

Descomposición Valores Singulares

Ejercicio 1.

1. **Estudio elementos que intervienen en descomposición SVD** (aplicación del comando svd de Matlab):

Se considera un diccionario de seis términos $\{t_1, t_2, \dots, t_6\}$ y una colección de cinco documentos de texto $\{d_1, d_2, \dots, d_5\}$. La matriz 6x5 de términos-documentos $G=(g_{ij})$ contiene la información sobre la presencia o no de los términos en los documentos. De esta forma, $g_{ij}=1$ indica que el término i está en el documento j , $g_{ij}=0$ en otro caso:

$$G = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

1.1. A partir de la matriz G , construir una matriz 6x5 que llamaremos

$A = (a_{ij})$ con $a_{ij} = \frac{g_{ij}}{\text{norm}(G(:, j))}$, cuyas columnas están normalizadas $\text{norm}(A(:, j)) = 1, \forall j=1:5$.

1.2. Calcular los valores singulares de A .

¿Cuántos son? Calcular el rango de la matriz A .

1.3. Calcular mediante el método de la potencia el autovalor dominante λ_{\max} de A^*A . ¿Coincide con el cuadrado del mayor valor singular de A ?

1.4. Dar las matrices U y V de vectores singulares por la izquierda y por la derecha, respectivamente, de A .

- Comprobar que las columnas de la matriz U son ortonormales: $(U(:, i))^* U(:, j)$ es igual a 1 si $i=j$ y 0 en otro caso).

- Dar una base ortonormal del espacio de columnas (o espacio imagen de A), ¿qué dimensión tiene dicho espacio?

2. **Aproximaciones de rango bajo** (descomposición SVD truncada, comando svds de Matlab). **Estudio del error.**

Se considera la matriz A del ejercicio anterior.

2.1. Construir las matrices $A_k = \sum_{i=1}^k s_i u_i v_i'$ de rango k , con $k=1:r$, que mejor aproximan a la matriz A

del problema anterior. Los vectores u_i y v_i denotan los vectores singulares por la izquierda y por la derecha asociados al valor singular s_i y $r=\text{rango}(A)$.

Observación: Esta operación se puede hacer también usando el comando svds de Matlab. Comprobar que se obtienen los mismos resultados para $k=3$.

2.2. Calcular el error relativo cometido en cada aproximación $\text{error}_k = \frac{\|A - A_k\|_{fro}}{\|A\|_{fro}}$,

$k=1, \dots, r$.

2.3. Representar en una misma gráfica dichos errores para los distintos valores de k y el vector de valores singulares relativo a la suma de los valores singulares $s_k / \sum_{i=1}^r s_i, k=1, \dots, r$. Comentar los resultados.

Ejercicio 2. Aplicación aproximaciones de rango bajo a compresión de imágenes.

Se considera la imagen photopress_2016.jpg en escala grises.

1. Leer y ver imagen. Obtener matriz A de datos. Estudio matriz A.

1.1. Leer la imagen utilizando el comando `imread('nombrefichero')` (si está en el directorio de Matlab) y guardarla en la variable `im` (con datos tipo `uint8`). Ver la imagen (`imagesc(im); colormap(gray)`).

- Para manipular la imagen como matriz convertimos los datos a doble precisión (`A=double(im);`)

- ¿Qué dimensión tiene A?

1.2.- Calcular el vector $s=(s_k)$ de los valores singulares de la matriz A (no mostrarlos). Representar gráficamente el vector s .

- ¿Cuál es el rango r de la matriz A? Justificar la respuesta.

- Representar gráficamente el vector $\text{cont_inf}=s/s_1$ cuya componente k -ésima viene dada por s_k/s_1 . Mostrar las 21 primeras componentes de dicho vector ¿A partir de qué valor de k aproximadamente los valores singulares se reducen a menos de la mitad de s_1 ? Para el valor de k calculado dar el valor de s_{k+1} .

- Completar los datos de la siguiente Tabla 1:

Rango r	s_1	k	s_{k+1}	s_{k+1}/s_1	$s_{11}/s_1, s_{21}/s_1, s_{31}/s_1$

2. Aproximaciones de rango bajo. Estudio matrices A_k .

2.1. Construir las matrices $A_k = \sum_{i=1}^k s_i u_i v_i'$ de rango k , con $k=10, 20, 30$, que mejor aproximan la matriz A del problema anterior (no mostrarlas), donde los vectores u_i y v_i denotan los vectores singulares por la izquierda y por la derecha asociados al valor singular s_i de A.

