

# TP N°3: Métodos Numéricos y Optimización

Zoe Velazquez Zorzi y Manuela Gomez Pazos

November 2023

## Abstract

En este trabajo realizamos un exhaustivo estudio enfocado en dos conjuntos de datos: uno que consiste en imágenes y otro que representa mediciones obtenidas a través de un sensor. El objetivo fue realizar una compresión de imágenes y una reducción de dimensionalidad para las mediciones del sensor, utilizando en el primer caso la técnica de Descomposición de Valores Singulares (SVD) con el propósito de obtener representaciones más compactas y eficientes, y en el segundo, la técnica de Cuadrados Mínimos. Los resultados y conclusiones revelaron el impacto de la compresión y reducción de dimensionalidad en la similitud entre muestras graficadas con mapas de calor, la importancia de las dimensiones respecto a los valores singulares, y la influencia del valor de  $d$  en la predicción y modelado de los datos.

## 1 Introducción

En nuestra investigación, hemos explorado dos escenarios distintos que involucran la aplicación de técnicas de análisis y reducción de dimensionalidad en conjuntos de datos diversos. Inicialmente, nos sumergimos en el ámbito de la compresión de imágenes, donde hemos trabajado con un conjunto de 16 imágenes representadas como matrices, cada una con números en tono blanco sobre fondos oscuros. En este contexto, llevamos a cabo un estudio individual de cada imagen y luego analizamos su representación matricial de forma conjunta.

La exploración colectiva de las imágenes nos permitió indagar acerca de la información contenida en los autovectores asociados a los valores singulares predominantes y mínimos de las respectivas matrices. Posteriormente, al adentrarnos en el análisis singular, nuestro objetivo fue reducir la matriz de cada imagen con el propósito de limitar el error entre la imagen original y su contraparte comprimida, asegurando que este error no superara el 10% bajo la norma de Frobenius. De este modo, se buscó identificar la dimensión mínima que permitiera esta reducción sin exceder el umbral de error especificado.

En la segunda parte de nuestro estudio, nos centramos en un conjunto de datos representado en formato matricial, denominado  $\mathbb{X}$ . Este conjunto consta de una serie de muestras  $(x_1, x_2, \dots, x_i, \dots, x_n)$  medidas a través de un sensor, donde las muestras no se distribuyen uniformemente en el espacio y presentan clusters con alta similitud entre sí.

Dado que se observó que algunas dimensiones podrían ser más ruidosas que otras en las muestras, se consideró ventajoso trabajar en un espacio de menor dimensionalidad, denotado por  $d$ . Esto llevó a realizar una descomposición de  $\mathbb{X}$  en sus valores singulares y a reducir la dimensión de esta representación, proyectando los vectores  $x$  al nuevo espacio reducido, es decir,  $\mathbb{V}_d^T x$ .

## 2 Métodos Utilizados

A lo largo del trabajo fueron empleados los siguientes métodos.

### 2.1 Descomposición en Valores Singulares (SVD)

La descomposición en valores singulares, mejor conocido como SVD, es un método que se basa en la descomposición de una matriz en un producto de tres matrices calculadas en base a sus valores singulares y los autovectores asociados a los mismos. Si se toma como ejemplo una matriz  $A$  perteneciente a  $\mathbb{R}^{n \times m}$  y se la descompone con SVD,  $A$  queda de la siguiente manera:

$$A = USV^T, \quad (1)$$

donde  $S$  pertenece a  $\mathbb{R}^{n \times m}$  es una matriz diagonal que contiene los valores singulares de  $A$ ,  $U$  pertenece a  $\mathbb{R}^{n \times n}$  y  $V^T$  pertenece a  $\mathbb{R}^{m \times m}$  matrices unitarias contenedoras de los vectores singulares asociados a los valores singulares. La diagonal de  $S$  contiene los valores singulares ordenados de mayor a menor magnitud, logrando así mantener

la importancia de los componentes principales a los que se asocian. Estos componentes son la proyección de las columnas de la matriz original sobre  $V^T$ . Guardan información sobre la variabilidad de los datos de A, mientras mayor sea el valor singular al que están asociados, más información tienen. Al querer comprimir la matriz de forma que queden los componentes más importantes, sucede la reducción del tamaño cada parte de la descomposición, así quedan  $U_d \in \mathbb{R}^{m \times d}$ ,  $S_d \in \mathbb{R}^{d \times d}$  y  $V_d^T x \in \mathbb{R}^{d \times d}$ , que generan  $A_d \in \mathbb{R}^{m \times n}$ , matriz aproximada de A, pero con una cantidad reducida de datos.

