable en términos de precisión y almacenamiento.

## 2.4 Regresión por Cuadrados Mínimos

La regresión por cuadrados mínimos es un método estadístico utilizado para ajustar un modelo lineal a un conjunto de datos, minimizando la suma de los cuadrados de los errores. Dado un conjunto de datos  $X \in \mathbb{R}^{n \times p}$  con  $n$  observaciones y  $p$  variables predictoras, y un vector de respuestas  $y \in \mathbb{R}^n$ , el objetivo es encontrar un vector de coeficientes  $\beta \in \mathbb{R}^p$  que minimice el error:

$$\min_{\beta} \|X\beta - y\|_2^2. \quad (11)$$

### 2.4.1 Formulación Matemática

El problema de minimización se resuelve encontrando  $\hat{\beta}$  que satisface:

$$\hat{\beta} = \arg \min_{\beta} \|X\beta - y\|_2^2. \quad (12)$$

La solución óptima  $\hat{\beta}$  se obtiene resolviendo las ecuaciones normales:

$$X^T X \hat{\beta} = X^T y. \quad (13)$$

Si  $X^T X$  es invertible, la solución explícita es:

$$\hat{\beta} = (X^T X)^{-1} X^T y. \quad (14)$$

### 2.4.2 Interpretación de la Solución

El vector de coeficientes  $\hat{\beta}$  representa la contribución de cada variable predictora al modelo lineal. Cada componente de  $\hat{\beta}$  indica el cambio esperado en la variable de respuesta  $y$  por unidad de cambio en la correspondiente variable predictora, manteniendo las demás variables constantes.

El ajuste por cuadrados mínimos busca minimizar la suma de los cuadrados de las diferencias entre los valores observados y los valores predichos por el modelo. Este método es eficiente y proporciona la mejor aproximación lineal en términos de error cuadrático medio, siempre que se cumplan las suposiciones de linealidad, independencia, homocedasticidad y normalidad de los errores.

## 2.5 Relación con PCA y SVD

El método de cuadrados mínimos puede beneficiarse del PCA y la SVD para mejorar la interpretación y la estabilidad numérica del modelo. Cuando los datos presentan multicolinealidad (alta correlación entre variables predictoras),  $X^T X$  puede ser casi singular, lo que conduce a instabilidades en la estimación de  $\hat{\beta}$ . En estos casos, es útil aplicar PCA para transformar  $X$  en un espacio de menor dimensión donde las variables sean ortogonales, mejorando así la estabilidad de la solución.

De manera similar, la SVD de  $X$  permite descomponer la matriz en:

$$X = U \Sigma V^T. \quad (15)$$

La solución por cuadrados mínimos puede expresarse en términos de SVD como:

$$\hat{\beta} = V \Sigma^{-1} U^T y, \quad (16)$$

donde  $\Sigma^{-1}$  es la pseudoinversa de la matriz diagonal  $\Sigma$ . Esta formulación es particularmente útil para la regularización y para el análisis de la contribución de cada componente en la predicción.

## 2.6 Varianza Explicada

La varianza explicada es un concepto fundamental en PCA que mide la cantidad de información o variabilidad en el conjunto de datos que es capturada por cada componente principal. Dado que el objetivo del PCA es encontrar un conjunto reducido de variables que retengan la mayor cantidad posible de información original, es importante evaluar qué porcentaje de la varianza total es explicado por cada componente.

Matemáticamente, la varianza explicada por la  $i$ -ésima componente principal está asociada al valor propio  $\lambda_i$  correspondiente de la matriz de covarianza. En términos de los valores singulares  $\sigma_i$  de  $X$  centrado, la varianza explicada se calcula como:

$$\text{Varianza explicada } (\sigma_i) = \frac{\sigma_i^2}{\sum_{j=1}^r \sigma_j^2} \times 100\%. \quad (17)$$

La suma de las varianzas explicadas por las primeras  $d$  componentes principales proporciona la varianza explicada acumulada, indicando el porcentaje de la varianza total retenida al reducir la dimensionalidad a  $d$  componentes. Este análisis permite seleccionar el número óptimo de componentes que balancea la retención de información y la simplicidad del modelo.

## 2.7 Algoritmo K-means

El algoritmo K-means es un método de clustering no supervisado que agrupa un conjunto de datos en  $k$  clusters distintos, donde cada muestra pertenece al cluster con el centroide más cercano. Este método minimiza la variación intra-cluster, es decir, la distancia entre las muestras y el centroide del cluster al que pertenecen. Dado un conjunto de datos  $X \in \mathbb{R}^{n \times p}$  y un número deseado de clusters  $k$ , el algoritmo K-means sigue los siguientes pasos:

1. **Inicialización:** Seleccionar aleatoriamente  $k$  centroides iniciales.
2. **Asignación:** Asignar cada muestra al centroide más cercano, formando  $k$  clusters.
3. **Actualización de centroides:** Recalcular los centroides de cada cluster como el promedio de las muestras asignadas a ese cluster.
4. **Iteración:** Repetir los pasos de asignación y actualización hasta que los centroides se stabilicen, es decir, hasta que no haya cambios significativos en la asignación de las muestras.

La función objetivo que minimiza el algoritmo es la suma de las distancias cuadráticas entre cada muestra y su centroide asignado:

$$\min \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2, \quad (18)$$

donde  $C_i$  representa el conjunto de muestras en el  $i$ -ésimo cluster y  $\mu_i$  es el centroide de ese cluster. El algoritmo K-means es sensible a la inicialización de los centroides, por lo que es común realizar múltiples ejecuciones y seleccionar la partición con el menor error total.

## 2.8 Función de Similaridad Gaussiana

En el análisis de similaridad entre pares de muestras, utilizamos una función de similaridad basada en una distribución Gaussiana, la

cual convierte las distancias entre muestras en valores de similaridad continua. Dada la distancia euclídea  $d_{ij} = \|x_i - x_j\|$  entre dos muestras  $x_i$  y  $x_j$ , definimos la similaridad  $K(x_i, x_j)$  como:

$$K(x_i, x_j) = \exp\left(-\frac{d_{ij}^2}{2\sigma^2}\right), \quad (19)$$

donde  $\sigma$  es un parámetro de escala que controla el grado de similaridad asignado en función de la distancia entre muestras. Un valor pequeño de  $\sigma$  hace que la función de similaridad decaiga rápidamente, asignando valores altos de similaridad solo a muestras muy cercanas, mientras que un  $\sigma$  grande suaviza el decaimiento, asignando valores altos de similaridad a una mayor cantidad de pares de muestras.

## 2.9 Criterios de Selección del Parámetro $\sigma$

La elección de  $\sigma$  es fundamental para obtener una medida de similaridad significativa. En este trabajo, consideramos tres criterios principales para su determinación:

### 1. Regla de Silverman:

$$\sigma = \hat{\sigma} \left( \frac{4}{d+2} \right)^{\frac{1}{d+4}} n^{-\frac{1}{d+4}}, \quad (20)$$

donde  $d$  es la dimensionalidad,  $n$  es el número de muestras y  $\hat{\sigma}$  es la desviación estándar de los datos. Este método es robusto para datos de alta dimensionalidad, pues considera tanto el número de características como el número de muestras.

