



PROCESAMIENTO DE IMÁGENES Y REPRESENTACIÓN DE CARACTERÍSTICAS

CIMAT

Introducción a Ciencia de Datos Proyecto final

Autor: Rodrigo Gonzaga Sierra[†]

Profesor: Dr. Marco Antonio Aquino López[†]

Centro de Investigación en Matemáticas A. C.[†]

1. Introducción

El análisis de imágenes digitales es una tarea fundamental en visión por computadora, aprendizaje automático y reconocimiento de patrones. Una característica inherente de este tipo de datos es su alta dimensionalidad: por ejemplo, una imagen de 28×28 píxeles puede verse como un vector en \mathbb{R}^{784} , mientras que bases más complejas como CIFAR-10 trabajan en espacios de más de 3 000 dimensiones. Esta elevada dimensionalidad incrementa los costos computacionales, dificulta la visualización directa y puede degradar el desempeño de modelos de aprendizaje debido a la *maldición de la dimensionalidad*.

Para abordar esta problemática, se emplean técnicas de reducción dimensional que permiten representar la información esencial del conjunto de datos en espacios de menor dimensión. Métodos lineales como PCA capturan la máxima varianza en direcciones ortogonales, mientras que NMF impone restricciones de no negatividad que favorecen interpretaciones basadas en partes. Por otro lado, métodos no lineales como t-SNE permiten revelar estructuras intrínsecas difíciles de observar en el espacio original.

El presente trabajo integra estas herramientas con el objetivo de estudiar sus capacidades para extraer características relevantes, visualizar patrones latentes y facilitar tareas posteriores como el agrupamiento no supervisado o la clasificación mediante redes neuronales.

1.1. Objetivos

- Implementar y analizar técnicas de reducción dimensional (PCA, NMF y t-SNE) aplicadas a bases de imágenes reales.
- Visualizar los componentes latentes obtenidos y discutir su interpretación en el contexto del contenido visual.
- Aplicar métodos de agrupamiento no supervisado sobre las representaciones reducidas y evaluar su coherencia con las clases originales.
- Comparar la efectividad de las diferentes técnicas frente a conjuntos de datos con distintos niveles de complejidad visual.
- Integrar posteriormente modelos de redes neuronales para evaluar el desempeño supervisado y contrastarlo con las representaciones no supervisadas.

2. Metodología

2.1. Conjuntos de Datos

Se trabajó con tres conjuntos de datos ampliamente utilizados en investigación de visión computacional debido a su simplicidad, estandarización y variedad de patrones visuales:

- **MNIST Digits:** Consta de 70,000 imágenes de dígitos escritos a mano, en escala de grises y tamaño 28×28 . A pesar de su simplicidad, continúa siendo referencia en tareas de reducción dimensional, clasificación y análisis exploratorio.
- **Fashion-MNIST:** También contiene 70,000 imágenes de tamaño 28×28 y escala de grises, pero correspondientes a prendas de vestir. Presenta mayor ambigüedad visual que MNIST, lo que lo convierte en un conjunto más desafiante y útil para evaluar la capacidad de los métodos para separar clases similares.
- **CIFAR-10:** Este conjunto incluye 60,000 imágenes a color de tamaño 32×32 distribuidas en 10 categorías (vehículos, animales, etc.). Para mantener compatibilidad con los métodos lineales aplicados, las imágenes se transformaron a escala de grises y posteriormente se vectorizaron a dimensión 1024. CIFAR-10 introduce variaciones más complejas, incluyendo iluminación, textura, postura y fondo, lo que permite evaluar el rendimiento de PCA, NMF y t-SNE en contextos visuales más realistas.

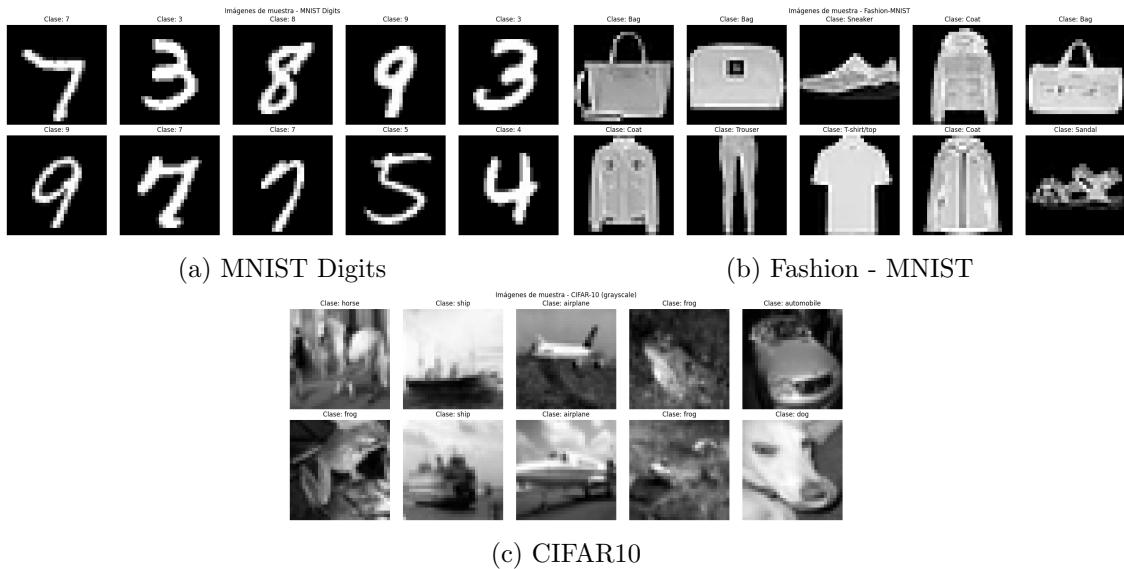


Figura 1: Algunos ejemplos de las bases de datos.