## 2.2 Norma de Frobenius

Esta es una medida de magnitud para una matriz, definida como la raíz cuadrada de la suma de los cuadrados de todos los elementos de la matriz. Para una matriz  $A \in \mathbb{R}^{m \times n}$ , la norma de Frobenius se calcula de la siguiente manera:

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

Su utilización permite evaluar la calidad de la aproximación de una matriz al usar SVD.

## 2.3 Similitud

Es posible medir la similitud entre un par de muestras  $x_i, x_j$  utilizando una función no-lineal de su distancia euclidiana con la siguiente fórmula:

$$K_1(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (3)$$

para algún valor de  $\sigma$ .

## 2.4 Cuadrados Mínimos

Es una técnica ampliamente utilizada para ajustar una función a un conjunto de datos experimentales de tal manera que minimice la diferencia entre la función ajustada y los datos observados. La regresión lineal específicamente es un método estadístico ampliamente utilizado para analizar y modelar la relación entre dos o más variables. Su enfoque principal consiste en establecer una relación lineal entre una variable independiente y una variable dependiente, que es la respuesta que deseamos modelar. Busca encontrar la mejor línea recta que se ajuste a los datos observados, de manera que minimice la discrepancia entre los valores reales y los valores predichos por el modelo. Esta línea recta se caracteriza mediante una ecuación lineal, donde los coeficientes representan la pendiente y la intersección en el eje vertical.

$$y = \beta_0 + \beta_1 x + \varepsilon, \quad (4)$$

donde  $y$  es la variable dependiente,  $x$  es la variable independiente,  $\beta_0$  es la intersección en el eje vertical,  $\beta_1$  es la pendiente de la línea, que representa la relación entre  $x$  e  $y$ ,  $\varepsilon$  es el error residual.

## 3 Otras Herramientas Utilizadas

Otras herramientas utilizadas en el desarrollo del trabajo fueron las siguientes.

### 3.1 Función de similitud del coseno

Es una medida común utilizada para comparar la similitud entre dos vectores o muestras en un espacio vectorial. En términos matemáticos, la similitud del coseno entre dos vectores A y B se calcula de la siguiente manera:

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \cdot \|B\|} \quad (5)$$

donde  $A \cdot B$  representa el producto escalar de los dos vectores A y B. A su vez  $\|A\|, \|B\|$  son las normas (longitudes) de los vectores A y B, respectivamente.

La similitud del coseno mide el coseno del ángulo entre los dos vectores. Cuanto más cercano sea el valor de la similitud del coseno a 1, más similares son estos. Un valor de 1 indica que los vectores son idénticos, mientras que un valor de -1 indica que son opuestos en dirección. Un valor de 0 indica que son ortogonales o no tienen similitud.

## 4 Desarrollo experimental y aplicaciones

El estudio realizado se centró en dos ejercicios distintos, el primero de ellos dirigido a la compresión de imágenes. Este proceso se llevó a cabo con el propósito de reducir la dimensionalidad de las imágenes, permitiendo una representación más eficiente. Inicialmente, se transformaron las imágenes en matrices, seguido por la aplicación de la SVD sobre la matriz original. A partir de este análisis, se realizaron reconstrucciones considerando tanto las primeras como las últimas dimensiones de la descomposición obtenida. Posteriormente, se visualizaron las imágenes originales junto con sus reconstrucciones basadas en las primeras y últimas  $d$  columnas, proporcionando un análisis comparativo entre las mismas. La exploración continuó con un análisis detallado de la matriz de valores singulares (S), la matriz  $VT$  y la matriz  $U$ , tanto en su forma original como en su versión reconstruida con una dimensionalidad reducida. Se representaron los mapas de calor para cada una de estas matrices, permitiendo visualizar y comparar las diferencias entre las originales y sus contrapartes de baja dimensión. Otro aspecto clave del análisis se enfocó en la medición de la similitud entre las imágenes a distintos valores de  $d$  (dimensión) utilizando la función de similitud del coseno. La evaluación detallada de la similaridad entre pares de imágenes permitió una comprensión más profunda de cómo la reducción de la dimensionalidad afecta la relación entre las imágenes. Finalmente, se abordó la determinación del número mínimo de dimensiones ( $d$ ) para comprimir una imagen, manteniendo un error por debajo del umbral del 10% bajo la norma de Frobenius. Este proceso involucró un cálculo iterativo para encontrar la dimensión  $d$  que minimiza el error de compresión, proporcionando una comprensión más precisa del compromiso entre la reducción de la dimensionalidad y la calidad de la imagen comprimida.