### 2. Mediana de distancias:

$$\sigma = \text{mediana} \left\{ \|x_i - x_j\| \right\}_{i,j=1}^n. \quad (21)$$

Este enfoque proporciona una estimación robusta de la escala típica de las distancias en el conjunto de datos, siendo menos sensible a valores atípicos.

### 3. Percentil de distancias:

$$\sigma = \text{percentil}_{15} \left\{ \|x_i - x_j\| \right\}_{i,j=1}^n. \quad (22)$$

Este método utiliza el percentil 15 de las distancias por pares, lo cual es útil cuando los datos presentan clusters de diferentes densidades.

La selección de estos criterios se fundamenta en su capacidad para adaptarse a diferentes escalas y distribuciones de datos, su robustez demostrada en la literatura para datos de alta dimensionalidad, y su eficiencia computacional, crucial para conjuntos de datos grandes.

## 2.10 Índice de Silueta

El índice de silueta es una métrica ampliamente utilizada para evaluar la calidad de los clusters obtenidos en un análisis de clustering. Esta métrica proporciona una medida de qué tan bien se han agrupado los puntos dentro de sus clusters y qué tan separados están de otros clusters.

### 2.10.1 Definición del Índice de Silueta

Para cada punto  $i$  en el conjunto de datos, el índice de silueta  $s(i)$  se calcula considerando:

- **Cohesión  $a(i)$ :** Es la distancia promedio entre el punto  $i$  y todos los otros puntos dentro de su cluster  $C_i$ :

$$a(i) = \frac{1}{|C_i| - 1} \sum_{\substack{j \in C_i \\ j \neq i}} d(i, j). \quad (23)$$

- **Separación  $b(i)$ :** Es la distancia promedio entre el punto  $i$  y todos los puntos del cluster más cercano  $C_k$  que no contiene a  $i$ :

$$b(i) = \min_{C_k \neq C_i} \left( \frac{1}{|C_k|} \sum_{j \in C_k} d(i, j) \right). \quad (24)$$

El índice de silueta para el punto  $i$  se define como:

$$s(i) = \frac{b(i) - a(i)}{\max \{ a(i), b(i) \}}. \quad (25)$$

### 2.10.2 Interpretación del Índice de Silueta

El valor de  $s(i)$  oscila entre -1 y 1:

- $s(i) \approx 1$ : El punto  $i$  está bien agrupado dentro de su cluster y bien separado de otros clusters.
- $s(i) \approx 0$ : El punto  $i$  se encuentra en el borde entre dos clusters.
- $s(i) \approx -1$ : El punto  $i$  está más cerca de otro cluster que de su propio cluster, lo cual sugiere que podría estar mal asignado.

### 2.10.3 Índice de Silueta Promedio

Para evaluar la calidad general del clustering, se calcula el índice de silueta promedio para todos los puntos:

$$\text{Índice de Silueta Promedio} = \frac{1}{n} \sum_{i=1}^n s(i), \quad (26)$$

donde  $n$  es el número total de puntos en el conjunto de datos. Un índice de silueta promedio alto indica clusters bien definidos y separados, mientras que valores bajos sugieren posibles problemas en el clustering.

En el contexto de este trabajo, el índice de silueta promedio nos permite cuantificar la calidad de los clusters obtenidos mediante el algoritmo K-means en distintos subespacios de menor dimensión. Al aplicar esta métrica, podemos evaluar si la reducción de dimensionalidad mejora la separación de los clusters o si algunos puntos podrían estar mal asignados en el proceso de agrupamiento.

## 3 Desarrollo Experimental

### 3.1 Compresión de Imágenes

El propósito de esta sección es reducir la dimensionalidad de un conjunto de imágenes mediante la técnica de Descomposición en Valores Singulares (1). A partir de un dataset que contiene 16 imágenes, cada una representada como una matriz de  $p \times p$ , buscamos obtener una representación comprimida que capture las características principales de cada imagen. El proceso consta de tres etapas: construcción de la matriz de datos, descomposición y compresión de las imágenes, y análisis de los errores de reconstrucción.

#### 3.1.1 Construcción de la Matriz de Datos

Cada imagen se convierte inicialmente en una matriz en escala de grises de tamaño  $28 \times 28$ , donde cada elemento representa la intensidad de un píxel en una escala de 0 a 255. Posteriormente, las matrices se apllanan en vectores y se apilan, generando una matriz de datos  $A \in \mathbb{R}^{16 \times (28 \times 28)}$ , donde cada fila representa una imagen en formato vectorial. Este formato es conveniente para aplicar SVD y facilita la manipulación matemática de las imágenes.

#### 3.1.2 Motivación para Aplanar y Agrupar las Imágenes

La razón principal para juntar todas las imágenes en una sola matriz y aplanar cada una de ellas en un vector unidimensional radica en la necesidad de representar el conjunto de datos de manera eficiente

y uniforme para su posterior análisis. Cada imagen original, que puede ser vista como una matriz bidimensional de píxeles, contiene información visual que deseamos explorar en términos de patrones o características comunes. Al aplanar cada imagen en un vector, transformamos la estructura bidimensional de las imágenes en una forma unidimensional, facilitando así su manipulación matemática. Agrupar estos vectores aplanados como filas en una única matriz nos permite tratar el conjunto de imágenes como un conjunto de puntos en un espacio de alta dimensión, donde cada fila representa una imagen en su forma vectorial. Esta representación es crucial para técnicas como la descomposición en valores singulares (SVD), ya que nos permite analizar las imágenes colectivamente, identificando las variaciones y patrones comunes entre todas ellas en un espacio compartido.

### 3.1.3 Descomposición en Valores Singulares

Aplicamos la descomposición en valores singulares (SVD) 1, donde  $U \in \mathbb{R}^{n \times n}$ ,  $\Sigma \in \mathbb{R}^{n \times (p \times p)}$  y  $V \in \mathbb{R}^{(p \times p) \times (p \times p)}$ . La matriz  $\Sigma$  contiene los valores singulares ordenados de mayor a menor, lo cual nos permite seleccionar las  $d$  componentes principales que capturan la mayor parte de la varianza de los datos. Al retener solo las primeras  $d$  columnas de  $U$  y  $d$  filas de  $V^T$ , obtenemos una aproximación de la matriz original  $A$  en una dimensión reducida.

### 3.1.4 Reconstrucción de Imágenes con Diferentes Dimensiones

Para analizar el impacto de la compresión, reconstruimos cada imagen a partir de distintas dimensiones  $d$  (por ejemplo,  $d = 1, 5, 10, 16$ ), utilizando solo las  $d$  componentes principales. Esta reconstrucción se realiza mediante:

$$A_d = U_d \Sigma_d V_d^T$$

donde  $U_d$ ,  $\Sigma_d$  y  $V_d^T$  representan las matrices truncadas con las  $d$  componentes principales. Visualizamos cada imagen reconstruida en formato matricial  $p \times p$  y comparamos los resultados para observar la pérdida de detalles a medida que disminuye la cantidad de componentes conservadas.