2.2. Preprocesamiento

Todas las imágenes fueron transformadas a vectores de dimensión fija mediante una simple vectorización por filas. Adicionalmente, se aplicó una normalización min–max definida como:

$$X_{\text{normalized}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}},$$

donde X representa la matriz de datos original. Este paso es fundamental porque **PCA** es sensible a magnitudes absolutas, es decir, valores grandes pueden dominar la varianza. **NMF** requiere datos estrictamente no negativos para garantizar una factorización válida. Finalmente, **t-SNE** conserva relaciones de vecindad, por lo que la escala uniforme evita sesgos en la construcción de las distribuciones de probabilidad.

En el caso de **CIFAR-10**, se realizó un paso adicional: conversión a escala de grises mediante una combinación ponderada de los canales RGB y posterior reducción de dimensionalidad mediante vectorización:

$$I_{\text{gris}} = 0.299R + 0.587G + 0.114B.$$

De esta manera, se garantiza compatibilidad con los métodos lineales utilizados y se reduce el ruido cromático.

3. Técnicas de Reducción Dimensional

En esta sección se describimos brevemente las técnicas empleadas para proyectar los datos de alta dimensión a espacios más manejables, destacando su interpretación y su efecto concreto sobre imágenes como las de MNIST y Fashion-MNIST.

3.1. PCA (Análisis de Componentes Principales)

PCA es un método lineal que busca las direcciones ortogonales en las que los datos presentan mayor varianza. Matemáticamente, se basa en la descomposición en valores singulares:

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

donde las columnas de V representan los *componentes principales*. En el contexto de imágenes cada componente puede interpretarse como una “plantilla” global que captura variaciones dominantes: bordes, trazos diagonales, simetrías, etc. La proyección de una imagen en estos componentes determina cuánto contribuye cada característica global al patrón visual. Los primeros componentes capturan estructura visual gruesa; los últimos, ruido y variación fina.

En imágenes de dígitos (MNIST), los primeros componentes representan curvaturas generales de los números, regiones oscuras centrales o periféricas, trazos horizontales y diagonales comunes entre clases.

3.2. NMF (Non-negative Matrix Factorization)

NMF factoriza la matriz de datos en dos matrices no negativas:

$$X \approx WH, \quad (2)$$

donde W contiene las combinaciones de activación para cada imagen y H almacena las *bases no negativas*, que se interpretan como partes aditivas. A diferencia de PCA, NMF produce componentes aditivos que son visualmente interpretables, representan partes locales de las imágenes, descomponen un dígito en trazos o segmentos elementales.

Para MNIST y Fashion-MNIST esto se traduce en segmentos verticales u horizontales, curvas parciales, partes características de prendas (mangas, suelas de zapatos, bolsos). Así, NMF responde a la pregunta: *¿de qué partes está hecha cada imagen?*

3.2.1. t-SNE (t-Distributed Stochastic Neighbor Embedding)

t-SNE es una técnica no lineal diseñada para visualización en dos o tres dimensiones. Busca preservar la estructura local de los datos minimizando la divergencia de Kullback-Leibler entre distribuciones de vecinos:

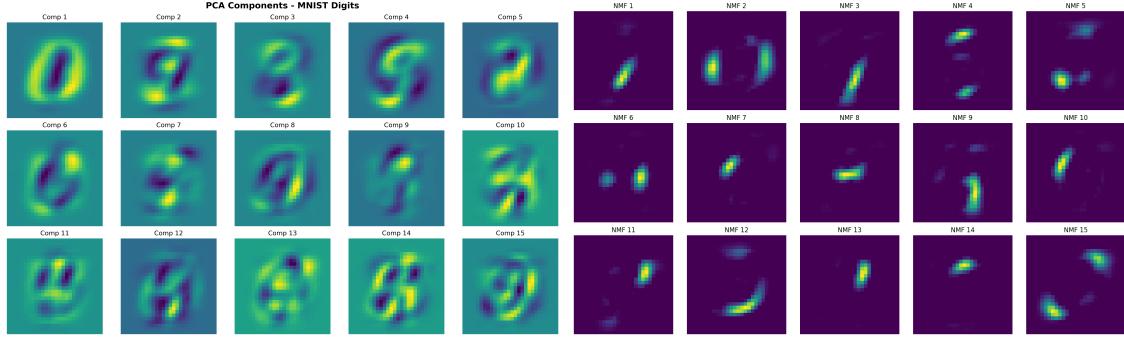
$$KL(P||Q) = \sum_{i,j} p_{ij} \log \left(\frac{p_{ij}}{q_{ij}} \right), \quad (3)$$

donde p_{ij} y q_{ij} representan afinidades entre pares de puntos en los espacios original y reducido, respectivamente.