El segundo estudio se centró en la reducción de la dimensionalidad y la aplicación de Cuadrados Mínimos a un conjunto de datos recopilados a través de un sensor. El dataset  $X$  se carga y se analiza para comprender la distribución de las muestras. Se supone que la distribución de las muestras no es uniforme, lo que puede dar lugar a la identificación de clusters o grupos con alta similitud entre ellos. Para reducir la alta dimensionalidad de los datos, implementamos la Descomposición en Valores Singulares (SVD) en  $X$ , reduciendo la dimensión del espacio a diferentes valores de  $d$ , como 2, 4, 6, 20 y la dimensión original  $p$ . Realizamos un mapeo de las proyecciones en el nuevo espacio reducido, visualizando así la distribución de las muestras en un espacio de menor dimensión. Esto nos permitió observar la agrupación de muestras y la relación entre ellas en los diferentes espacios dimensionales. Además, evaluamos la similitud entre las muestras en el espacio original y en el espacio reducido. Para esto último empleamos la función de similitud coseno para comparar las muestras a distintos valores de  $d$ . Estos análisis comparativos brindan una visión general de cómo la reducción de la dimensionalidad afecta la relación entre las muestras y su distribución en el espacio. Otro punto de interés es identificar las dimensiones más representativas en comparación con las dimensiones reducidas obtenidas por SVD. Se examina la importancia de las dimensiones originales en relación con las dimensiones reducidas para cada valor de  $d$ , permitiendo un análisis más profundo de cómo las dimensiones originales se reflejan en el espacio de menor dimensión. Finalmente, realizamos un análisis basado en Cuadrados Mínimos, intentando encontrar el mejor valor de  $d$  que minimizara el error de predicción. Se ajustó un modelo lineal a los datos y se realizaron predicciones. El proceso incluyó la selección de la dimensión  $d$  que optimizara el error de predicción, proporcionando una mejor comprensión de cómo la reducción de la dimensionalidad afectaba la capacidad predictiva del modelo. Cada paso de ambas experimentaciones se apoyó en gráficos, visualizaciones y un análisis detallado para respaldar y fundamentar los hallazgos y conclusiones obtenidas en el desarrollo del estudio.

## 5 Análisis de los resultados

### 5.1 Compresión de Imágenes

Para este estudio, utilizamos un conjunto de imágenes en escala de grises. Inicialmente, aplicamos SVD a las matrices de las imágenes, con el fin de obtener sus componentes singulares. Luego, llevamos a cabo distintos experimentos con el propósito de evaluar la calidad de la reconstrucción de las imágenes originales a partir de su representación en dimensiones reducidas. Los experimentos presentaron los siguientes hallazgos clave: En primer lugar, observamos que la calidad de las imágenes reconstruidas variaba significativamente con la cantidad de componentes singulares retenidos. Como podemos ver en la figura 1, las reconstrucciones utilizando un menor número de componentes presentaban una pérdida notable en los detalles finos, mientras que las reconstrucciones con un mayor número de componentes conservaban una mayor fidelidad a las imágenes originales.

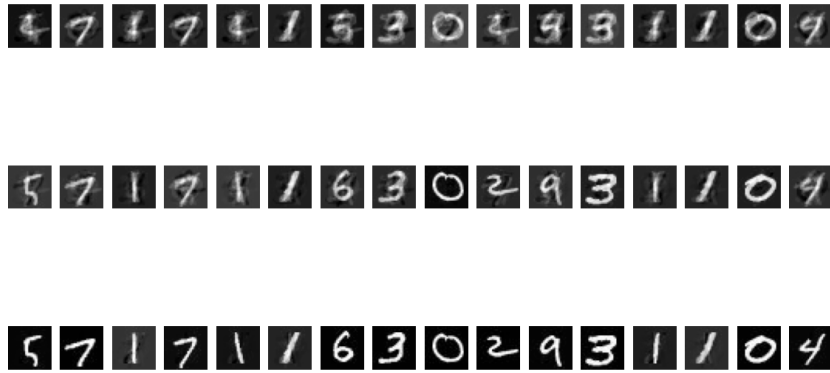


Figure 1: Reconstrucciones de las imágenes originales utilizando las  $d$  primeras dimensiones, con  $d$  entre 5, 10 y 15 (de arriba a abajo).

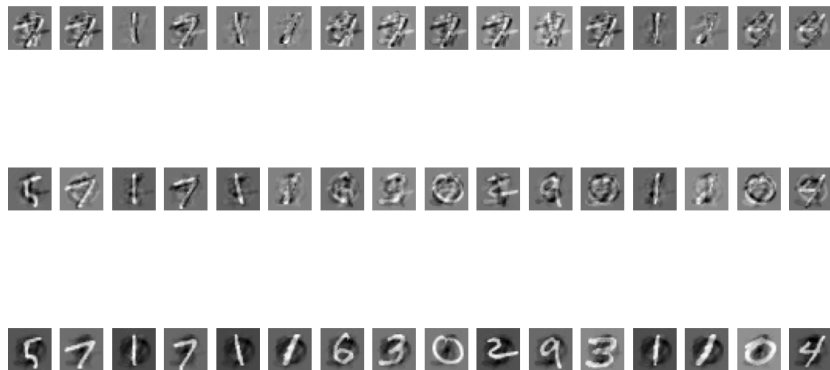


Figure 2: Reconstrucciones de las imágenes originales utilizando las  $d$  últimas dimensiones, con  $d$  entre 5, 10 y 15 (de arriba a abajo).