### 3.1.5 Análisis del Error de Reconstrucción

Para evaluar la precisión de la compresión, calculamos el error entre las imágenes comprimidas y las originales usando la norma de Frobenius (6), donde  $A$  es la matriz de datos original y  $A_d$  es la aproximación comprimida.

### 3.1.6 Optimización de la Dimensionalidad $d$ para la Compresión de Imágenes

Nuestro objetivo aquí es determinar el número óptimo de componentes principales  $d$  necesario para reducir el error de reconstrucción de un conjunto de imágenes comprimidas mediante la Descomposición en Valores Singulares (SVD) a menos del 10%. Para lograrlo, hemos realizado los siguientes pasos: aplicar SVD, reconstruir las imágenes comprimidas con diferentes valores de  $d$ , y calcular el error de reconstrucción utilizando la norma de Frobenius (3.1.5).

### 3.1.7 Extracción de Patrones Visuales en Imágenes usando SVD

El objetivo de esta sección es identificar y visualizar los patrones más representativos en un conjunto de imágenes en escala de grises mediante la descomposición en valores singulares (SVD). Este análisis permite descubrir patrones comunes en el conjunto de datos, proporcionando una representación de las características visuales predominantes en las imágenes. Cada patrón, definido por los principales autovectores de la matriz de datos, corresponde a estructuras o texturas presentes en las imágenes, y nos ayuda a entender cómo se distribuye la información visual en el conjunto.

Para este análisis, partimos de una colección de 16 imágenes en escala de grises, las cuales convertimos a una forma matricial apta

para el procesamiento matemático. Cada imagen fue aplanada en un vector unidimensional, y todos estos vectores se apilaron para formar una matriz de datos  $A \in \mathbb{R}^{16 \times n}$ , donde  $n$  representa el número de píxeles de cada imagen.

### 3.1.8 Aplicación de SVD y Extracción de Patrones Principales

El primer paso en nuestro proceso fue realizar una descomposición en valores singulares (SVD) sobre la matriz  $A$  de datos de las imágenes. Esta técnica nos permite descomponer la matriz  $A$  en tres componentes principales:  $U$ ,  $S$ , y  $V^T$ , donde  $V^T$  contiene los autovectores o componentes principales en el espacio de las columnas de  $A$ . Cada autovector corresponde a un patrón característico que representa la dirección de mayor varianza en el conjunto de imágenes.

Para este análisis, nos enfocamos en los primeros 12 autovectores, los cuales corresponden a los patrones más significativos o 'importantes'. Estos autovectores fueron seleccionados debido a que capturan la mayoría de la varianza en los datos, lo cual implica que reflejan las características visuales predominantes. A través de estos patrones, podemos explorar la estructura subyacente y las similitudes entre las imágenes del conjunto.

### 3.1.9 Visualización de Patrones

Con el fin de interpretar visualmente los patrones extraídos, se reconstruyeron las imágenes de los autovectores seleccionados. Cada patrón fue reestructurado en una matriz bidimensional que representa su forma visual original. Los patrones fueron normalizados para asegurar que sus valores estuvieran dentro del rango adecuado de intensidades y fueran visualmente interpretables. La visualización de estos patrones permite identificar las características predominantes que emergen en el conjunto de imágenes, como estructuras repetitivas o texturas comunes.

Finalmente, se presentaron los primeros 12 patrones en una cuadrícula de imágenes, donde cada imagen corresponde a un autovector reestructurado. Esta visualización facilita la interpretación de las características visuales más relevantes en el conjunto de datos, ayudando a reconocer patrones que podrían estar presentes en todas las imágenes o que se repiten a lo largo del conjunto.

## 3.2 Análisis de Similaridad y Cuadrados Mínimos

En esta sección, abordamos el análisis de un conjunto de datos de alta dimensionalidad para identificar patrones y relaciones internas mediante técnicas de reducción de dimensionalidad y regresión. Contamos con una matriz de datos  $X \in \mathbb{R}^{2000 \times 206}$ , que representa 2000 muestras con 206 características cada una, y un vector de respuestas  $y \in \mathbb{R}^{2000}$ , ambos de los cuales desconocemos que representan los datos.

El objetivo principal es explorar la estructura interna de  $X$ , identificar posibles agrupamientos de las muestras y evaluar la capacidad de predecir  $y$  a partir de  $X$  utilizando regresión lineal, tanto en el espacio original como en espacios de dimensión reducida obtenidos mediante Análisis de Componentes Principales (PCA).

### 3.2.1 Análisis Exploratorio de los Datos

Comenzamos con un análisis exploratorio de la matriz de datos original  $X$ . Para ello, realizamos un gráfico de  $X$  para visualizar patrones o estructuras que pudieran ser evidentes a simple vista. Esta visualización nos permite tener una primera impresión sobre la distribución de los datos y posibles correlaciones entre las características.

### 3.2.2 Análisis de Valores Singulares

Con el fin de comprender la contribución de cada componente a la variabilidad total de los datos, calculamos los valores singulares de  $X$  y analizamos la varianza explicada 17 por cada uno. Esto nos permite determinar cuántas componentes principales son necesarias para capturar la mayor parte de la información contenida en  $X$ .

### 3.2.3 Reducción de Dimensionalidad y Matrices de Similaridad

Procedimos a reducir la dimensionalidad de  $X$  utilizando PCA 4, proyectando los datos en espacios de dimensión  $d = 2, 6, 10$  y  $p = 206$  (la dimensión original). Para cada valor de  $d$ , construimos las matrices de similaridad entre muestras en los espacios reducidos, utilizando la función de similaridad no lineal definida en el marco teórico:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right),$$

donde  $\sigma$  fue elegido según el criterio de la mediana 21 ya que al intentar . Estas matrices nos permiten analizar cómo se preservan las relaciones de similaridad entre muestras al reducir la dimensionalidad.

El parámetro  $\sigma$  fue elegido según el criterio de la mediana, debido a que resulta sencilla de calcular y es robusta. Esto asegura que el cálculo de  $\sigma$  sea eficiente y no se vea afectado por valores atípicos en los datos. A su vez, al utilizar la mediana como criterio, el parámetro  $\sigma$  introduce un adecuado nivel de contraste en la matriz de similaridad. Esto facilita la diferenciación entre pares de elementos que son muy similares y aquellos que no lo son, contribuyendo a una mejor discriminación en la estructura de los datos.

### 3.2.4 Proyección y Visualización en Dimensión 2

Para visualizar de manera intuitiva la distribución de las muestras, proyectamos los datos en un espacio de dimensión  $d = 2$  utilizando las dos primeras componentes principales. Esta proyección nos permite observar si existen agrupamientos naturales o patrones discernibles en los datos cuando se representan en dos dimensiones.