Respecto a la aplicación de imágenes, esta técnica agrupa imágenes con estructura visual muy similar, separa clases cuando tienen patrones locales muy distintivos, permite observar clusters naturales sin necesidad de etiquetas. En MNIST, por ejemplo los dígitos 1, 0 y 7 suelen formar clusters bien separados; los dígitos similares como 3 y 5 pueden solaparse parcialmente.

3.3. Visualización de Componentes

3.4. Resultados en MNIST Digits



(a) Componentes PCA - MNIST

(b) Componentes NMF - MNIST

Figura 2: Comparación de componentes aprendidos por PCA y NMF en MNIST

En la Figura 2 se observa una comparación directa entre los componentes aprendidos por PCA y NMF sobre el conjunto de datos MNIST. En el caso de **PCA**, cada componente corresponde a una combinación lineal ortogonal de los píxeles. Esto produce patrones globales que capturan variaciones amplias: zonas más oscuras donde algunos dígitos suelen presentar trazos, o gradientes horizontales y verticales que separan clases. Estos componentes reflejan tendencias **colectivas**. Por el contrario, **NMF** genera bases no negativas que representan partes constitutivas de los dígitos. En lugar de plantillas globales, los componentes se asemejan a trazos parciales: segmentos verticales, curvas laterales o pequeñas porciones de bordes. Esto se alinea con la naturaleza aditiva del método y lo hace más interpretable visualmente. Cada dígito puede verse como la suma ponderada de estos fragmentos básicos.

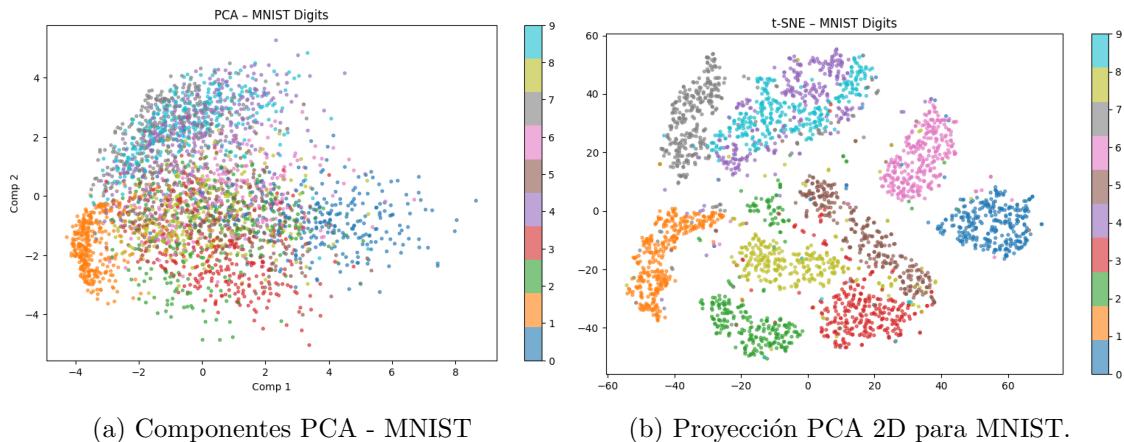


Figura 3: Proyección t-SNE aplicada sobre las primeras 50 componentes PCA de MNIST.

En la Figura 3 se muestran las proyecciones bidimensionales obtenidas al reducir la dimensionalidad mediante PCA, utilizando únicamente las dos primeras componentes principales. Para el conjunto **MNIST**, los dígitos tienden a formar grupos relativamente bien definidos. Esto se debe a que la variación dominante presente en las imágenes está asociada a la forma global de

los trazos, lo cual PCA captura de manera eficiente. Las clases como el dígito 1 aparecen más concentradas, mientras que dígitos como 3, 5 y 8 muestran mayor dispersión debido a similitudes en sus estructuras.

3.5. Resultados en Fashion MNIST

La base de datos Fashion-MNIST representa un desafío un poco mayor en comparación con MNIST Digits, debido a la naturaleza más compleja de sus patrones visuales. Cada imagen contiene información de textura, bordes y formas menos definidas que los dígitos manuscritos, lo cual afecta de forma significativa el comportamiento de los métodos de reducción dimensional.

