Entrenamiento, optimización y evaluación de clasificadores lineales para clasificación de imágenes

# Introducción

El presente trabajo se centra en la experimentación y evaluación de clasificadores lineales para la clasificación de imágenes utilizando dos conjuntos de datos populares: Fashion-MNIST y CIFAR-10/100. Estos conjuntos de datos han sido ampliamente utilizados en la comunidad de aprendizaje automático debido a su diversidad y desafíos inherentes, lo que los hace ideales para este estudio.

En este estudio, se han elegido dos clasificadores lineales: Naive Bayes Gaussiano y Regresión Logística. Cada uno de estos clasificadores tiene sus propias características y suposiciones, lo que los hace únicos en su forma de abordar el problema de clasificación.

Para cada conjunto de datos, se ha proporcionado una descripción detallada y una visualización de la variabilidad de los datos. Posteriormente, se han desarrollado experimentos específicos para cada clasificador, evaluando su rendimiento en función de la dimensionalidad de las imágenes después de aplicar la SVD. Además de lo encomendado en la tarea, se ha añadido un pequeño estudio sobre la confusión por clases de los modelos ante una complejidad escogida para observar más de cerca que ocurre con las predicciones de los modelos.

# Fashion-MNIST

El análisis de Componentes Principales (PCA) se llevó a cabo para investigar la estructura subyacente y la varianza en el conjunto de datos. Durante este análisis, se centraron los datos restando la media y se calculó la matriz de covarianza. A partir de la matriz de covarianza, se obtuvieron los valores y vectores propios que indican las direcciones de máxima varianza en los datos.

La visualización de la varianza explicada acumulada mostró que no es necesario utilizar todas las componentes para capturar una gran parte de la varianza en el conjunto de datos. Esto es esencial para entender cuántos componentes principales son necesarios para retener la mayoría de la "información" en el conjunto de datos. Además, las primeras 10 *eigen*-imágenes, que representan las direcciones de máxima varianza, ofrecieron una idea de las características o patrones más dominantes en el conjunto de datos.

Una representación bidimensional de los datos, utilizando las dos primeras componentes principales, proporcionó una perspectiva visual de cómo están distribuidas las diferentes clases en el espacio reducido. Esta proyección en dos dimensiones puede ayudar a entender las relaciones y separaciones entre las diferentes clases de ropa en el conjunto de datos. Las representaciones anteriores las podemos encontrar en el archivo *pca.ipynb* de la carpeta *fashion-mnist* además de una adicional con tres componentes en un espacio tridimensional.

## Clasificador naive Bayes Gaussiano

En el archivo *gaussian\_naive\_bayes.ipynb* de la carpeta *fashion-mnist* podemos descubrir los experimentos realizados con el clasificador que lleva su nombre. En primer lugar, se han extraído lo datos, e han dividido en entrenamiento y evaluación para luego ser pasados por el cálculo SVD para las posteriores pruebas sobre dimensionalidad.

A partir de esto, se ha realizado un experimento donde se han comprobado la precisión de diferentes modelos dependiendo de la complejidad de los datos, desde una sola dimensión hasta cincuenta. En este estudio se han ido sumando componentes al análisis desde aquellas con más variabilidad hasta las que menos. Además, se ha recogido información no solo de la precisión obtenida respecto a la complejidad sino además del tiempo de entrenamiento.

Con estos resultados se han representado las gráficas de la Ilustración 1 donde podemos apreciar cómo el modelo cuenta con un punto de inflexión en cuanto a su precisión. Este punto se encuentra aproximadamente entre la componente catorce y la dieciséis donde a partir de este no encontramos una mejora significativa de la precisión al añadir complejidad al modelo. Esto es de gran ayuda porque nos indica la posibilidad de poder optimizar nuestros recursos y conseguir un modelo que, además de preciso, sea conciso y no demasiado complejo. Además, es sabido que un modelo demasiado complejo es probable que no consiga generalizar datos no vistos.

Por otro lado, también podemos apreciar que, conforme aumenta la complejidad también lo hace el tiempo de entrenamiento, razón adicional para tener en cuenta la idea de utilizar un modelo simple frente a uno mayor complejo que no ofrezca una mejora significativa en precisión.