### 3.2.5 Aplicación de K-Means y Evaluación con el Índice de Silueta

Aplicamos el algoritmo de agrupamiento no supervisado  $K$ -means 10 sobre los datos proyectados en  $d = 2$ , con  $k = 2$  clusters, basándonos en la observación de posibles agrupamientos en la visualización anterior. Para evaluar la calidad de los clusters obtenidos, utilizamos el índice de silueta 26. Este índice nos proporciona una medida de qué tan bien se separan los clusters y la cohesión interna de las muestras dentro de cada uno.

### 3.2.6 Análisis del primer Vector Singular y el Vector $\beta$

Para profundizar en el entendimiento de las características más influyentes, analizamos el primer vector singular de la descomposición SVD de  $X$ , identificando las columnas (características) que tienen mayor peso en este vector. Esto nos indica cuáles son las variables que más contribuyen a la variabilidad principal de los datos.

Además, realizamos un ajuste de mínimos cuadrados en el espacio original para obtener el vector de coeficientes  $\beta$  que minimiza la norma  $\|X\beta - y\|^2$ . Analizamos los valores de  $\beta$  para identificar las características que tienen mayor impacto en la predicción de  $y$ .

### 3.2.7 Regresión Lineal en Espacios de Dimensión Reducida

Con el objetivo de evaluar la capacidad predictiva de los modelos lineales en espacios de dimensión reducida, aplicamos regresión lineal sobre las proyecciones de  $X$  en espacios de diferentes dimensiones  $d$ . Para ello, implementamos dos métodos de ajuste:

1. **Método 80/20:** Dividimos aleatoriamente la matriz reducida  $Z$  en un conjunto de entrenamiento (80% de las muestras) y un conjunto de prueba (20% restante). Ajustamos el modelo de regresión lineal utilizando el conjunto de entrenamiento y evaluamos su desempeño prediciendo  $y$  en el conjunto de prueba. Este método

simula un escenario real de predicción sobre datos no vistos durante el entrenamiento.

2. **Método 100/20:** Utilizamos toda la matriz reducida  $Z$  para ajustar el modelo de regresión lineal y luego predecimos  $y$  en un 20% aleatorio de las muestras. Este enfoque nos permite evaluar el modelo utilizando toda la información disponible para el ajuste.

### 3.2.8 Determinación de la Dimensionalidad Óptima

Para encontrar el valor óptimo de  $d$  que balancea la complejidad del modelo y su capacidad predictiva, realizamos iteraciones de PCA para distintos valores de  $d$ , aplicando el método 80/20 en cada caso. Calculamos el error de predicción en cada iteración y comparamos los resultados para identificar el valor de  $d$  que minimiza el error.

### 3.2.9 Comparación entre $d = 2$ y $d = 206$

Realizamos un análisis comparativo entre los modelos ajustados en el espacio reducido con  $d = 2$  y en el espacio original con  $d = 206$ . Para ambos casos, ejecutamos 50 repeticiones del método 80/20, calculando el error de predicción en cada iteración. Este análisis nos permite evaluar el impacto de la reducción de dimensionalidad en el desempeño del modelo de regresión.

### 3.2.10 Análisis de los Errores de Predicción

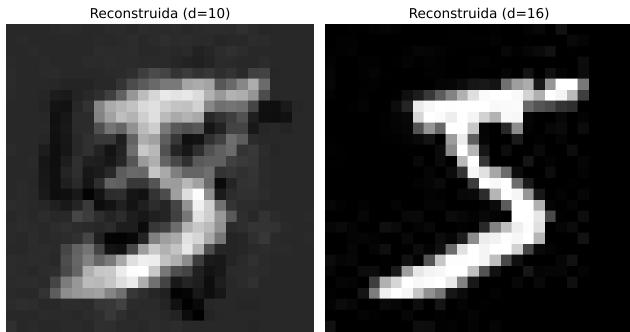
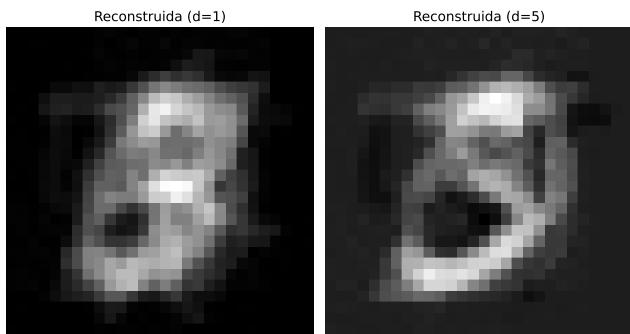
Comparando los errores obtenidos con los métodos 80/20 y 100/20 para  $d = 2$  y  $d = 206$ , evaluamos la consistencia y precisión de los modelos en diferentes escenarios de ajuste y prueba. Esto nos ayuda a comprender la influencia de la dimensionalidad y del tamaño del conjunto de entrenamiento en la capacidad predictiva del modelo.

### 3.2.11 Promedio de Coeficientes $\beta$ sobre Múltiples Iteraciones

Para investigar la estabilidad de los coeficientes  $\beta$ , realizamos 50 iteraciones del ajuste de mínimos cuadrados con el método 80/20 para  $d = 2$  y  $d = 206$ . Calculamos el promedio de los vectores  $\beta$  obtenidos en cada iteración, lo cual nos permite analizar la variabilidad de los coeficientes y determinar qué características mantienen consistentemente un peso significativo en la predicción de  $y$ .

## 4 Resultados

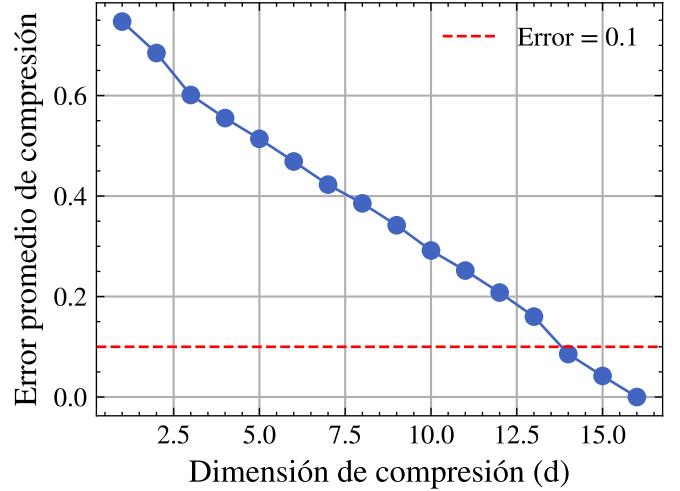
### 4.0.1 Imágenes reconstruidas con distintos valores de $d$



**Fig. 1:** Error vs dimension

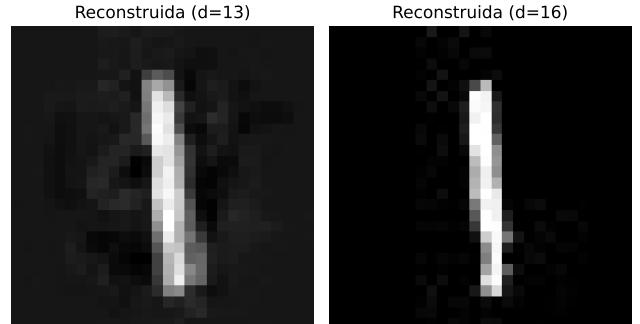
En la Figura 1, se presenta la imagen 00 reconstruida utilizando diferentes valores de  $d$ . Observamos que para valores bajos de  $d$ , la imagen presenta una reconstrucción aproximada en la que se distinguen las formas generales, pero falta detalle en las características finas. A medida que incrementamos  $d$ , la calidad de la imagen mejora, recuperando gradualmente los detalles hasta alcanzar la imagen exacta cuando  $d = 16$ . Sin embargo, a medida que se incrementan las dimensiones las mejoras marginales en la reconstrucción son cada vez menores, lo que coincide con los resultados de la Figura 2.