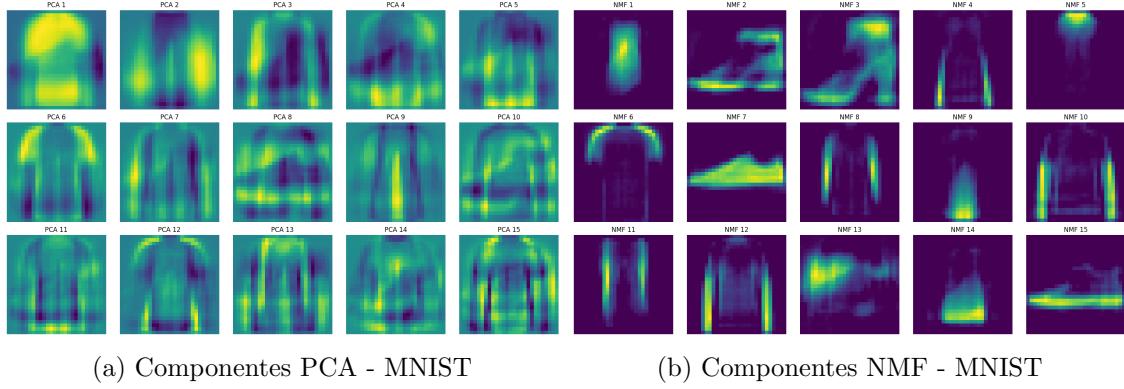


Figura 4: Comparación de componentes aprendidos por PCA y NMF en MNIST

En la Figura 4 se observan las primeras componentes extraídas por PCA y NMF. A diferencia de MNIST, donde las componentes son fácilmente interpretables como trazos o partes de dígitos. **PCA** genera componentes que capturan variaciones globales en iluminación, contornos difusos y contrastes amplios entre distintas prendas. Las primeras componentes tienden a representar diferencias generales entre objetos voluminosos (como *Coat* o *Pullover*) y artículos más planos o simétricos (como *Sneaker* o *Bag*). Sin embargo, debido a la linealidad del método, estas componentes mezclan múltiples atributos visuales, lo que dificulta su interpretación directa. Por otra parte, **NMF**, al imponer no negatividad, produce componentes más localizadas que pueden relacionarse con partes específicas de la ropa: mangas, cuellos, suelas o patrones de textura. No obstante, dichas partes no son tan claramente separables como en los dígitos de MNIST, por lo que los componentes tienden a ser combinaciones de estructuras que aparecen en varias clases.

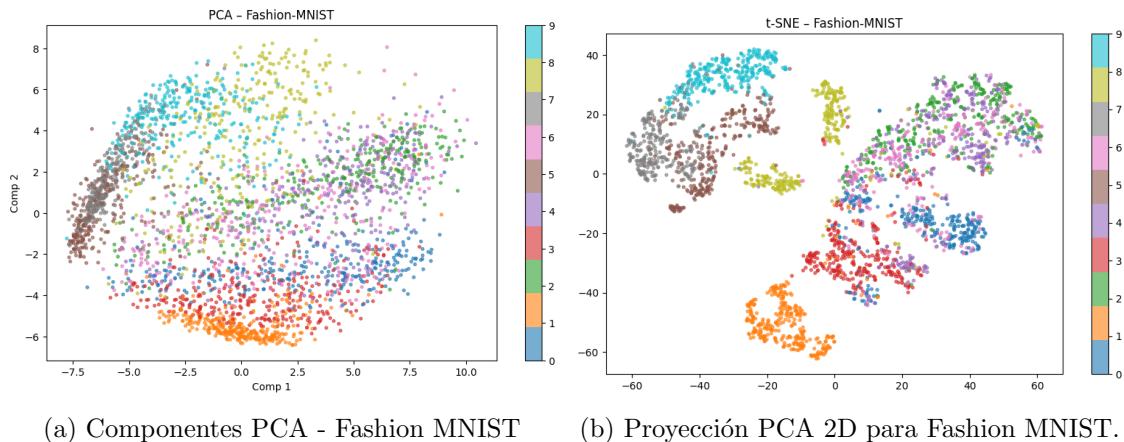


Figura 5: Proyección t-SNE aplicada sobre las primeras 50 componentes PCA de Fashion MNIST.

La Figura 5 compara las proyecciones obtenidas mediante PCA y t-SNE. La primera muestra que la separación entre clases es limitada. Grupos como *T-shirt/top*, *Shirt* y *Coat* aparecen

fuertemente solapados, reflejando la similitud visual entre estas prendas. Esto confirma que las dos primeras componentes capturan sólo variaciones globales, sin distinguir rasgos locales relevantes para la clasificación. Por otra parte, al aplicar **t-SNE** sobre las primeras 50 componentes principales, se obtiene una estructura más clara. Algunas clases, como *Sneaker* y *Bag*, forman conglomerados más compactos. Sin embargo, otras siguen entremezcladas debido a la variación intraclase y la ausencia de bordes distintivos.

3.6. Resultados en CIFAR-10

El análisis de CIFAR-10 representa un caso más complejo debido a la variabilidad inherente de las imágenes: distintos tipos de objetos, texturas más ricas, mayor ruido en el fondo y relaciones espaciales menos estructuradas que en MNIST.

