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 { t1,t2,…,t6 } y una colección de cinco documentos de texto { d1, d2,…,d5}. La matriz 6x5 de términos-documentos G=(gij) contiene la información sobre la presencia o no de los términos en los documentos. De esta forma, gij=1 indica que el término i está en el documento j, gij=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 $A = \left(a_{ij}\right)$ con $a_{ij} = \dfrac{g_{ij}}{norm(G(:,j))}$, cuyas columnas están normalizadas $norm(A(:,j)) = 1$, $\forall$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 lambda_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 ui y vi denotan los vectores singulares por la izquierda y por la derecha asociados al valor singular si y r=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 $error\_k = \|A - A_k\|_{fro} \big/ \|A\|_{fro}$, k=1,..,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 $s_k \big/ \sum_{i=1}^{r} s_i$, k=1,...,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 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	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 ui y vi denotan los vectores singulares por la izquierda y por la derecha asociados al valor singular si de A.

2.2. Completar los datos pedidos en Tabla 2:

k	Rango(A_k)	Menor valor singular de A_k
10		
20		
30		

¿Hay alguna relación entre los valores de la columna segunda de la Tabla 2 y los de la cuarta columna 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 Ak: 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 $error\_k = \|A - A_k\|_2 / \|A\|_2$

- *Ganancia_k*=k(1+n+m)/(nm), indica la razón de datos a almacenar para representar Ak respecto del volumen de datos de A, siendo nxm 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 2ª columna de la Tabla 3 y la última columna de la Tabla 1?

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

<u>Ejercicio 3.</u> Calcular los valores singulares y el rango de las matrices que se construyen a continuación:
1. Se considera el vector w=[0,2 0,1 0,3 0,25 0,15], construir la matriz de ratios W=(wij) con wij=wi/wj.
- 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.
2. Se considera el vector v=[20 10 30 25 0]', construir la matriz V=(vi-vj)=ve'-ev' donde e= ones(5,1).
- Calcular los valores singulares y el rango de V.

<u>Ejercicio 4.</u> Se considera la matriz H de Hilbert de dimensión 6 ($H(i,j) = \dfrac{1}{i+j-1}$, H=hilb(6)).

1. 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.
2. Calcular el nº de condición de la matriz H, cond(H)=s1/sr, siendo s1 y sr el mayor y menor valor singular respectivamente de H ¿Está bien condicionada la matriz H?

<u>Ejercicio 5</u>. Se considera la matriz a=(aij) 7X9 cuyas filas recogen la información de la expresión de 9 genes en 7 muestras de tejidos:

```
a(1,:)=[0.9323190  1.2113278 -0.4616314 -0.88042240 -1.09941170 -0.96955658 -0.06758821
0.7638985 1.1145246];
a(2,:)=[0.8975893  1.4624217  0.1259876 -0.43860412 -1.22377397 -1.49015370  0.15747862
1.0964449 0.9901505];
a(3,:)=[0.6811663  1.1275117 -0.2017863  0.02108253 -0.81164718 -1.02881566 -0.53623375
1.1793956 1.1355446];
a(4,:)=[-0.8451067 -1.2319309 -0.9265317 -0.53071756 -0.05634140 -0.27472142  1.04447944
0.7132720 0.7550505];
a(5,:)=[-1.6228992 -0.7777483 -0.7868679 -0.67712937  0.19922605 0.11591201  1.19523024
1.1565408 0.7884622];
a(6,:)=[1.1016 1.0564 1.0067 0.73326  1.2255  1.07  -1.0598  -0.99542 -1.0524];
a(7,:)=[0.64996 0.94287 0.83373 0.80416 0.76872 0.89329 -1.4005 -0.80715 -0.89599];
```

1. A partir de la matriz a, construir la matriz X=(xij) cuyas columnas se obtienen a partir de los datos de las columnas de la matriz a restándoles la media de cada columna. Esto es

$$x_{ij} = a_{ij} - \frac{1}{9}\sum_{k=1}^{9} a_{kj}$$

2. Calcular la matriz (9X9) de covarianza cov= $X^t X$. Calcular la descomposición en valores singulares de cov ([Uc, Varc, Vc]=svd(cov)).
– Construir el vector de valores singulares, que llamaremos var. Hacer una representación del vector var de valores singulares mediante un diagrama de sectores (pie (var)) ¿Qué porcentaje representa var_1 de sum(var) ¿Y var_1+var_2 sobre la suma de valores singulares?
- Mostrar los vectores columnas primero y segundo que recogen los vectores singulares por la izquierda asociados a los dos mayores valores singulares (var1 y var 2) y que denominaremos v1 y v2. v1 y v2 son las denominadas direcciones principales.
Observad que por ser la matriz cov cuadrada las matrices de vectores singulares por la izquierda U y por la derecha V coinciden.

3. Calcular la descomposición SVD de X ([U, S, V]=svd(X)).
- ¿Qué relación tienen los vectores singulares de X con los de $X^t X$?
- ¿Qué relación hay entre la matriz de vectores singulares por la derecha de X (matriz V) y la correspondiente matriz de $X^t X$ (matriz Vc) calculada en el apartado anterior? Fijarse por ejemplo en las dos primeras columnas.

4. Atendiendo a que los dos primeros valores singulares llevan un porcentaje muy alto del peso en las distintas direcciones, se van a proyectar los datos de la matriz X sobre el espacio generado por las dos primeras columnas de V (componentes principales). Esto es, se calcula XV2=U2S2, con [U2,S2,V2]=svds(X,2). Las columnas de esta nueva matriz XV2, que llamaremos PC1 y PC2, representan las coordenadas de los datos (muestras de tejidos) en las direcciones principales (v1 y v2). Construir PC1 y PC2.
Representar gráficamente los puntos en el plano cuya primera coordenada está en PC1 y la segunda coordenada es el correspondiente elemento de PC2: plot(PC1, PC2,'*`). Etiquetar los puntos en la gráfica. Observar la disposición de los puntos en el plano y comentar.