

Evaluación de Análisis Multivariante

Nombres:

Objetivo: Realizar un clustering de productos y determinar los N usuarios a los cuales conviene recomendar un grupo de productos.

Descripción del dataset: El conjunto de datos corresponde a los votos de varios usuarios sobre un conjunto de productos. **Existen 11 usuarios y 15 productos.** Cada fila corresponde a los votos de un usuario sobre todos los productos. Cada columna es un producto. De tal manera que el valor de la fila 0 y columna 0 corresponde al voto del usuario 0 al producto 0, y el valor de la fila 5 y columna 6, corresponde al voto del usuario 5 al producto 6.

	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	i11	i12	i13	i14	i15
0	5	5	5	5	5	0	1	0	0	0	1	0	0	0	0
1	5	5	5	5	5	0	0	1	0	0	0	0	0	0	0
2	5	5	0	5	5	0	0	0	0	0	0	3	0	0	0
3	0	0	1	0	0	5	5	5	5	5	0	1	0	2	0
4	0	3	0	0	2	5	5	0	5	5	0	2	0	4	0
5	0	1	0	0	0	5	5	5	5	5	0	0	0	0	0
6	0	0	0	4	0	0	0	0	1	0	5	4	5	5	5
7	0	1	0	0	0	0	4	0	0	0	5	5	5	4	5
8	0	0	0	0	0	3	0	0	3	0	5	4	5	5	5
9	5	5	5	5	5	1	5	2	1	5	0	0	0	0	0
10	0	1	0	0	0	5	5	5	5	5	0	0	5	0	0

1. Cargar el dataset “datatoy” -> Mostrar el dataframe cuyas dimensiones serían UxI
2. Transponer el dataframe para realizar una reducción de dimensionalidad, en este caso tomar en cuenta los usuarios como dimensiones a reducir. El dataframe quedaría IxU (Dataframe de Inicio)
3. Reducir la dimensionalidad del dataframe IxU y mostrar el análisis de componentes principales (determinación de la cantidad óptima de componentes por el porcentaje de la varianza y mediante gráficas). Cada gráfica: eje X el identificador del componente principal y eje Y la varianza. La salida más relevante será la matriz de componentes de principales con dimensión I x #CP. (#CP es la cantidad de componentes principales)
4. Desde la matriz de componentes principales de PCA con 2 componentes, realizar un clustering de ítems utilizando el algoritmo kmeans, indicar 3 grupos (K=3). Esto dará como salida el vector IDX que indicará el grupo al cual cada producto pertenece. Visualizar los grupos en gráficas en 2D (desde matriz de PCA con 2 componentes).
5. Desde la matriz de componentes principales de PCA con 3 componentes, realizar un clustering de ítems utilizando el algoritmo kmeans, indicar 3 grupos (K=3). Esto dará como salida el vector IDX que indicará el grupo al cual cada producto pertenece. Visualizar los grupos en gráficas en 3D (desde matriz de PCA con 3 componentes).
6. Desde el dataframe de Inicio obtener las filas de los productos que pertenecen al grupo 1 obteniendo el dataframe del grupo.
7. Obtener las medidas descriptivas del dataframe del grupo. Listar el top N usuarios con media más alta.