

Práctica 6. PCA

Objetivo

Utilizar técnicas de reducción de dimensión basadas en el análisis de componentes principales. En la primera parte evaluaremos el uso de PCA para la reconstrucción de imágenes y su clasificación. En la segunda parte, usaremos PCA cuando las imágenes tienen ruido Gaussiano para eliminar ruido.

Estudio previo

Repasa las transparencias de clase y estudia las funciones auxiliares proporcionadas para esta práctica. Repasa los conceptos de valores singulares/proprios y vectores singulares/proprios y su interpretación geométrica.

Desarrollo de la práctica

1. Reducción de la dimensión en **Fashion-MNIST**: En esta práctica se empleará uno de los clasificadores Bayesianos desarrollados en la Práctica 5. Sin embargo, vamos emplear el dataset **Fashion-MNIST**, el cual es ligeramente más complejo que el empleado en la Práctica 5 (MNIST). En el template `P6_template.ipynb` ya aportamos el clasificador que usaremos. Además, en dicho template se aporta código para descargar el dataset y cargarlo directamente en la RAM. En esta parte, deberéis realizar:
 - a) Si utilizas PCA para reducir las dimensiones de los datos, ¿con cuántas de las 784 componentes te puedes quedar sin alterar los resultados de la clasificación? Explica detalladamente tu respuesta.
 - b) Si reduces las dimensiones de los datos para mantener el 99% de la variabilidad, ¿qué tan buenos son los resultados de la clasificación? Explica detalladamente tu respuesta.
 - c) Reconstruye alguna imagen del dataset de entrenamiento. Haz reconstrucciones para un número creciente de componentes principales. ¿Qué observas? Además, también puedes emplear la función (ya aportada) `calculate_psnr` (ver **PSNR**) para evaluar cada reconstrucción de forma cuantitativa.
2. Filtrado de ruido en imágenes. En el template `P6_template.ipynb`, también se aporta código para aplicar ruido Gaussiano a las imágenes.
 - a) Utiliza el widget para evaluar cómo PCA elimina el ruido de las imágenes. Verás que el widget utiliza la función que calcula el **PSNR**. Experimenta con distintos niveles de ruido. Razona tus decisiones y explica detalladamente los resultados.
 - b) Evalúa el clasificador de la práctica anterior (y utilizado en el apartado 1) en imágenes de test ruidosas y en el espacio PCA. ¿Cómo cambian los resultados? Detalla tu respuesta.
3. Interpreta la visualización 3D aportada en `P6_template.ipynb`.
 - a) ¿Qué determina la ubicación de las imágenes en 3D?
 - b) ¿Qué efecto tiene PCA en la distribución de las imágenes?

- c) ¿Se observa el mismo tipo de distribución si se emplean las imágenes ruidosas? ¿Por qué?

A entregar en Moodle

- o **Un notebook** `P6.ipynb` con el código de cada apartado, los resultados, su interpretación y las conclusiones que hayas obtenido. **Esta es la opción que más recomendamos debido a que os pedimos interactuar con herramientas particulares de los Jupyter Notbooks, como widgets interactivos.**
- o O, si lo prefieres, dos ficheros `P61.py` y `P62.py` con el código y otro `P6.pdf` con los resultados de todos los apartados, su interpretación y las conclusiones que hayas obtenido.