1 k-means clustering

k-means clustering is a data clustering method which clusters input data into k different classes. The classes are represented by the class means μ_i and points are considered to be in a class S_i if the squared distance to the class mean is the minimum compared to the squared distance to the other class means. Formally:

$$S_i = \{x \in X : ||x - \mu_i||^2 \leq ||x - \mu_j||^2 \forall 1 \leq j \leq k\}$$

k-means finds the placement of the class means by minimization of the summed squared distance of all class points to the class mean for all k classes:

$$\arg\min_S \sum_{i=1}^{k} \sum_{x \in S_i} ||x - \mu_i||^2$$

A common algorithm to find this is Lloyd's algorithm, which iteratively classifies points according to current class means and updates them with the average of all classified points until convergence.