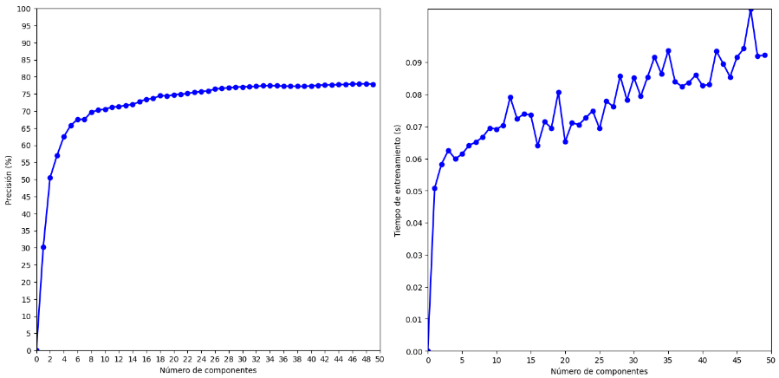


Ilustración - Precisión y tiempo de entrenamiento de clasificadores naive Bayes Gaussianos sobre Fashion-MNIST

Además de esto, se ha realizado una estimación de forma analítica de este punto de *codo* a través de los datos resultantes. Para eso se ha considerado que allá en la precisión donde se encuentre la primera mejora inferior al uno por ciento, es donde se haya el punto de inflexión. Como es de costumbre en este tipo de aproximaciones, cuesta encontrar un buen balance entre la precisión del modelo escogido y la complejidad de este.

Con la aproximación escogida se apuesta por un modelo relativamente simple, aunque no demasiado preciso ya que, como podemos observar en los resultados del documento de código, se ha acabado escogiendo un a precisión de nueve componentes. Esta es relativamente alta para el transcurso de la gráfica, pero, como se ha comentado anteriormente, se puede llegar a considerar a partir de un estudio visual de la representación que este punto se encuentre entre las componentes catorce a la dieciséis. Es objeto de debate una u otra aproximación analítica a este valor, pero eso escapa al alcance de este trabajo.

A partir de esta complejidad, se ha formado la representación de una matriz de confusión completa que podemos apreciar en la Ilustración 2. El resultado nos muestra como el modelo consigue predecir correctamente si vemos esos altos valores en la diagonal, aunque también muestra algunas curiosidades. Es interesante entrever como el modelo relaciona clases que en la realidad son similares como la clase *Shirt* con *T-shirt/top* y, a veces, las confunde de forma significativa. Esto nos puede llevar a pensar que estas clases tienen características muy similares y seguramente haría falta un modelo que consiguiera discernir las diferencias entre estas para abordar la incertidumbre que el modelo que hemos generado tiene sobre ellas.

Imagen que contiene Tabla

Descripción generada automáticamente

Ilustración - Matriz de confusión completa de un clasificador naive Bayes Gaussiano con nueve componentes para el corpus de datos Fashion-MNIST

## Regresión logística

Los experimentos realizados son equivalentes que los descritos en el apartado de Clasificador naive Bayes Gaussiano así que por brevedad nos centraremos únicamente en los resultados obtenidos.

Como podemos observar en la Ilustración 3, este experimento ha resultado en una mayor precisión para número de componentes por encima del *codo*. Visualmente podemos ver que a partir de unas diez componentes dejamos de encontrar una mejora significativa en la precisión del modelo.

A diferencia del anterior tipo de clasificador, en este no encontramos unos tiempos de entrenamiento con tendencia claramente creciente. En su lugar, podemos observar como algunos números de componentes se consiguen entrenar el clasificador en relativamente poco tiempo mientras que otros tardan significativamente más. Estos picos en tiempo de entrenamiento se pueden llegar a deber a aquellas veces que no se ha podido conseguir la convergencia del modelo y, por tanto, se ha entrenado hasta el punto de parada establecido.

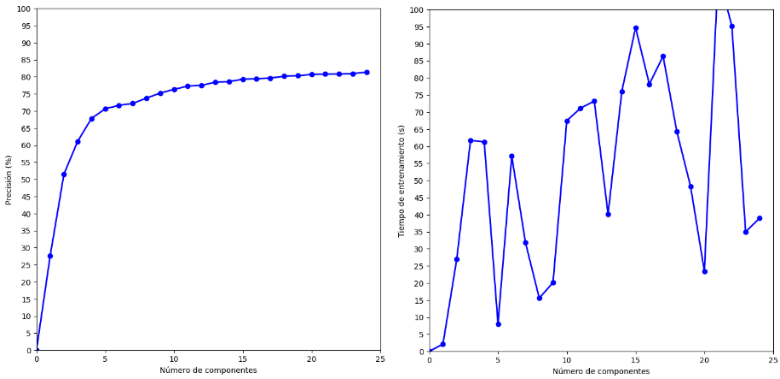


Ilustración - Precisión y tiempo de entrenamiento de clasificadores de regresión logística sobre Fashion-MNIST

Siguiendo la línea del experimento con el anterior clasificador, se ha obtenido que una posible buena complejidad sería escoger seis componentes. A partir de esto, se ha entrenado un clasificador y se ha obtenido la gráfica de la Ilustración 4 donde podemos apreciar un resultado similar al anteriormente obtenido, pero ligeramente de menor confusión de forma general con todas las clases.