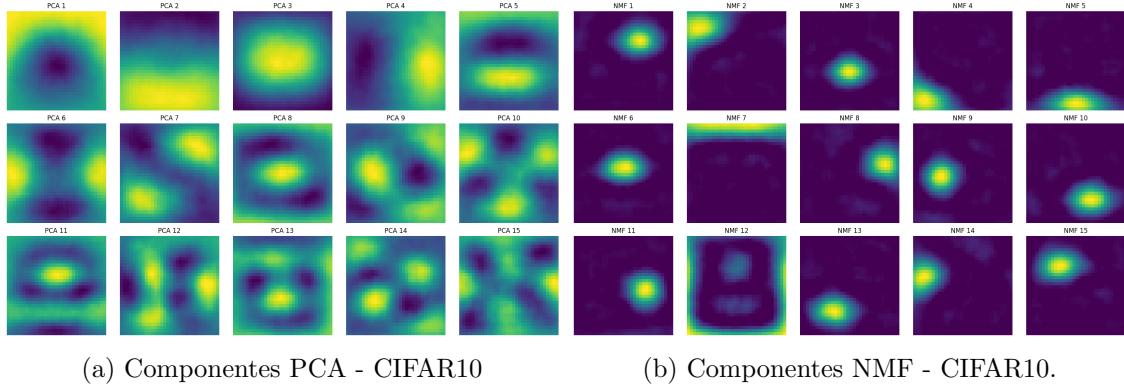


Figura 6: Comparación de componentes aprendidos por PCA y NMF en CIFAR10

Por lo observado en la Figura 6, las componentes PCA muestran patrones globales más difusos y menos interpretables que en MNIST o Fashion-MNIST, lo cual es esperable dado que la variabilidad interclase es más alta y la estructura visual más heterogénea. Por otro lado, NMF produce componentes base que capturan regiones localizadas como bordes, sombras o contornos aproximados de animales y vehículos.

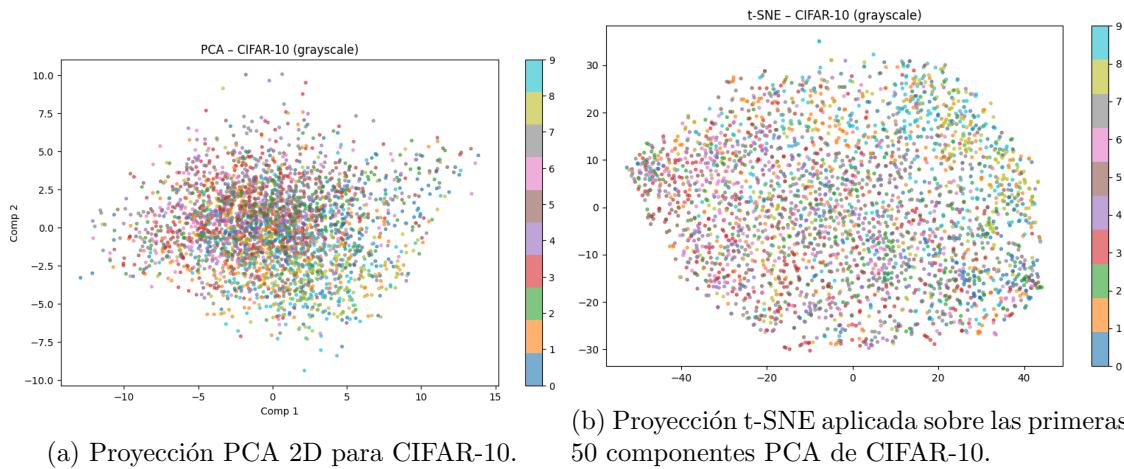


Figura 7: Proyección t-SNE aplicada sobre las primeras 50 componentes PCA de CIFAR10.

La Figura 7, muestra la proyección 2D mediante PCA muestra clusters menos definidos comparados con MNIST. Esto sugiere que la varianza principal en CIFAR-10 no separa de manera clara las categorías, sino que está influenciada por factores como iluminación, textura y color original. El uso de t-SNE produce agrupamientos más discernibles, aunque todavía con mayor traslape que

en MNIST o Fashion-MNIST. Esto confirma que CIFAR-10 presenta una estructura más compleja que requiere métodos no lineales más sofisticados.

3.7. Evaluación

La Tabla 1 resume el desempeño de los métodos de reducción dimensional PCA y NMF como representaciones previas al agrupamiento no supervisado. Se reportan dos métricas estándar: el *Adjusted Rand Score* (ARS), que mide el grado de coincidencia entre las etiquetas verdaderas y las etiquetas obtenidas por clustering, y el *Silhouette Score*, que evalúa la compactación interna de los grupos y su separación respecto a otros grupos.

Tabla 1: Métricas de evaluación del agrupamiento no supervisado para MNIST, Fashion-MNIST y CIFAR-10 utilizando PCA y NMF

Dataset	Método	Adjusted Rand Score	Silhouette Score
MNIST	PCA	0.358	0.116
	NMF	0.165	0.132
Fashion-MNIST	PCA	0.371	0.222
	NMF	0.248	0.279
CIFAR-10	PCA	0.028	0.090
	NMF	0.020	0.094

En el caso del conjunto **MNIST Digits**, PCA supera a NMF en ARS, indicando que las primeras componentes principales preservan suficiente estructura para aproximar las clases verdaderas. Sin embargo, NMF alcanza un Silhouette ligeramente mayor, señalando que los grupos formados son algo más compactos, aunque menos consistentes con las etiquetas reales. Para **Fashion-MNIST**, se observa un comportamiento interesante: PCA logra mejor ARS, pero NMF obtiene superior Silhouette Score. Esto refleja que NMF captura partes locales de las imágenes que favorecen la compactación de grupos, mientras que PCA preserva más la variación global relevante para distinguir clases. En el conjunto **CIFAR-10**, ambos métodos muestran valores muy bajos debido a la alta complejidad del dataset y a que técnicas lineales (PCA) o aditivas (NMF) no capturan adecuadamente la estructura de imágenes a color. Aun así, los valores de Silhouette ligeramente mayores en NMF sugieren que puede formar grupos con algo más de cohesión relativa, aunque sin alinearse con las clases originales.