Siguiendo este hilo realizamos comparaciones entre las reconstrucciones basadas en las primeras y últimas columnas de la matriz SVD. Se encontró que las reconstrucciones a partir de las primeras columnas (figura 1) mostraban una mayor fidelidad a las imágenes originales en comparación con las últimas columnas (figura 2), lo que nos indica que estas últimas no capturaban de manera adecuada los detalles esenciales de las imágenes. Paralelamente decidimos realizar mapas de calor de las matrices de la descomposición SVD tanto de una imagen original, como de su versión reconstruida

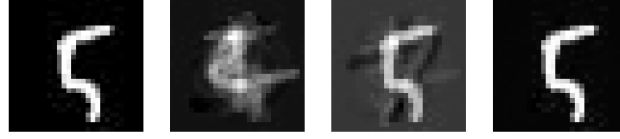


Figure 3: Comparación entre la imagen original y sus respectivas reconstrucciones utilizando las primeras  $d$  componentes principales ( $d=5, 10$  y  $15$  respectivamente).

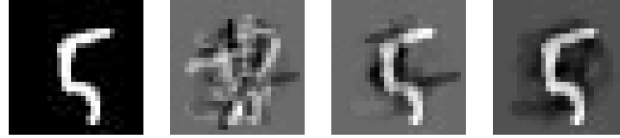


Figure 4: Comparación entre la imagen original y sus respectivas reconstrucciones utilizando las últimas  $d$  componentes principales ( $d=5, 10$  y  $15$  respectivamente).

En las figuras 3 y 4, podemos profundizar y consolidar las observaciones obtenidas en las figuras 1 y 2. Al considerar un valor pequeño de  $d$  para los primeros componentes, se evidencian zonas de la imagen reconstruida que aparecen borrosas y amalgamadas. A medida que se incrementa el valor de  $d$ , se comienza a percibir, de forma progresiva, la aparición de la imagen original. Esta se va nítidamente delineando hasta lograr una representación prácticamente idéntica a la imagen original. Por otro lado, al utilizar los últimos  $d$  en lugar de los primeros, se observa la formación de la imagen original, aunque con mayor error y una notoria pérdida de detalles. A medida que se reduce la cantidad de componentes, se genera una mayor distancia con la imagen original y disminuye la fidelidad de la representación, mostrando una reconstrucción con menos precisión y menor nivel de detalle. Continuando con este análisis, realizamos uno de similitud entre pares de imágenes proyectadas en espacios de baja dimensión. Calculamos la similitud entre cada imagen utilizando una métrica de coseno y generando una matriz de similitud. En estas matrices los tonos más claros representan valores más bajos y los más oscuros corresponden a valores más altos de similitud, que es efectivamente lo que podemos ver en la figura 5. En este gráfico vemos como con un  $d$  pequeño, la matriz presenta tonos mucho más claros (amarillos), y mientras más aumentamos el valor de  $d$ , más oscura se hace la matriz, certificando así una mayor similitud. La diagonal de estas matrices representa la similitud de cada imagen consigo misma y es por eso que son completamente amarillas, porque cada imagen debería ser idéntica a sí misma.

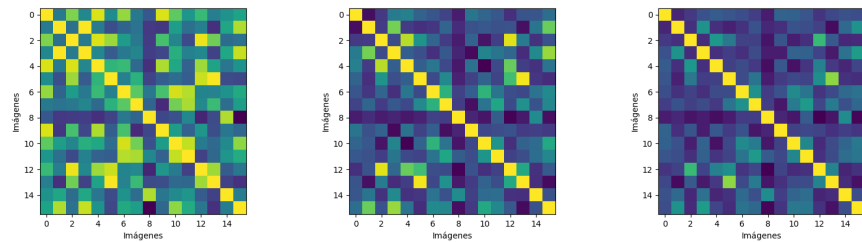


Figure 5: Representación de la similitud par a par entre imágenes a través de distintos espacios de baja dimensión  $d$ , generados a partir de SVD. Cada subfigura muestra la matriz de similitud entre imágenes para un valor específico de  $d$  (5, 10 y 25 en orden).