Imagen que contiene Tabla

Descripción generada automáticamente

Ilustración - Matriz de confusión completa de un clasificador de regresión logística con seis componentes para el corpus de datos Fashion-MNIST

# CIFAR-10/100

Al igual que con el *Fashion-MNIST*, se aplicó PCA a *CIFAR-10* para identificar la estructura subyacente. A través del proceso de PCA, se extrajeron valores y vectores propios, ofreciendo una visión de las direcciones de máxima varianza en el conjunto de datos. La visualización de la varianza explicada acumulada indicó que, al igual que con *Fashion-MNIST*, no se requieren todas las componentes para retener la mayoría de la información del conjunto. Las *eigen*-imágenes, en este caso, proporcionaron una perspectiva de las características dominantes, revelando patrones de objetos naturales en lugar de artículos de ropa.

La representación bidimensional de *CIFAR-10* en las dos primeras componentes principales mostró la distribución de las clases en el espacio reducido, similar a lo observado previamente con Fashion-MNIST. Esta visualización es crucial para entender la relación y separación entre las categorías de objetos, como podemos ver en el archivo *pca.ipynb* de la carpeta *cifar-10\_100*, entre otras.

En cuanto a la varianza, la observación de que no todas las componentes son esenciales para capturar la varianza sugiere que, al igual que con *Fashion-MNIST*, los modelos entrenados en un conjunto de datos reducido de este corpus podrían beneficiarse de tiempos de entrenamiento más cortos y una menor propensión al sobreajuste.

## Clasificador naive Bayes Gaussiano

Equivalentemente a lo comentado sobre *Fashion-MNIST*, se van a analizar los resultados de los experimentos realizados sobre el corpus *CIFAR-10*.

Aunque la variabilidad del nuevo conjunto de datos es similar a la de *Fashion-MNIST*, nos encontramos con que los resultados de precisión son sorprendentemente peores que los obtenidos en el mismo clasificador para el otro corpus. Esto puede deberse a la suposición de independencia que realiza este tipo de clasificadores sobre las características (en nuestro caso los píxeles de la imagen), lo cuál suele ser complicado que se cumpla en imágenes. Adicionalmente, las imágenes de este corpus tienen un nivel de complejidad y de detalle mayor que el del anterior, lo puede comportar también esta pérdida en precisión.

En cuanto a los tiempos de entrenamiento no encontramos ninguna sorpresa más allá de que estos son relativamente constantes a lo largo de toda la complejidad en contraposición a los del anterior corpus donde eran claramente crecientes. Lo comentado anteriormente puede verse representado en la Ilustración 5.

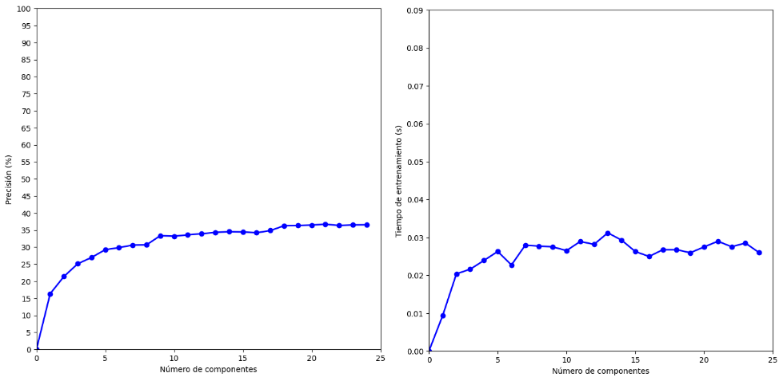


Ilustración - Precisión y tiempo de entrenamiento de clasificadores naive Bayes Gaussianos sobre CIFAR-10

Hemos utilizado un valor de complejidad de seis componentes para realizar la matriz de confusión por clases. Como podemos observar en la Ilustración 6, el resultado de la matriz es bastante peor que el encontrado en la equivalente para *Fashion-MNIST* en la Ilustración 2. La baja precisión del modelo se encuentra en las clases de *automobile*, *bird* y *cat* que son las que más están penalizando a su precisión ya que prácticamente no las distingue de otras clases en algunos casos. Sorprendentemente podemos apreciar que consigue buenos resultados para las clases *airplane*, *frog*, *ship* y *truck* con valores realmente altos en comparación al resto.

