Algoritmo de las K-Means (K-Medias)

Hay varios algoritmos propios del aprendizaje no supervisado. Sin embargo, vamos a concentrarnos en K-Mean (K-Medias) que permite hacer clasificaciones en una cantidad de grupos de predicción conocida y en una cantidad pequeña de datos (inferior a 10 000).

El clustering (fragmentación de datos)

La fragmentación de datos, también llamada clustering, es la acción de distribuir el conjunto de las observaciones en grupos pequeños que tienen características comunes.

Esta fragmentación de los datos es el resultado esperado de cada algoritmo de aprendizaje no supervisado. Así, cuando se le proponga al algoritmo una observación nueva para clasificar, él se encargará de colocarla en uno de los grupos que tendrá determinados. Nosotros nos ocuparemos de la definición de la etiqueta del grupo que da sentido a la clasificación (albaricoque, cereza…). La máquina no es capaz de hacerlo porque no es consciente del significado de los datos que utiliza durante su aprendizaje (para ella solo son números).