Para este experimento específico, optamos por un análisis adicional: seleccionamos la primera imagen del conjunto de 16 imágenes, aplicamos un procedimiento de compresión utilizando la técnica SVD, y nos propusimos encontrar un valor óptimo de  $d$ . Este valor permitiría una compresión con un error menor al 10% en comparación con la imagen original. Una vez determinado el valor óptimo de  $d$  para la primera imagen, empleamos este mismo valor para comprimir las demás imágenes del conjunto. Al examinar el gráfico resultante, observamos que la imagen para la cual se obtuvo el  $d$  óptimo se reconstruyó de manera precisa. Sin embargo, al observar las demás imágenes, notamos similitudes con sus respectivas originales, aunque no logran alcanzar la misma fidelidad encontrada en la reconstrucción de la primera imagen.

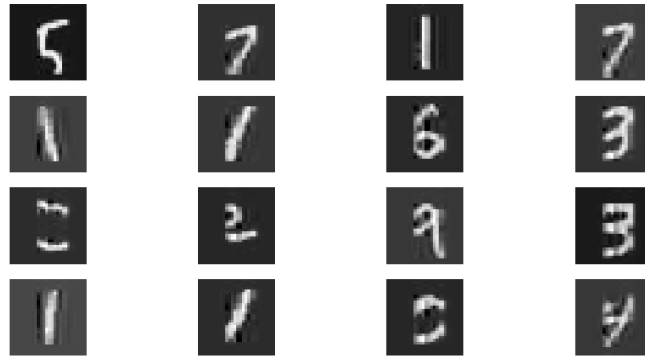


Figure 6: Representación visual de todas las imágenes comprimidas a una dimensión óptima  $d$  con un error inferior al 10%.

## 5.2 Reducción de la dimensionalidad y Cuadrados Mínimos

En esta sección, analizamos un conjunto de datos contenidos en el archivo '.csv' proporcionado por la cátedra. Este conjunto de datos fue cargado y almacenado en una matriz llamada  $X$ . A continuación, realizamos una reducción de dimensionalidad para diferentes valores de  $d$  utilizando PCA. Los valores de  $d$  considerados fueron 2, 4, 6, 20 y la dimensionalidad original ' $p = 107$ '. Descompusimos  $X$  en sus valores singulares (SVD), y luego seleccionamos las  $d$  componentes principales para obtener una representación reducida del conjunto de datos.

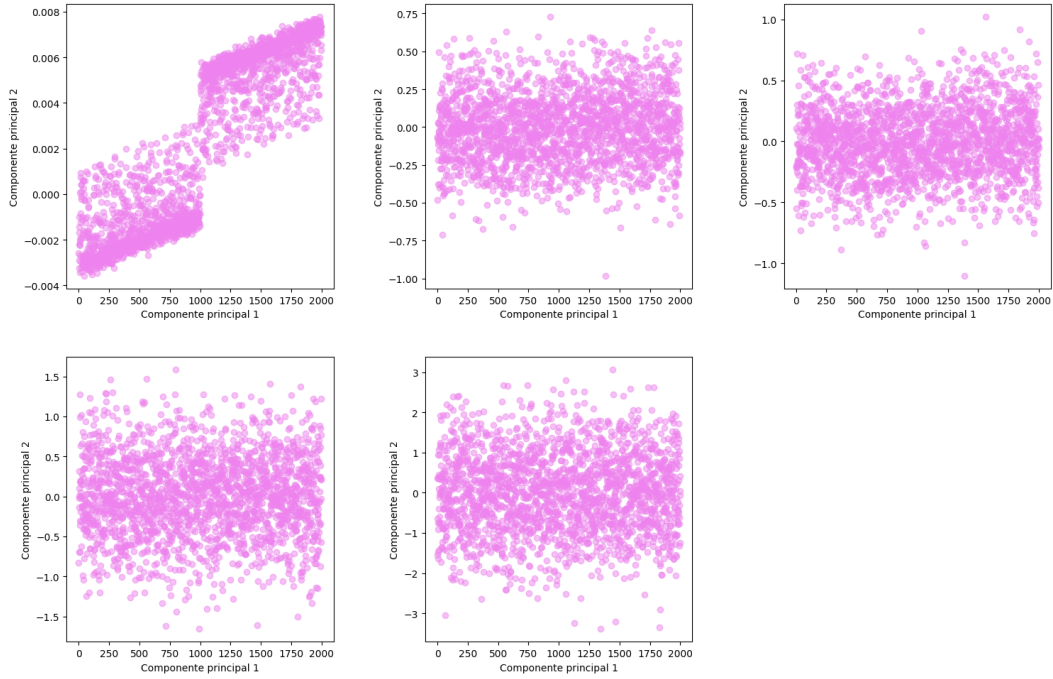


Figure 7: Proyección de las muestras en el espacio de dimensiones reducidas.