### 4.0.2 Error en función de la dimensión



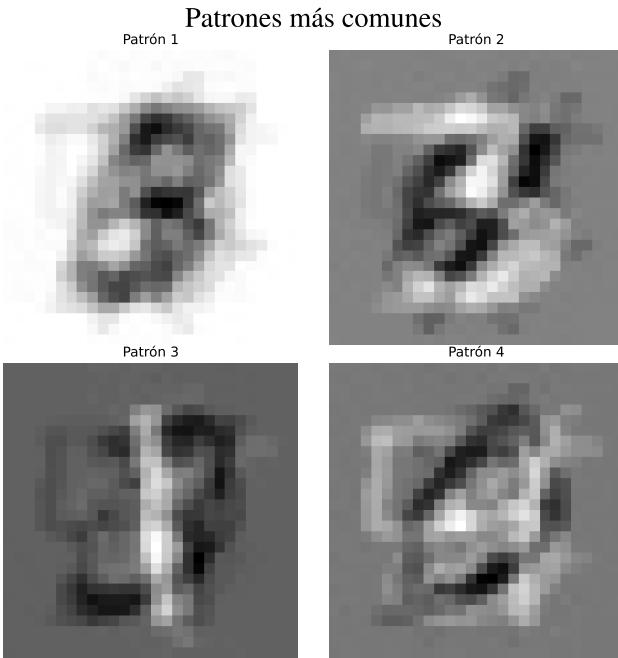
**Fig. 2:** Error vs dimensión

En la Figura 2, se observa el comportamiento del error de reconstrucción en función de la dimensión  $d$  utilizada en la representación reducida. Notamos que a medida que  $d$  aumenta, el error disminuye y en la dimensión 13 baja del umbral de 0.1 (10%). Lo que significa que al usar las primeras 13 dimensiones vamos a poder generar representaciones con una precisión mayor al (90%). En la figura 3 se puede ver como la reconstrucción es muy buena.

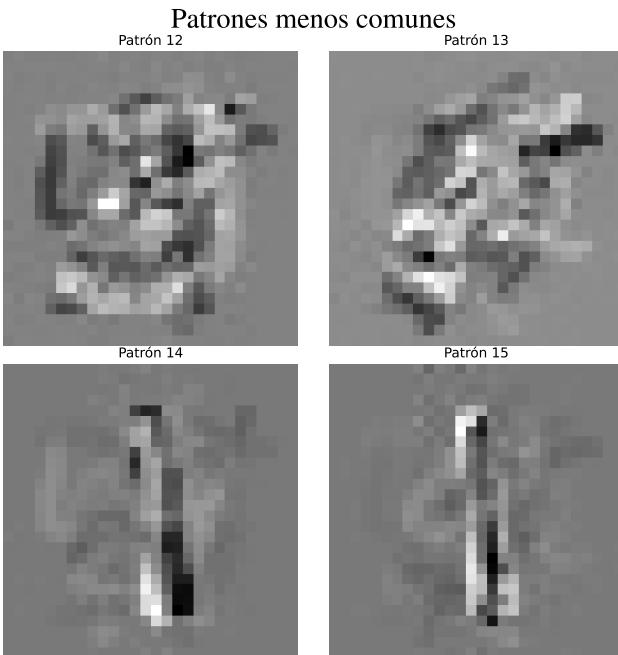


**Fig. 3:** Comparación  $d = 13$  y  $d = 16$

#### 4.0.3 Reconstrucción de patrones comunes



**Fig. 4:** Reconstrucción de patrones menos comunes



**Fig. 5:** Reconstrucción de patrones más comunes

Al analizar los autovectores correspondientes en la figura 4 a los valores singulares más grandes (es decir, los patrones más relevantes), observamos que estos reflejan formas discernibles que permiten identificar ciertos números presentes en las imágenes originales. Esto sugiere que estos patrones capturan características comunes y significativas del conjunto de datos, tales como contornos o trazos específicos de los números.

Por el contrario, los autovectores asociados a los valores singulares más pequeños (los patrones menos relevantes) en la figura 4 muestran estructuras caóticas y difíciles de interpretar. Este comportamiento es

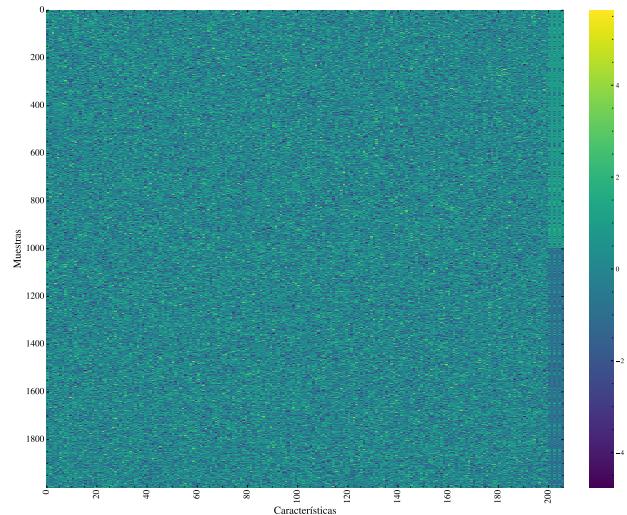
esperable, ya que estos patrones representan información de menor importancia que, en la práctica, podría corresponder al ruido o a variaciones mínimas en las imágenes.

#### 4.1 Análisis de similaridad y Cuadrados Mínimos

En esta sección, presentamos los resultados obtenidos del análisis del conjunto de datos  $X$  y del vector de respuestas  $y$ , aplicando técnicas de reducción de dimensionalidad y regresión lineal. A través de visualizaciones y métricas, exploramos la estructura interna de los datos y evaluamos la eficacia de los modelos predictivos en diferentes espacios de dimensión reducida.