En conjunto, los resultados indican que la efectividad de cada método depende del tipo de imágenes y la estructura latente del dataset.

4. Redes Neuronales para Clasificación de Imágenes

En las secciones anteriores se emplearon métodos lineales y no lineales de reducción de dimensión como PCA, NMF y t-SNE para analizar la estructura latente de distintos conjuntos de imágenes. Tales técnicas permiten una interpretación geométrica clara del espacio de características, así como la extracción de patrones relevantes. Sin embargo, estos métodos no optimizan explícitamente una función de pérdida de clasificación. Por otro lado, las **redes neuronales artificiales** ofrecen un enfoque complementario y ampliamente utilizado en visión computacional. Su capacidad para aproximar funciones no lineales y aprender representaciones jerárquicas las convierte en una herramienta natural para tareas de clasificación supervisada. En esta sección se presenta un estudio introductorio al uso de redes neuronales en PyTorch. Se diseñaron, 4 redes neuronales distintas para las bases de datos:

- Para la base **MNIST Digits** se diseñó una red neuronal de una sola capa oculta, la cual tendrá 64 neuronas y finalmente la capa de salida tiene 10 neuronas ya que se desea clasificar para 10 clases que son cada uno de los dígitos
- Para la base **Fashion MNIST**, se diseñó una red neuronal de dos capas ocultas para la clasificación de las imágenes, cada una con diferente función de activación y use una función de pérdida en la que el error de clasificar mal un calzado sea el doble que el resto de prendas.
- Para la base **CIFAR10**, se diseñaron dos redes neuronales. La primera solo con una capa oculta completamente conectada de al menos 10,000 neuronas para la clasificación de las imágenes. La segunda fue usando la estructura de LeNet5.

La heuristica de implementación se encuentra en el Apéndice A. Además, de su replicación usando código de Colab del repositorio.

4.1. Resultados

La Tabla 2 muestra una tendencia clara en relación con la complejidad de los conjuntos de datos y la capacidad de generalización de los modelos empleados. En **MNIST**, una red totalmente conectada logra más del 92 % de precisión, reflejando que este conjunto de dígitos escritos a mano posee patrones simples, bien alineados y con baja variabilidad intra-clase. Para **Fashion-MNIST**, la precisión disminuye a aproximadamente 76 %, lo cual era esperado dado que las prendas presentan mayor complejidad visual, diferencias sutiles en textura y formas menos estandarizadas.

Tabla 2: Resumen de desempeño de las redes neuronales para cada conjunto de datos

Modelo	Precisión (%)	Precisión (métrica)	Exhaustividad
MNIST	92.35	0.9235	0.9235
Fashion-MNIST	76.68	0.7571	0.7668
CIFAR-10 (FC 10,000)	48.74	0.4874	0.4874
CIFAR-10 (LeNet-5)	44.35	0.4416	0.4435

La situación se vuelve más notable en **CIFAR-10**. A pesar de incrementar la dimensionalidad de la capa oculta a 10 000 neuronas, la red totalmente conectada apenas alcanza una precisión del 48 %, mostrando que modelos densos no son apropiados para imágenes a color con alta variabilidad en textura, iluminación, perspectiva y fondo. Incluso al emplear una arquitectura convolucional clásica como *LeNet-5*, diseñada originalmente para imágenes simples en escala de grises, el desempeño mejora muy poco, alcanzando únicamente un 44 %. Este resultado evidencia que las CNN poco profundas son insuficientes para la complejidad de CIFAR-10, donde la discriminación entre clases requiere una jerarquía más profunda de filtros convolucionales capaces de capturar bordes, texturas y patrones de alto nivel.