Como se puede observar en la figura 8, la reducción de la dimensionalidad influye en la distribución de las muestras en el espacio de características. La comparación de similitudes en espacios de diferentes dimensiones demuestra como, un valor bajo de  $d$  hace que se corra el riesgo de perder información importante. A su vez, a mayor  $d$ , se puede notar una disminución en la dispersión de puntos. Por lo tanto en la elección de este parámetro se debe tomar en cuenta la obtención de una buena visualización de los datos y una correcta retención de la información relevante.

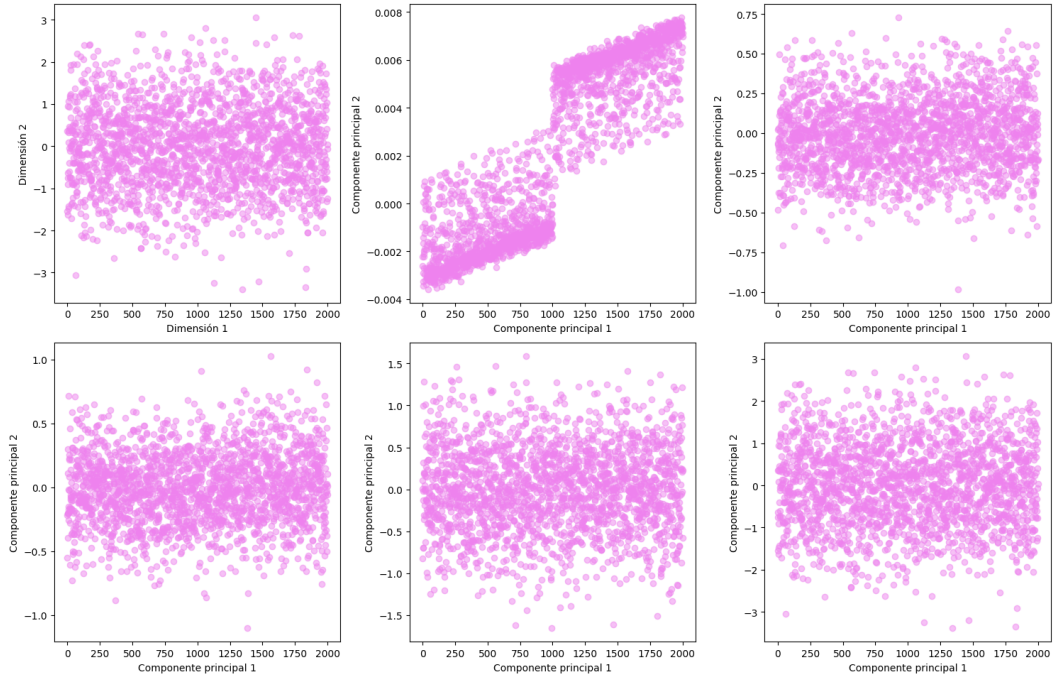


Figure 8: Comparación de similitud en espacios de dimensiones original y reducidas.

Calculamos e identificamos las dimensiones más importantes en relación con las dimensiones reducidas obtenidas

por SVD, ordenandolas de manera descendente según su importancia, para así optimizar futuros usos, como puede observarse en la figura 9.

```
Dimensiones más importantes para d=2 según SVD:
Dimensión 1 - Importancia: 0.03318130501746835
Dimensión 0 - Importancia: 0.009370097621454156

Dimensiones más importantes para d=4 según SVD:
Dimensión 3 - Importancia: 0.0750299411834454
Dimensión 2 - Importancia: 0.05465034531595771
Dimensión 1 - Importancia: 0.03318130501746835
Dimensión 0 - Importancia: 0.009370097621454156

Dimensiones más importantes para d=6 según SVD:
Dimensión 4 - Importancia: 0.07932231560814607
Dimensión 3 - Importancia: 0.0750299411834454
Dimensión 5 - Importancia: 0.07352339329116567
Dimensión 2 - Importancia: 0.05465034531595771
Dimensión 1 - Importancia: 0.03318130501746835

Dimensiones más importantes para d=20 según SVD:
Dimensión 8 - Importancia: 0.07935916677125614
Dimensión 4 - Importancia: 0.07932231560814607
Dimensión 19 - Importancia: 0.07850405734628153
Dimensión 7 - Importancia: 0.07833049052675738
Dimensión 15 - Importancia: 0.07824110723603037

Dimensiones más importantes para d=107 según SVD:
Dimensión 45 - Importancia: 0.07998910278166065
Dimensión 102 - Importancia: 0.07983895420309388
Dimensión 8 - Importancia: 0.07935916677125614
Dimensión 4 - Importancia: 0.07932231560814607
Dimensión 64 - Importancia: 0.07923088800538583
```