##### 4.1.1 Análisis Exploratorio de la Matriz Original

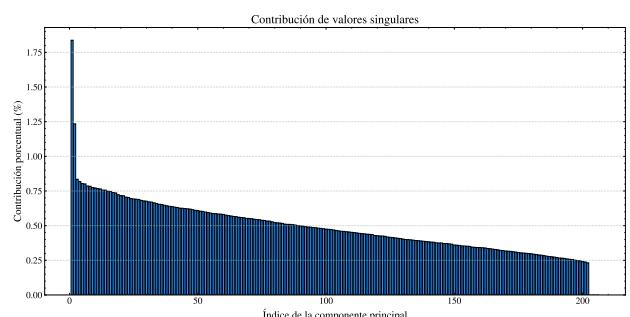
La Figura 6 muestra la representación gráfica de la matriz de datos original  $X$ . A simple vista, se puede observar que las primeras 200 columnas parecen contener principalmente ruido, mientras que las últimas 6 columnas muestran información relevante. Además, se aprecia una clara diferencia entre las primeras 1000 filas y las segundas 1000 filas, lo que sugiere la existencia de dos grupos distintos dentro de las muestras.



**Fig. 6:** Representación gráfica de la matriz de datos original  $X$ .

##### 4.1.2 Análisis de Valores Singulares

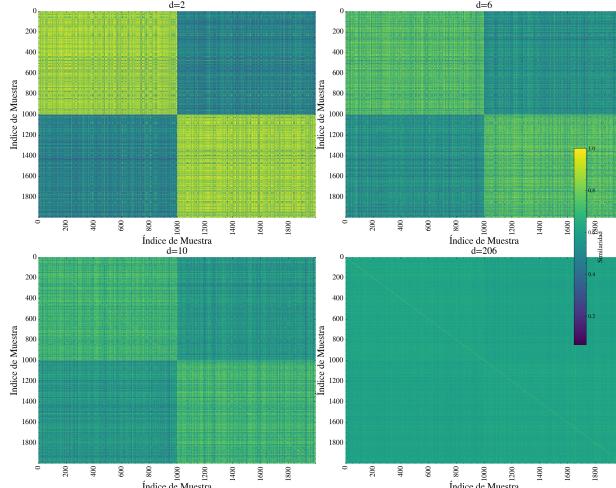
En la Figura 7, se presenta la varianza explicada por cada uno de los valores singulares de  $X$ . Como era de esperarse, los primeros valores singulares son los que más influyen en la variabilidad total de los datos. El primer valor singular explica casi un 2% de la varianza total, el segundo un poco más del 1.25%, y los restantes contribuyen con menos del 1%, decreciendo progresivamente.



**Fig. 7:** Porcentaje de varianza explicada por cada valor singular de  $X$ .

#### 4.1.3 Matrices de Similaridad en Diferentes Dimensiones

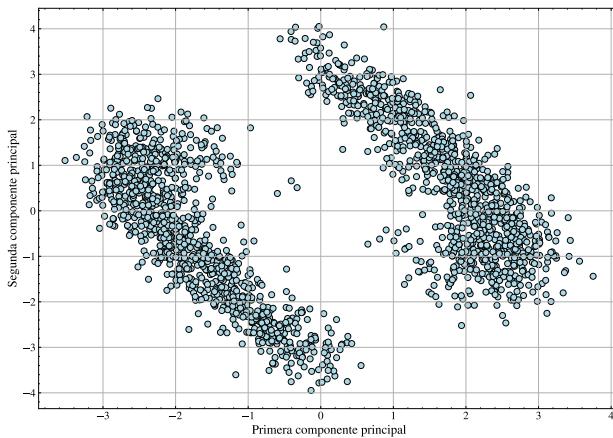
La Figura 8 muestra las matrices de similaridad para  $d = 2$ ,  $d = 6$ ,  $d = 10$  y  $d = p = 206$ . Principalmente para valores pequeños de  $d$ , se observan dos clusters claros: las primeras 1000 muestras son similares entre sí, al igual que las últimas 1000. A medida que  $d$  aumenta, se incluyen componentes principales con menor varianza explicada, lo que diluye el contraste entre los clusters al añadir ruido y variaciones menos relevantes a la proyección.



**Fig. 8:** Matrices de similaridad para  $d = 2, 6, 10$  y  $d = p = 206$ .

#### 4.1.4 Proyección en Dimensión 2

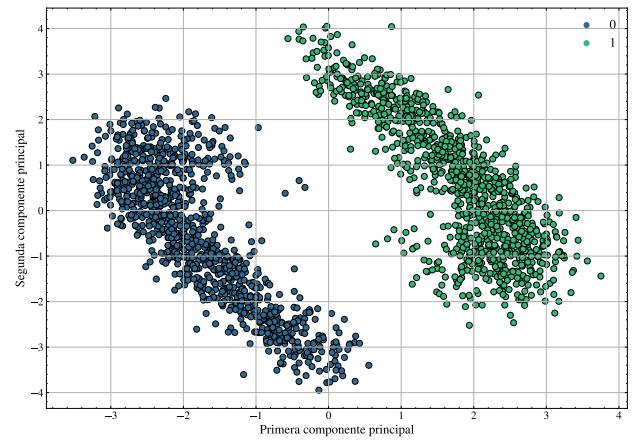
La Figura 9 muestra la proyección de las muestras en el espacio bidimensional formado por las dos primeras componentes principales. Se observa claramente la separación en dos clusters, lo que coincide con los análisis anteriores y sugiere que las primeras dos componentes capturan la mayor parte de la variabilidad relevante para distinguir los grupos.



**Fig. 9:** Proyección de las muestras en dimensión  $d = 2$ .

#### 4.1.5 Aplicación de K-Means

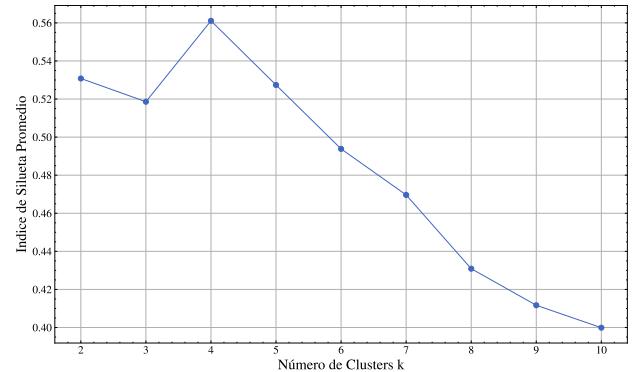
En la Figura 10, se muestra la misma proyección anterior pero con los clusters identificados por el algoritmo K-means para  $k = 2$ . Se puede observar que K-means detecta efectivamente la división de clusters, asignando correctamente las muestras a cada grupo. Esto valida la eficacia del algoritmo en este contexto y refuerza la presencia de dos agrupamientos naturales en los datos.



**Fig. 10:** Clusters identificados por K-means en la proyección de dimensión  $d = 2$ .

#### 4.1.6 Índice de Silueta

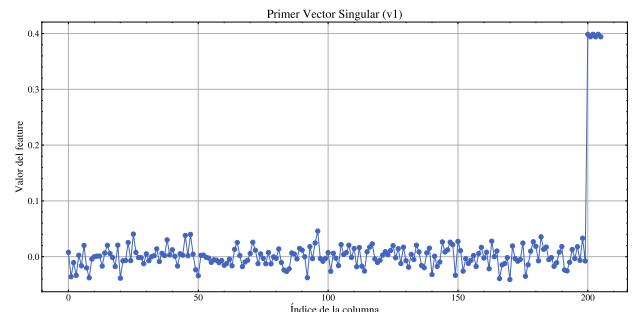
La Figura 11 presenta la evolución del índice de silueta en función de  $k$  (número de clusters) desde 2 hasta 10. Se observa que el valor óptimo del índice se alcanza en  $k = 4$  y luego disminuye. Sin embargo, dado que la separación más clara y significativa se encuentra en  $k = 2$ , esto sugiere que los clusters adicionales podrían estar capturando subdivisiones menos relevantes o ruido.