2.2. Completar los datos pedidos en Tabla 2:

k	Rango(A_k)	Valor singular dominante de A_k (s_{k1})	s_{k1}/s_1
10			
20			
30			

¿Hay alguna relación entre los valores de la columna tercera de la tabla anterior y los valores singulares de A? ¿Qué relación hay entre los valores de la columna tercera de la Tabla 2 y los de la columna última de la Tabla 1? Justificar las respuestas.

3. Visualizar aproximaciones anteriores. Estudiamos los errores y la compresión obtenida.

3.1. Reconstruir y visualizar las imágenes aproximadas imk a partir de las matrices A_k : Para ello, previamente convertimos los datos tipo `double` a tipo `uint8` (`imk=uint8(Ak);`). Posteriormente vemos las imágenes en pantalla mediante el comando `imagesc(imk); colormap(gray)`. Presentar todas las imágenes incluida la original en una ventana gráfica mediante el comando `subplot(2,2,*)`, donde cada subgráfica debe llevar un título indicando qué imagen muestra ('Imagen original', 'Aproximación de rango 10', 'Aproximación de rango 20',....) ¿Se reconoce la imagen original en dichas imágenes? Comentar los resultados.

3.2. Rellenar los datos de la siguiente Tabla 3 para las aproximaciones de rango bajo del apartado anterior:

k	Error_relativo_k	Ganancia_k
-----	------------------	------------

10		
20		
30		

donde: - $Error_relativo_k$ cometido en cada aproximación $error_k = \frac{\|A - A_k\|_2}{\|A\|_2}$

- $Ganancia_k = k(1+n+m)/(nm)$, indica la razón de datos a almacenar para representar A_k respecto del volumen de datos de A , siendo $n \times m$ el tamaño de la matriz A .
- Representar en una misma gráfica los vectores de las dos últimas columnas de la tabla anterior (2ª columna azul, 3ª columna rojo).
- ¿Qué relación hay entre la últimas columnas de la Tabla 2 y la Tabla 3?

4. Conclusiones del ejercicio. Comentar los resultados de la Tabla 3, la gráfica y las imágenes del apartado 3.

Ejercicio 3. Calcular los valores singulares y el rango de las matrices que se construyen a continuación:

- Se considera el vector $w = [0,2 \ 0,1 \ 0,3 \ 0,25 \ 0,15]$, construir la matriz de ratios $W = (w_{ij})$ con $w_{ij} = w_i/w_j$.
 - Calcular los valores singulares y el rango de W .
 - Calcular el autovalor dominante de W y un autovector positivo asociado cuyas componentes sumen 1. Comentar los resultados.
- Se considera el vector $v = [20 \ 10 \ 30 \ 25 \ 0]^T$, construir la matriz $V = (v_i - v_j) = v e' - e v'$ donde $e = \text{ones}(5,1)$.
 - Calcular los valores singulares y el rango de V .

Ejercicio 4. Se considera la matriz H de Hilbert de dimensión 6 ($H(i, j) = \frac{1}{i+j-1}$, $H = \text{hilb}(6)$).

- Calcular el rango de la matriz:
 - A partir del nº de valores singulares.
 - Utilizando el comando `rank(H)` de Matlab.
 - Calcular de nuevo los valores singulares utilizando formato extendido (en "Preferences" seleccionar "format long "). Comentar los resultados.
- Calcular el nº de condición de la matriz H , $\text{cond}(H) = s_1/s_r$, siendo s_1 y s_r el mayor y menor valor singular respectivamente de H ¿Está bien condicionada la matriz H ?

Conferencia TRY IT 2019

Aplicaciones del Data Science que (probablemente) desconocías/What coders do when they are not coding

"¿Qué tienen en común la predicción del Alzheimer, los análisis del mercado bursátil, la lectura de dígitos manuscritos o el estudio del cambio climático? La respuesta es el análisis de componentes principales (PCA). Esta técnica, básica en el data science, sirve para ilustrar la transversalidad de esta disciplina. Acompáñanos en un viaje a través de ejemplos de aplicación del PCA de la mano de Pablo López, data scientist". Un programador/a se dedica a programar, obviamente. Pero el mundo laboral hoy día exige mucho más que conocer un lenguaje de programación y ser capaz de utilizarlo. Las metodologías ágiles de trabajo como Scrum, Kanban o Design Thinking, el Test Driven Development, la definición de la arquitectura de un proyecto, el uso de microservicios o de git... son factores fundamentales para el éxito de un proyecto actualmente.

Ponente: PABLO LOPEZ

ETS Asset Management Factory

Data Scientist y ganador del premio al mejor Data Scientist de España.