Figure 9: Importancia de las dimensiones

Como observamos en la figura 10, que representa la variación del error de predicción en función de la dimensionalidad  $d$ , a mayor dimensionalidad se logra minimizar significativamente la cantidad de error obtenida. Esto respalda lo mencionado anteriormente acerca de como a pesar de que a primera vista, un  $d$  más bajo parece más efectivo, este podría hacernos perder información y tener un mayor error. Teniendo en cuenta esto, se llega a la resolución de que el  $d$  que minimiza mejor el error es  $d = 20$ .

Para confirmar la elección de  $d$ , realizamos el gráfico correspondiente a la figura 10. En el mismo podemos observar la distribución de los errores de predicción para cada  $d$  en la lista de dimensiones, confirmando así un mayor error en la predicción para  $ds$  más pequeños y un decrecimiento explosivo con el aumento de  $d$ .

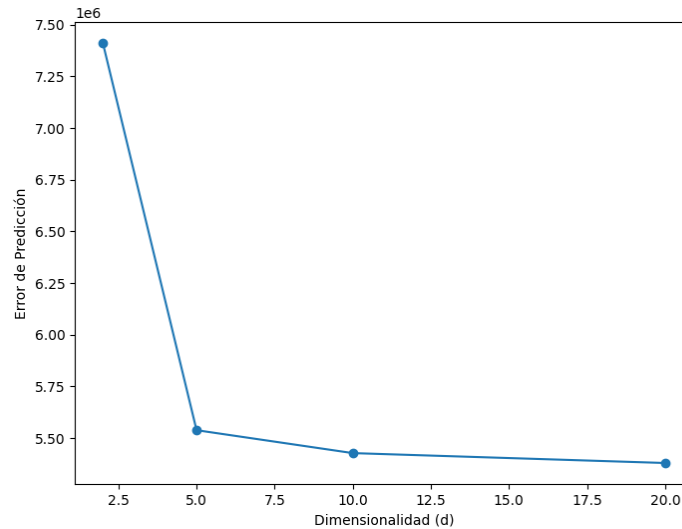


Figure 10: Comparación de error de predicción vs dimensionalidad.



Además realizamos un gráfico de barras donde cada barra representa el valor de los coeficientes  $\beta$  para diferentes atributos, para cada dimensión  $d$ , como se observa en la figura 11. Los coeficientes  $\beta$  se muestran para todos los atributos, y las barras de diferentes colores representan diferentes valores de  $d$ . Este gráfico nos aporta una idea de qué atributos tienen más peso en las predicciones a medida que cambia la dimensionalidad. Esto se debe a que atributos con coeficientes  $\beta$  más grandes tienen un mayor impacto en las predicciones.

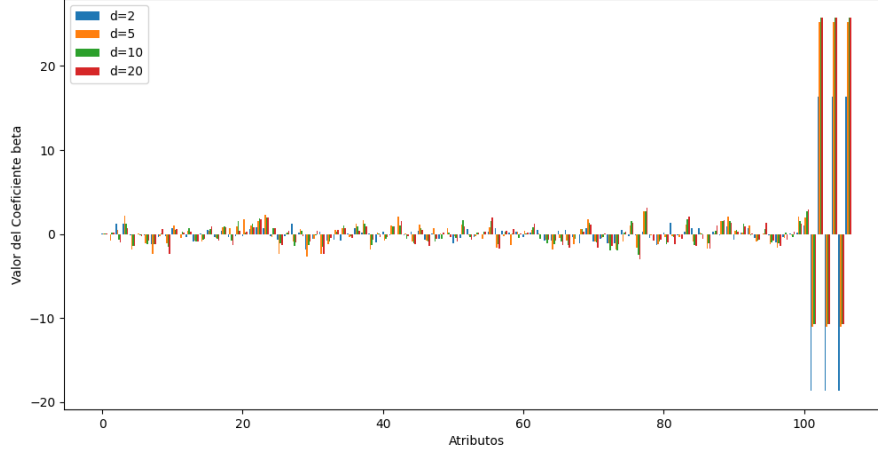


Figure 11: Coeficientes beta para distintas dimensiones.

