



K-medoides

K-medoides

- ▶ La partición alrededor de los “medoides” (se suele llamar también Partition around medoids - PAM) es otro algoritmo de partición.
- ▶ Esencialmente, PAM es una modificación del algoritmo K —medias.
- ▶ Este algoritmo busca K "objetos representativos" en lugar de los centroides entre las observaciones en el conjunto de datos.
- ▶ Entonces, se espera que el método sea más robusto ante anomalías o atípicos.
- ▶ Una desventaja del algoritmo es que, aunque funciona bien en pequeños conjuntos de datos, no son lo suficientemente eficientes como para utilizarlos para agrupar grandes conjuntos de datos.

K-medoides

► El algoritmo es el siguiente:

1. Sea \mathbf{x}_i para $i = 1, \dots, n$ el conjunto de observaciones de la matriz de datos.
2. Calcular $D = \{d_{ij}, \text{tal que } i, j = 1, \dots, n\}$, una matriz que contiene las distancias entre las n observaciones.
3. Elegir K observaciones como los “medoides” de los K grupos iniciales.
4. Asignar cada observación a su “medoide” más cercano usando la matriz de distancias D .

K-medoides

► El algoritmo es el siguiente:

5. Para cada clúster, buscar la observación \mathbf{x}_j (si existe) que proporciona la mayor reducción de

$$SCDG = \sum_{k=1}^K \sum_{c(i)=k} d_{ij}^2$$

donde $c(i)$ es el clúster que contiene a \mathbf{x}_i

Seleccionar la observación \mathbf{x}_j que reduce $SCDG$ como el nuevo “medoide” para ese clúster (note que $SCDG$ sólo considera las distancias entre las observaciones del clúster en cuestión y su medoide).

6. Repetir pasos (4) y (5) hasta que no se reduzca $SCDG$.