**Fig. 11:** Evolución del índice de silueta en función de  $k$ .

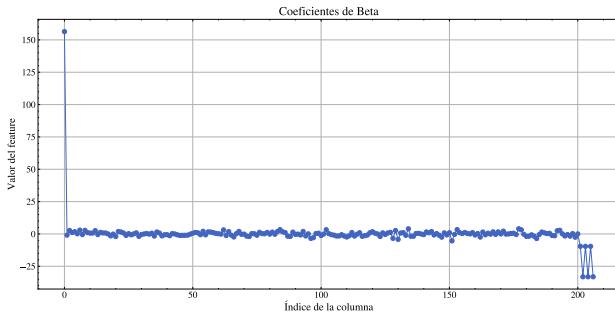
#### 4.1.7 Análisis de los Vectores Singulares y el Vector $\beta$

En la Figura 12, se muestra el primer vector singular de  $X$ , representando los pares de coordenada y valor. Se puede observar que las primeras 200 coordenadas tienen poco peso, mientras que las últimas 6 tienen valores significativamente mayores, lo que coincide con el análisis previo de que las últimas columnas contienen información más relevante.



**Fig. 12:** Primer vector singular de  $X$ : pares coordenada-valor.

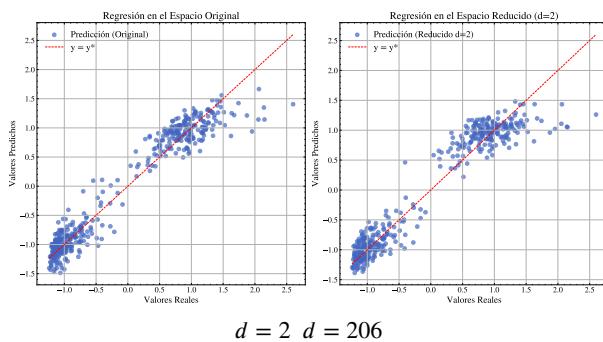
La Figura 13 muestra el vector  $\beta$  obtenido del ajuste de mínimos cuadrados, también en términos de coordenada y valor. Se observa un comportamiento inverso al del primer vector singular: las primeras 200 coordenadas tienen valores significativos, mientras que las últimas 6 tienen pesos menores. Esto puede deberse a que las primeras 200 características, aunque ruidosas, contribuyen a la predicción de  $y$  en el modelo lineal, posiblemente capturando variaciones que el vector singular no considera relevantes para la variabilidad total.



**Fig. 13:** Vector  $\beta$  del ajuste de mínimos cuadrados: pares coordenada-valor.

#### 4.1.8 Regresión Lineal y Análisis de Residuos

Comenzamos comparando las dimensiones  $d = 2$  y  $d = 206$  en el ajuste de regresión lineal. La Figura 14 muestra los gráficos de comparación entre los valores reales y predichos para ambas dimensiones. El gráfico de la izquierda corresponde a  $d = 2$  y el de la derecha a  $d = 206$ . Los puntos cercanos a la línea de identidad indican predicciones precisas. Aunque en ambos casos el modelo aproxima adecuadamente los valores de  $y$ , no es evidente a simple vista cuál dimensión ofrece un mejor ajuste.

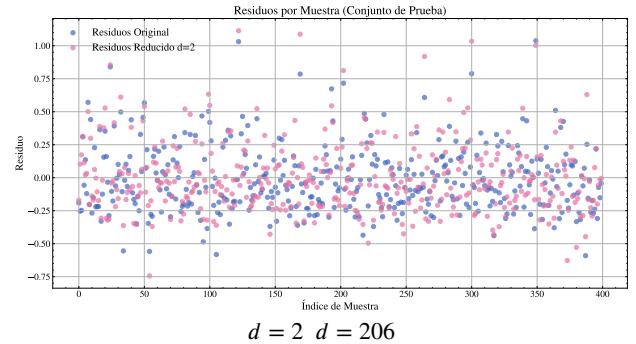


**Fig. 14:** Comparación entre valores reales y predichos para  $d = 2$  y  $d = 206$ .

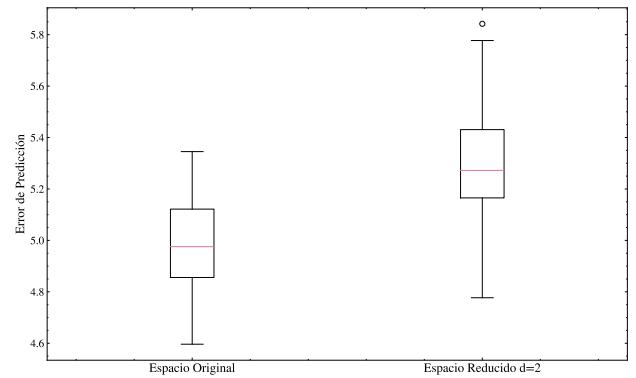
La Figura 15 presenta los gráficos de residuos para cada muestra en ambas dimensiones. Los residuos representan la diferencia entre los valores reales y los predichos. Los gráficos permiten identificar patrones de error o posibles sesgos en el modelo. Nuevamente, no es inmediatamente evidente cuál modelo es superior, lo que motiva un análisis más detallado de los errores.

#### 4.1.9 Análisis de Errores

En la Figura 16, se muestra un gráfico de barras comparando los errores promedio para el espacio reducido  $d = 2$  y el espacio original  $d = 206$ . Se observa que el error en el espacio original es menor, aunque la diferencia promedio es pequeña, aproximadamente 0.2. Esto indica que, si bien la reducción de dimensionalidad simplifica el modelo, puede conllevar una ligera pérdida en la precisión predictiva.



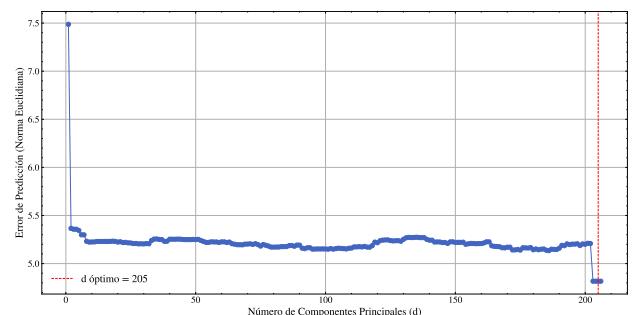
**Fig. 15:** Gráficos de residuos para  $d = 2$  y  $d = 206$ .