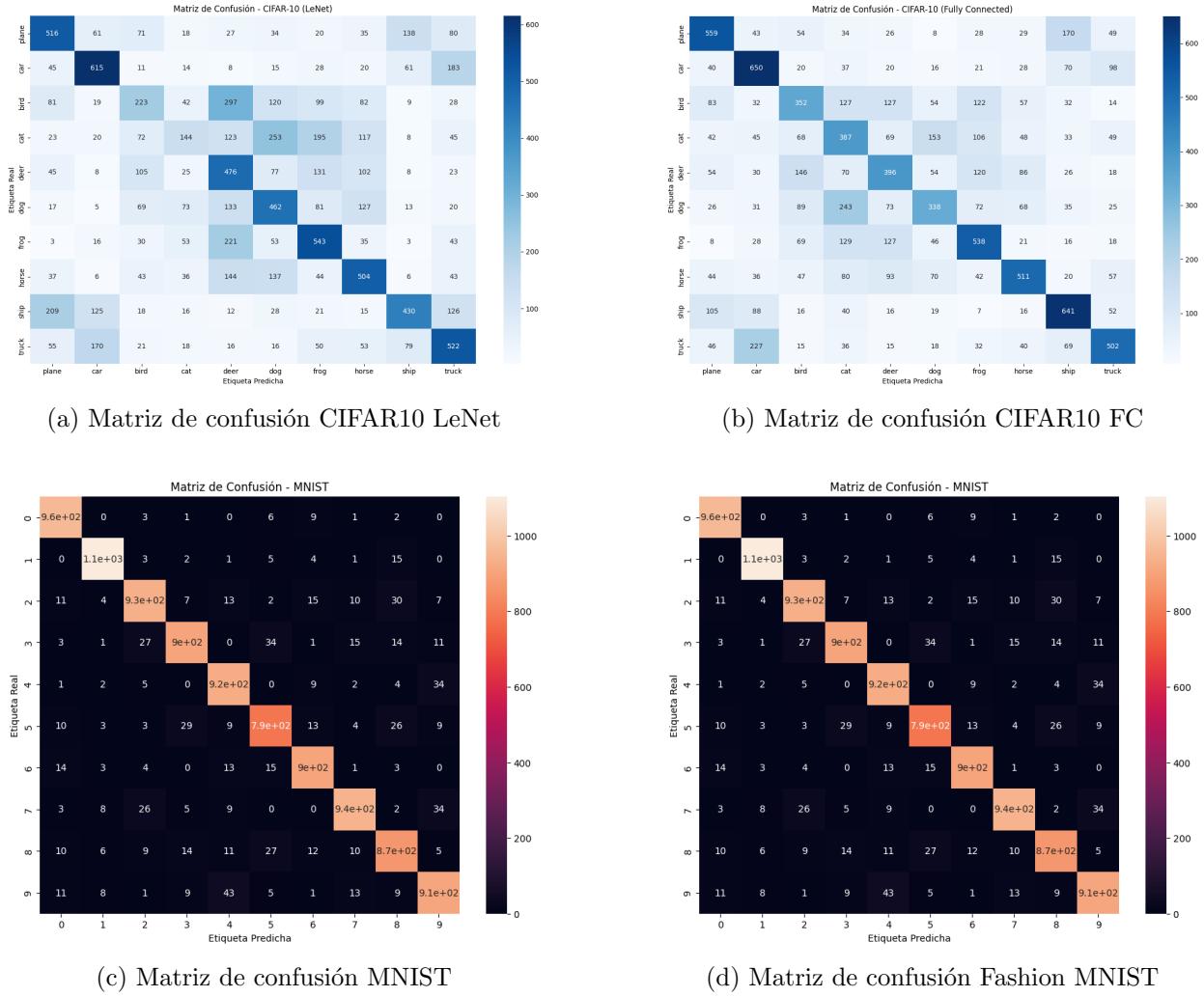


Figura 8: Matrices de confusión para los diferentes modelos y conjuntos de datos

En las figuras 8d y 8c, se puede ver que tiene valores muy altos en la diagonal, lo que significa que la mayoría de las prendas y dígitos se pudieron clasificar con precisión aceptable. Sin embargo, prendas como *Shirt* tiene problemas para decidir con precisión, de hecho se ve que el clasificador confunde *Shirt* con *T-Shirt* y también con *Coat*. En las figuras 8b y 8a, muestran que a pesar de tener valores altos en la diagonal, sigue teniendo confusión con varias imágenes de alrededor.

4.2. Relación con Métodos de Reducción de Dimensión

La incorporación de redes neuronales complementa el análisis previo basado en PCA, NMF y t-SNE. Mientras que los métodos lineales buscan representación latente optimizando variabilidad o restricciones de no negatividad, las redes neuronales optimizan una función de pérdida supervisada. Este contraste permite observar cómo las redes aprenden características jerárquicas. Así como la reducción dimensional no supervisada puede ayudar como etapa preliminar y cómo los modelos complejos como LeNet-5 logran mejorar la precisión en conjuntos más difíciles como CIFAR-10.

5. Discusión

5.1. Interpretación de Componentes

El análisis de componentes latentes revela diferencias importantes entre los datasets debido a su nivel de complejidad. En MNIST, las componentes principales obtenidas mediante PCA capturan trazos globales como curvas amplias o bordes diagonales, reflejando la estructura relativamente

simple y homogénea de los dígitos escritos a mano. En contraste, NMF descompone las imágenes en partes aditivas más interpretables, tales como segmentos de líneas verticales, horizontales o pequeñas curvas.

En Fashion-MNIST, aunque PCA y NMF detectan texturas y contornos, la presencia de clases muy similares (por ejemplo, *Shirt*, *T-shirt/top* y *Coat*) produce componentes menos distintivas. Esto sugiere que la varianza relevante para la clasificación no está tan concentrada como en MNIST, dificultando la separación lineal entre categorías. Para CIFAR-10, las componentes PCA resultan difícilmente interpretables; los vectores propios representan patrones dominados por ruido, texturas de fondo y variaciones no esenciales. NMF logra componentes algo más localizadas, pero aun así la complejidad visual (iluminación, color, perspectiva, fondos diversos) limita la obtención de bases significativas. Esto confirma que las técnicas lineales no son suficientes para capturar la estructura altamente no lineal del dataset.

