Machine Learning and Artificial Intelligence

Lab 06 – K-Means clustering

13/04/2021

Unsupervised Learning

- Unsupervised? We Cannot provide information about the ground truth state of the features to the model.
- Why? We do not have annotations for our data.
- Clustering: Study of the features of our patterns with the aim of creating different *groups.*

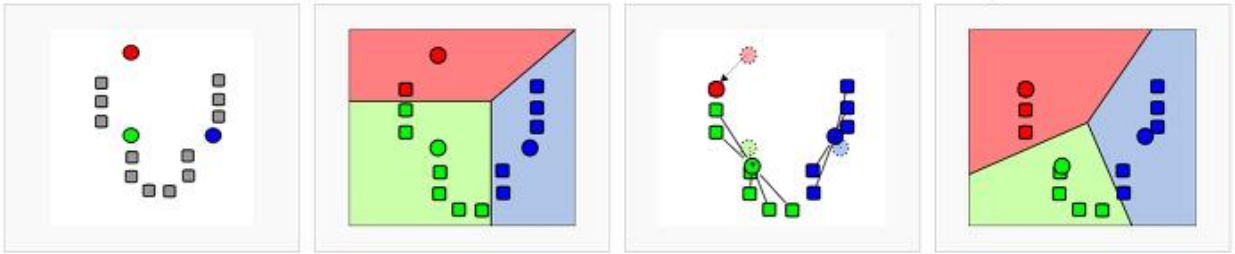
Clustering Approaches

- Partitioning methods:
  - Partitions objects into K clusters while minimizing a cost function.
  - Example: **K-Means**
- Density-based methods:
  - Find regions with high *density* that are separated from other regions with low *density*.
  - Example : **Mean-Shift**
- Hierarchical methods:
  - They form a tree structure (dendrogram) of clusters where a new cluster consists of a merge (or division) of previously defined clusters.
  - Es: **Linkage**

K-means

- A priori: Decide the number of clusters. This is a hyperparameter.
- Each cluster is represented by the average of the points that are part of it (centroid).
- How does it work?
  - Start with an initial, random guess of centroids (cluster centers).
  - Each iteration assigns each pattern to the nearest centroid.
  - Update the centroids as the mean of the observations belonging to that cluster.
  - Repeat until convergence.

Demonstration of the standard algorithm



1. *k* initial "means" (in this case *k*=3) are randomly generated within the data domain (shown in color).

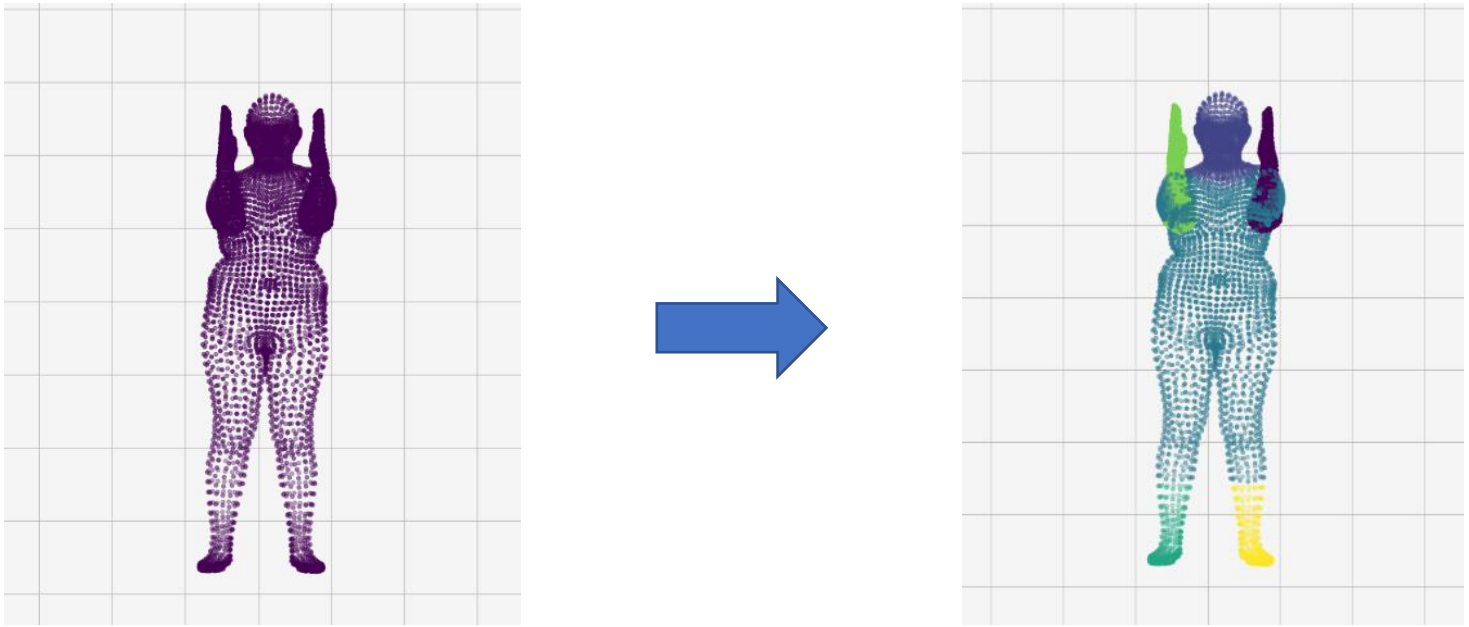
2. *k* clusters are created by associating every observation with the nearest mean. The partitions here represent the Voronoi diagram generated by the means.

3. The centroid of each of the *k* clusters becomes the new mean.

4. Steps 2 and 3 are repeated until convergence has been reached.

A concrete problem

- 3D shape segmentation: identify different parts of the body



Exercises