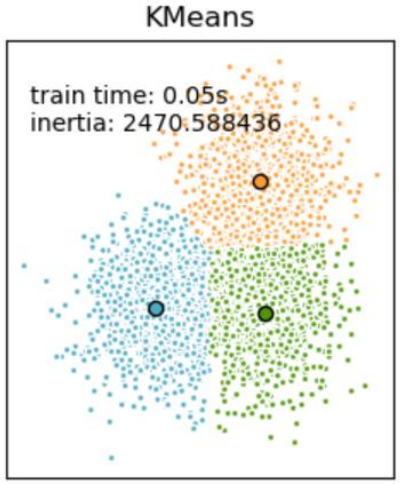
Algorithm



k means

$$J = \sum_{i=1}^{k} \sum_{\mathbf{x}_j \in S_i} \|\mathbf{x}_j - \boldsymbol{\mu}_i\|^2$$

standard score

$$Z = \frac{X - \mu}{\sigma}$$

μ = mean
σ = standard deviation

Algorithm

1. Specify number of clusters K
2. Randomly select K distinct centroid (new data points as cluster initialization)
3. Measure the distance (Euclidean distance) between each point and the centroid
4. Assign the each point to the nearest cluster
5. Calculate the mean of each cluster as new centroid
6. Repeat step 3–5 with the new center of cluster
7. Repeat until stop condition reached
    - Convergence (no further changes) or maximum number of iterations