Por último, en la figura 12 representamos un gráfico de densidad de puntos, también conocido como Kernel Density Estimation (KDE). Este nos aportó una representación visual de la distribución de los puntos en un espacio bidimensional. Es sumamente útil para comprender cómo los datos se distribuyen en diferentes dimensiones después de realizar la reducción SVD. Como es posible apreciar en la figura, en los tres valores distintos de  $d$  se pueden visualizar *clusters* en el espacio reducido. Por otra parte, la figura nos permite comparar visualmente como la reducción de dimensionalidad afecta la distribución de los datos en el espacio bidimensional y cómo esta distribución cambia al modificar el valor de  $d$ . Como último parámetro a tener en cuenta, estos gráficos nos ayudan a tener un mejor entendimiento de la disminución de la varianza. Esto se debe a que nos permiten comprender si la dimensión en la dimensionalidad está manteniendo la estructura de datos original o si se está perdiendo información importante.

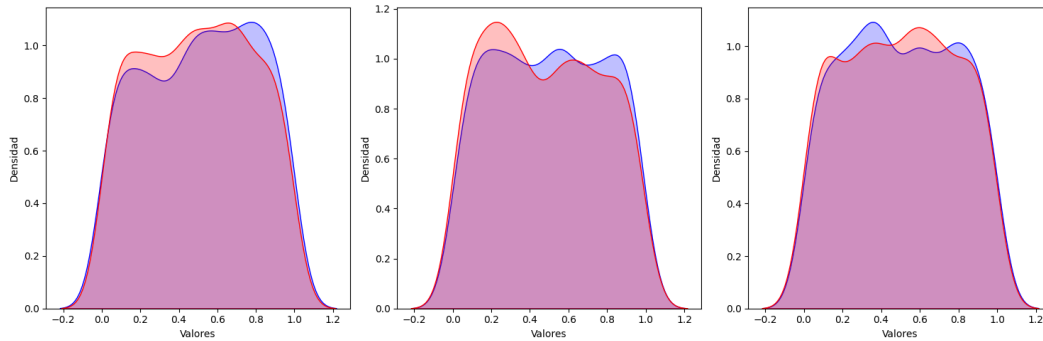


Figure 12: Gráfico de densidad para  $d=2$ ,  $d=4$  y  $d=6$  respectivamente.

Todas las figuras mencionadas anteriormente nos permitieron analizar y determinar cuál es el valor de  $d$  más óptimo, tomando en cuenta los distintos factores a considerar a la hora de preservar información mientras se buscaba el menor error posible.

## 6 Conclusiones

El uso de la Descomposición en Valores Singulares (SVD) ha demostrado ser una herramienta versátil y poderosa en la reducción de dimensionalidad, análisis de imágenes y modelos de datos. Con respecto a la reducción de dimensionalidad, pudimos analizar el impacto de  $d$  en la representación de imágenes, y llegamos a la conclusión de que es crítico a la hora de realizar estas reconstrucciones a partir de una matriz SVD reducida. Observamos que al utilizar un número mayor de componentes principales ( $d$ ), logramos una representación más fiel de las imágenes originales en el espacio de baja dimension; pero no solo eso, sino que también llegamos a la conclusión de que la reproducción más fiel se logra utilizand las primeras componentes principales y no las últimas, ya que estas con generan imagenes tan precisas. Con respecto a la similitud entre pares de imagenes en un espacio de baja dimensión, notamos que presenta una relación directa con la precisión de la reconstrucción de las imagenes originales entonces de la misma manera, mientras  $d$  más aumente, hallamos una mayor semejanza entre las imagenes reconstruidas y las originales.

Por otro lado, específicamente para la visualización de datos, las representaciones de estos en espacios de baja dimensión (figuras 7, 8 y 12) nos mostraron como se agrupan los datos en dimensiones reducidas y como, de nuevo, la variación de  $d$  influye en gran cantidad en la compresión de la información. Al mismo tiempo, una vez más, mostramos que a mayor  $d$  mayor similitud entre los datos originales y los datos reducidos. Sabiendo esto último, en la figura 9 luego de recopilar las dimensiones más importantes, podemos concluir que las que tienen una mayor importancia generalmente indican que retienen una mayir cantidad de información de los datos originales, es decir que son fundamentales para la representación de los datos. También, al seleccionar la dimensionalidad para la reducción, los valores de importancia pueden ayudar a identificar el número óptimo de dimensiones que mantienen una cantidad significativa de información. En resumen,  $d$  es crítico en la reducción de dimensionalidad: elegir un 'd' adecuado implica encontrar un equilibrio entre la cantidad de información retenida y la complejidad del modelo.

## 7 Bibliografía

- Burden, Richard L., J. Douglas Faires, and Annette M. Burden. Numerical
- Strang, Gilbert. Linear algebra and learning from data. Vol. 4. Cambridge: Wellesley-Cambridge Press, 2019.
- The Fundamentals of Kernel Density Estimation(2023). Aptech. <https://www.aptech.com/blog/the-fundamentals-of-kernel-density-estimation/>