El uso de t-SNE introduce una mejora visual en la separación de muestras, pero aun así no genera clusters bien formados en datasets complejos. Esto sugiere que la reducción no lineal ayuda a preservar distancias locales, pero no sustituye un modelado más sofisticado. En general, los resultados enfatizan que la aplicabilidad de PCA y NMF depende estrictamente de la estructura geométrica de los datos. Para bases con estructura casi lineal (MNIST), son suficientes; para imágenes naturales, su desempeño decae de forma considerable.

5.2. Implicaciones prácticas y limitaciones del estudio

Los hallazgos tienen varias consecuencias para el trabajo con imágenes. El **preprocesamiento**: métodos lineales pueden servir como etapa inicial de compresión o denoising, especialmente en datos simples. La **visualización** por PCA y t-SNE son herramientas efectivas para explorar datos y detectar outliers o patrones latentes cuando la estructura es moderadamente compleja. El **Modelado** para datasets complejos, la reducción dimensional debe apoyarse en modelos no lineales, idealmente redes neuronales (autoencoders, CNNs, transformadores).

Aunque el análisis es informativo, presenta limitaciones. Sobre la **Linealidad de los modelos**, PCA y NMF no pueden capturar relaciones no lineales presentes en imágenes naturales. La **Dimensionalidad fija**, pues la elección de 20–50 componentes puede no ser óptima para todos los datasets. Por último, la ausencia de implementación de métodos más modernos, no se incluyen comparaciones con UMAP, autoencoders ni representaciones convolucionales, que ofrecerían resultados superiores en CIFAR-10.

6. Conclusión

Este proyecto muestra que las técnicas de reducción dimensional constituyen herramientas fundamentales para el análisis exploratorio de datos de imágenes, aunque su efectividad depende fuertemente de la estructura intrínseca del conjunto. Los resultados permiten concluir lo que ya se ha estado mencionando, **MNIST** presenta una estructura de varianza altamente concentrada y patrones visuales simples. **Fashion-MNIST** exhibe clases visualmente más similares y mayor complejidad. **CIFAR-10** demuestra que los métodos lineales tienen límites claros frente a imágenes naturales complejas.

En términos generales, PCA, NMF y t-SNE son herramientas valiosas para compresión, exploración visual y análisis preliminar. Sin embargo, sus capacidades son insuficientes en contextos donde la representación visual requiere modelar relaciones no lineales.

Como trabajo futuro, se recomienda incorporar técnicas más avanzadas como UMAP, autoencoders variacionales, redes convolucionales auto-supervisadas o modelos generativos, que pueden capturar la estructura profunda de los datos y mejorar tanto la visualización como la calidad del agrupamiento. Estas conclusiones subrayan la importancia de seleccionar adecuadamente la técnica de reducción dimensional en función del tipo de datos y del objetivo analítico.

Referencias

- [Bishop and Nasrabadi, 2006] Bishop, C. M. and Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*, volume 4. Springer.
- [Jolliffe, 2011] Jolliffe, I. (2011). Principal component analysis. In *International encyclopedia of statistical science*, pages 1094–1096. Springer.
- [Krizhevsky et al., 2009] Krizhevsky, A., Hinton, G., et al. (2009). Learning multiple layers of features from tiny images.
- [LeCun, 1998] LeCun, Y. (1998). The mnist database of handwritten digits. <http://yann.lecun.com/exdb/mnist/>.
- [Lee and Seung, 1999] Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *nature*, 401(6755):788–791.
- [PyTorch Team, 2023] PyTorch Team (2023). Training a classifier — pytorch tutorials. PyTorch Official Documentation.
- [Xiao et al., 2017] Xiao, H., Rasul, K., and Vollgraf, R. (2017). Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*.

Apéndice A Redes Neuronales

A.1 Arquitectura General de una Red Neuronal

Una red neuronal totalmente conectada se define como una composición de capas lineales seguidas de funciones de activación no lineales. Para una imagen vectorizada $x \in \mathbb{R}^d$, una red con una capa oculta puede escribirse como:

$$f(x) = W_2 \sigma(W_1 x + b_1) + b_2,$$

donde σ es una función de activación como ReLU o sigmoide. El entrenamiento consiste en minimizar una función de pérdida, típicamente la entropía cruzada, mediante descenso de gradiente estocástico:

$$\ell = - \sum_{i=1}^n \log p_\theta(y_i | x_i).$$

En los problemas planteados en este proyecto se construyen redes neuronales completamente conectadas, redes convolucionales tipo LeNet-5 y variantes con costos ponderados que penalizan de manera diferente ciertos errores de clasificación.

A.2 Implementación en PyTorch

El uso de PyTorch facilita la construcción modular de modelos, permitiendo definir clases que heredan de `nn.Module`. La estrategia básica para todas las redes es:

1. Cargar la base de datos (MNIST, Fashion-MNIST o CIFAR-10).
2. Transformar las imágenes a tensores normalizados.
3. Definir el modelo neuronal.
4. Elegir una función de pérdida y un optimizador.
5. Entrenar la red durante un número determinado de épocas.
6. Evaluarla con un conjunto de prueba.
7. Analizar las métricas obtenidas (precisión, matriz de confusión, curvas de pérdida).