

## Guía 6: Reducción de Dimensiones

1. Dados los siguientes puntos en 4 dimensiones:  $p_1:(3,2,0,3)$   $p_2:(0,1,5,4)$   $p_3:(2,0,6,3)$   $p_4:(2,2,1,2)$ .

Aplicando la SVD obtenemos la siguiente descomposición (por filas):

$U = [-0.2942, 0.7870, -0.0268, -0.5417; -0.6193, -0.2611, 0.7362, -0.0794; -0.6715, -0.3012, -0.6759, -0.0395; -0.2811, 0.4709, 0.0207, 0.8359]$

$S = [10.0500, 0, 0, 0; 0, 4.6033, 0, 0; 0, 0, 1.8316, 0; 0, 0, 0, 0.6727]$

$V = [-0.2774, 0.5866, -0.7594, -0.0477; -0.1761, 0.4898, 0.3953, 0.7569; -0.7370, -0.5738, -0.1929, 0.3006; -0.5907, 0.2944, 0.4795, -0.5784]$

- a. Graficar los puntos en dos dimensiones.
  - b. Interpretar el resultado de la SVD en base a los datos.
2. Se tiene un set de datos sobre autos en donde cada columna mide en una escala de 0 a 1 diferentes datos de performance. En total tenemos 52 mediciones por cada vehículo. Se usa la SVD para descomponer la matriz de datos.
    - a. Explique de que forma podemos estimar a cuántas dimensiones podríamos reducir el set de datos para mantener un 99.5% de la información en el mismo.
    - b. Suponiendo que el punto "a" nos indique que podemos usar 5 dimensiones. ¿De qué forma reducimos el set de datos a 5 dimensiones?
    - c. Explique que representan las primeras "k" columnas de la matriz "V"
  3. Luego de aplicar la SVD a un cierto set de datos obtenemos las siguientes matrices (por filas):  $U = [-0.36, 0.25, -0.68, -0.57, -0.08; -0.29, 0.51, 0.60, -0.36, 0.39; -0.51, 0.11, 0.27, 0.18, -0.79; -0.59, -0.74, 0.09, -0.10, 0.29; -0.40, 0.33, -0.29, 0.70, 0.37]$   $S = [2.41, 0, 0, 0, 0; 0, 0.90, 0, 0, 0; 0, 0, 0.78, 0, 0; 0, 0, 0, 0.25, 0; 0, 0, 0, 0, 0.001]$   $V = [-0.43, -0.41, 0.44, -0.33, -0.58; -0.55, -0.03, -0.54, 0.56, -0.30; -0.40, 0.46, 0.64, 0.38, 0.24; -0.34, 0.65, -0.28, -0.60, -0.10; -0.49, -0.43, -0.13, -0.24, 0.71]$ . Se pide:
    - a. Graficar el set de datos reducido en 2d.
    - b. Indicar la reconstrucción de la matriz original si usamos  $k=1$
    - c. ¿Son todos los números del dataset original positivos? ¿Por qué?
  4. Determinar si las siguientes afirmaciones son V / F justificando la respuesta:
    - a. Tenemos una matriz con las calificaciones de películas x usuario. La matriz es de 50000 películas por 150000 usuarios. La dimensionalidad intrínseca de este set de datos seguramente es menor a 150000.
    - b. La SVD puede usarse tanto para reducir las dimensiones de un conjunto de puntos como para reducir la cantidad de puntos en un set de datos.

- c. Si queremos representar un set de datos en dos dimensiones para visualizarlo entonces la SVD es siempre la mejor opción.
- d. Si tenemos una matriz de  $m \times n$  y queremos reducirla a  $k$  dimensiones entonces podemos aplicar la SVD y luego calcular  $X = U \cdot S \cdot V^T$  y quedarnos las primeras  $k$  columnas de  $X$
- e. Si sabemos que la dimensionalidad intrínseca de los datos es 1 entonces podemos aplicar la SVD y quedarnos con la primer columna de  $U$  para representar a los mismos.