**Fig. 16:** Comparación de errores promedio entre  $d = 2$  y  $d = 206$ .

#### 4.1.10 Determinación de la Dimensionalidad Óptima

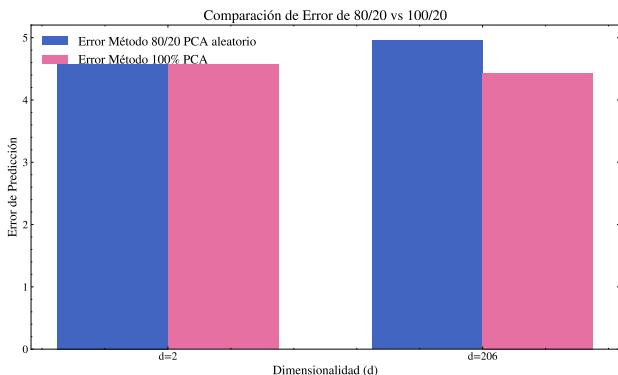
Para encontrar el valor óptimo de  $d$ , analizamos la evolución del error en función de  $d$  desde 2 hasta 206. La Figura 17 muestra esta relación. Se aprecia que el error se mantiene bastante estable para los primeros 200 valores de  $d$ , disminuyendo ligeramente en los últimos. El cambio no es drástico, pero el error mínimo se alcanza en  $d = 205$ , sugiriendo que incluir casi todas las componentes principales ofrece una mejora marginal en la predicción.



**Fig. 17:** Evolución del error de predicción en función de  $d$ .

#### 4.1.11 Comparación entre Métodos 80/20 y 100/20

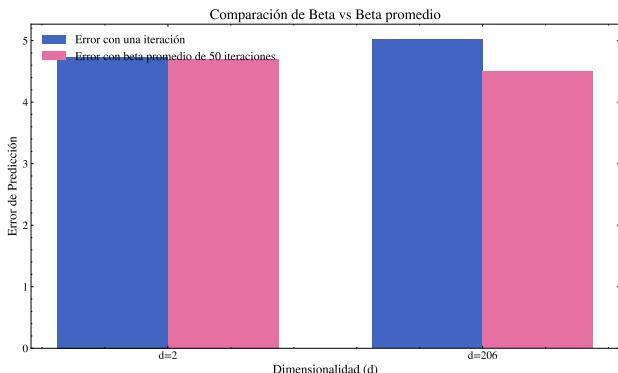
Realizamos el análisis comparativo de los métodos 80/20 y 100/20 para  $d = 2$  y  $d = 206$ . Los resultados se muestran en la Figura 18. Para  $d = 2$ , ambos métodos arrojan un error promedio de aproximadamente 4.5. Para  $d = 206$ , el método 100/20 obtiene un error ligeramente menor (4.5) en comparación con el método 80/20 (5). Esto podría deberse a que el uso de todo el conjunto de datos para el ajuste en el método 100/20 permite capturar mejor las relaciones, pero también puede aumentar el riesgo de sobreajuste, especialmente en alta dimensionalidad.



**Fig. 18:** Comparación de errores entre los métodos 80/20 y 100/20 para  $d = 2$  y  $d = 206$ .

#### 4.1.12 Análisis del Vector $\beta$ Promedio en Múltiples Iteraciones

Para evaluar la estabilidad del vector  $\beta$ , comparamos el error obtenido al usar el  $\beta$  de una sola iteración con el promedio de  $\beta$  sobre 50 iteraciones, utilizando el método 80/20. La Figura 19 muestra los resultados para  $d = 2$  y  $d = 206$ . Se observa que para  $d = 2$  el error es prácticamente igual en ambos casos, mientras que para  $d = 206$  el error obtenido al promediar  $\beta$  sobre 50 iteraciones es ligeramente menor que el error de una sola iteración. Esto sugiere que en alta dimensionalidad, el promedio de  $\beta$  ayuda a mitigar la sensibilidad del modelo a las particularidades del conjunto de entrenamiento, mejorando así el desempeño predictivo.



**Fig. 19:** Comparación de errores usando  $\beta$  de una iteración y el promedio de 50 iteraciones.

## 5 Conclusiones

En este trabajo, exploramos la aplicación de técnicas de reducción de dimensionalidad, específicamente la Descomposición en Valores Singulares (SVD) y el Análisis de Componentes Principales (PCA), en diferentes contextos de datos de alta dimensionalidad. A través de estos métodos, abordamos tanto la compresión y análisis de imágenes como el estudio de conjuntos de datos numéricos con el objetivo de identificar patrones internos y mejorar la capacidad predictiva de modelos de regresión.

En el caso del análisis de imágenes, la representación comprimida mediante SVD nos permitió reducir significativamente la dimensionalidad de las imágenes mientras preservábamos la mayor parte de la información visual relevante. Las imágenes reconstruidas con un bajo número de componentes principales ( $d$ ) capturaron las características generales, aunque con pérdida de detalles finos. A medida que  $d$  aumentó, la calidad de la reconstrucción mejoró. El análisis de los patrones visuales más relevantes mostró que los autovectores

principales representan características comunes del conjunto de datos, como contornos y trazos significativos. Esto confirma que la técnica SVD no solo permite compresión, sino también la extracción de información visual interpretativa. Los valores singulares más pequeños y sus correspondientes autovectores mostraron patrones caóticos y de menor relevancia, destacando su relación con el ruido o variaciones menores en las imágenes originales.

En el análisis del conjunto de datos numéricos, los métodos aplicados, como el cálculo de matrices de similaridad y la aplicación de  $K$ -means, resultaron efectivos para explorar la naturaleza y estructura interna de los datos. La utilización de PCA permitió distinguir de manera más clara entre las características relevantes y el ruido presente en el conjunto de datos, facilitando la identificación de patrones y agrupamientos. En cuanto a la regresión mediante mínimos cuadrados, se demostró que es un método efectivo para predecir la variable  $y$  a partir de  $X$ . Aunque el modelo de regresión con la dimensión completa ( $d = 206$ ) ofreció un rendimiento ligeramente mejor que con  $d = 2$ , la diferencia en precisión fue marginal. Esto sugiere que, para conjuntos de datos aún más grandes donde el costo computacional es crítico, es recomendable reducir la dimensionalidad. La reducción de dimensiones permite disminuir significativamente el tiempo de cómputo y los recursos necesarios, sin comprometer de manera sustancial la precisión del modelo predictivo.

En resumen, las técnicas de SVD y PCA demostraron ser efectivas tanto para la compresión y análisis de imágenes como para el procesamiento y análisis de datos numéricos de alta dimensionalidad. Lograron un balance entre reducción dimensional y preservación de información, facilitando la extracción de características relevantes y mejorando la interpretabilidad de los datos. Estos resultados refuerzan la utilidad práctica de estas técnicas en diversas aplicaciones, donde es esencial manejar grandes volúmenes de datos de manera eficiente sin perder información significativa. Balancear la complejidad del modelo y su capacidad predictiva mediante técnicas de reducción de dimensionalidad es, por tanto, una estrategia eficaz en el procesamiento y análisis de datos de alta dimensionalidad.