6.1 Introducción

En aprendizaje no supervisado trabajamos con datos sin etiquetas. El sistema "descubre" estructura (patrones, grupos) basándose en similitud entre ejemplos.

El significado de los grupos lo pone el diseñador: la máquina agrupa, tu interpretas (por ejemplo: "este grupos son tallas M", "este segmento son clientes frecuentes").

Conceptos clave: - Medida de similitud/distancia adecuada al problema - Función objetivo que cuantifique qué tan buena es la agrupación - Criterios para elegir el número de grupos

Ejemplos de usos: - Segmentación de clietnes - Detección de anomalías /outliers - Compresión de colores en imágenes - Agrupación de documentos por temática

6.2 Método K-medias

Queremos k clústeres donde cada punto pertenezca al centroide más cercano, minimizando la suma de distancias al centroide del clúster.

$$J\big(c^{(1)}, \dots, c^{(m)}, \mu_1, \dots, \mu_K\big) = \frac{1}{m}\sum_{i=1}^{m} \left\|x^{(i)} - \mu_{c^{(i)}}\right\|^2$$

y el problema es

$$\min_{c^{(1)},\dots,c^{(m)},\ \mu_1,\dots,\mu_K} J.$$

- $x^{(i)} \in \mathbb{R}^n$: el punto (ejemplo) número i del conjunto de datos. Hay m puntos en total.
- K: número de clústeres que queremos encontrar.
- $c^{(i)} \in \{1, \dots, K\}$: índice del clúster asignado al punto $x^{(i)}$.
- $\mu_k \in \mathbb{R}^n$: el centroide (media) del clúster (k).
- $\mu_{c^{(i)}}$: el centroide del clúster al que se asignó (x^{(i)}).
- $\|\cdot\|$: norma euclídea. $\|x-y\|^2$ es el cuadrado de la distancia euclídea entre dos puntos.

La función J mide la dispersión interna de los clústeres: suma (promedio) de las distancias cuadráticas de cada punto a su centroide. Minimizar J equivale a: - Hacer que cada punto esté lo más cerca posible de su "centro representativo". - Minimizar la variabilidad intra-clúster.

En otras palabras: queremos clústeres compactos.

Supón datos 1D: $x = \{1, 2, 8, 9\}$ y $K = 2$. - Inicialmente centroides $\mu_1 = 1, \mu_2 = 9$. - Asignaciones: {1,2} → clúster 1; {8,9} → clúster 2. - Coste:

$$J = \frac{1}{4}\left[(1-1)^2 + (2-1)^2 + (8-9)^2 + (9-9)^2\right] = \frac{1}{4}(0+1+1+0) = 0.5$$

- Recalcular centroides: $\mu_1 = (1+2)/2 = 1.5,\ \mu_2 = (8+9)/2 = 8.5$. - Nuevo coste:

$$J = \frac{1}{4}[(1-1.5)^2+(2-1.5)^2+(8-8.5)^2+(9-8.5)^2] = \frac{1}{4}(0.25+0.25+0.25+0.25) = 0.25$$

- Ya las asignaciones no cambian, convergió en un mínimo local (que aquí es también el global).

El algoritmo funciona de la siguiente forma: 1. Inicializa k centroides (aleatorios en principio) 2. Asignación: asigna a cada punto el centroide más cercano 3. Actualización: recalcula cada centroide como la media de los puntos asignados 4. Repite 2-3 hasta la convergencia

Ejemplo: Datos 2D (altura/peso simplificado):
$A(1,1), B(1.5,2), C(3,4), D(5,7), E(3.5,5), F(4.5,5), G(3.5,4.5)$.
$k = 2$, inicializa centroides en $A(1,1) y D(5,7)$.

- Asignación (distancia euclídea):
  - Cerca de $A : A, B$
  - Cerca de $D : C, D, E, F, G$
- Actualización de centroides:
  - $\mu 1 = media(A, B) = ((1 + 1.5)/2, (1 + 2)/2) = (1.25, 1.5)$
  - $\mu 2 = media(C, D, E, F, G) = (3.9, 5.1)$
- Siguiente iteración: vuelve a asignar con los nuevos centroides y repite hasta estabilizar.

6.3 Elegir el número de clústeres

Empleamos el método del codo, donde se calcula la inercia para $k = 1, 2, ...k$. El codo es el punto donde añadir más clústeres ya apenas reduce el coste.

Por ejemplo no podemos hacer una talla por persona ($k = n$) ni una sola talla para todos ($k = 1$). Buscas el mínimo k que capte bien la variabilidad de la población con un coste razonable.

6.4 Buenas prácticas

Eliminar los outilers, pueden arrastar los centroides, por lo que es importante detectatlos y filtrarlos.