Tabla

Descripción generada automáticamente con confianza media

Ilustración - Matriz de confusión completa de un clasificador naive Bayes Gaussiano con seis componentes para el corpus de datos CIFAR-10

## Regresión logística

Para el caso del clasificador basado en regresión logística nos encontramos un escenario francamente parecido en cuanto a la precisión para este corpus de imágenes. Aún así se puede apreciar un ligero descenso en la precisión a lo largo de todas las componentes y, de nuevo, una frenada en el aumento de precisión respecto a las componentes sobre una complejidad de nueve elementos aproximadamente.

En la Ilustración 7, además de lo anterior, se aprecia que, diferencia de *Fashion-MNIST*, los tiempos de entrenamiento tienen una clara tendencia alcista, lo que puede llegar a suponer que en los entrenamientos conseguían hacer converger al modelo en la mayoría de los casos, consiguiendo así, por lo general, tiempos de entrenamiento más cortos que los observados en el anterior corpus.

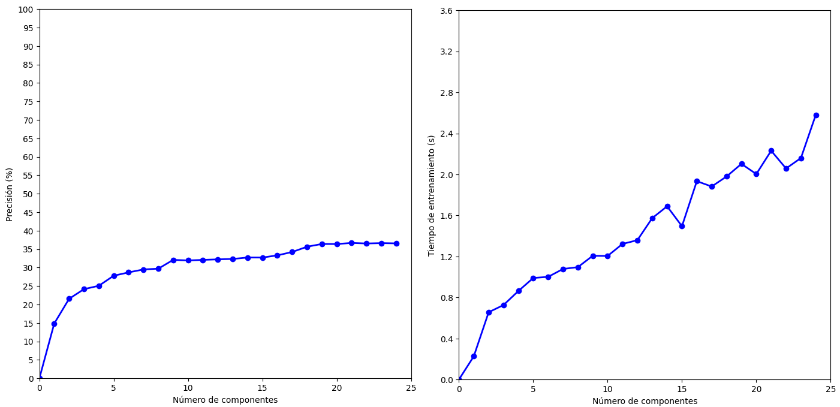


Ilustración - Precisión y tiempo de entrenamiento de clasificadores de regresión logística sobre CIFAR-10

Siguiendo la línea de investigación, se ha desarrollado la matriz de confusión por clases que podemos observar en la Ilustración 8 y que nos deja entrever lo esperable, una peor capacidad de predecir las clases de forma generalizada. Aquí aún se aprecia con más fuerza lo mismo que se observó en la Ilustración 6 pero con aún peores resultados y es evidente que el modelo tiene dificultades para distinguir entre ciertos objetos que pueden tener formas y colores similares en las imágenes.

Interfaz de usuario gráfica, Gráfico, Aplicación, Gráfico de rectángulos

Descripción generada automáticamente

Ilustración - Matriz de confusión completa de un clasificador de regresión logística con cuatro componentes para el corpus de datos CIFAR-10

# Conclusiones

La experimentación con clasificadores lineales en conjuntos de datos de imágenes ha revelado la importancia de la selección de características y la dimensionalidad en la clasificación. Es evidente que no todas las características o componentes son igualmente informativas para la tarea de clasificación.

Para el conjunto de datos Fashion-MNIST, ambos clasificadores mostraron un desempeño decente, aunque la regresión logística pareció tener una ligera ventaja en general. Sin embargo, para el conjunto de datos más complejo CIFAR-10, ambos clasificadores tuvieron dificultades, lo que sugiere que métodos más avanzados o redes neuronales podrían ser más adecuados para este tipo de tareas.

Los resultados también subrayan la importancia de la visualización y la interpretación en el aprendizaje automático. Las matrices de confusión proporcionaron información valiosa sobre las clases que el modelo confundía regularmente, lo que podría ser útil para mejoras futuras.

En general, aunque los clasificadores lineales son herramientas útiles y pueden ser eficientes en términos de tiempo de entrenamiento y predicción, su rendimiento puede ser limitado en conjuntos de datos más complejos y en tareas donde las relaciones entre características no son lineales.