

Agrupamiento con `scipy.cluster` y `scikit-learn`

Partimos de que tenemos los datos en una matriz X Cada fila es un punto de la muestra; cada columna es el valor de una variable.

Jerárquico (`scipy.cluster.hierarchy`)

1. Hacemos el agrupamiento con `Z=linkage(X)`
2. Evaluar los agrupamientos: `inconsistent(Z)`
Hay que fijarse en la cuarta columna. Habrá valores muy bajos y otros altos. Localiza un umbral que separe con la mayor claridad posible
3. Crear el agrupamiento: `T = fcluster(Z, umbral, criterion='inconsistent')`
 T es una columna con tantas filas como puntos de la muestra, en la que se indica el número de grupo que le corresponde

Medias (`scipy.cluster.vq`)

1. Normaliza los datos: `X=whiten(X)`
2. Vamos a iterar variando el número de grupos, para encontrar la cantidad mejor
 1. Crea el agrupamiento con: `centros, distor = kmeans(X, numgrupos)`
 2. Identifica el grupo de cada punto con `idx, distor=vq(X, centros)`
 idx es la columna que tiene por cada fila el número de grupos que le corresponde
 3. Evalua el agrupamiento con (`sklearn.metrics`):
`silh = silhouette_samples(X, idx)`
En un buen agrupamiento deben ser todos los valores de $silh$, o los más posibles, bastante altos (mayores que 0.5)