AI & Robotics

Unsupervised Learning

PXL AI & ROBOTICS LAB

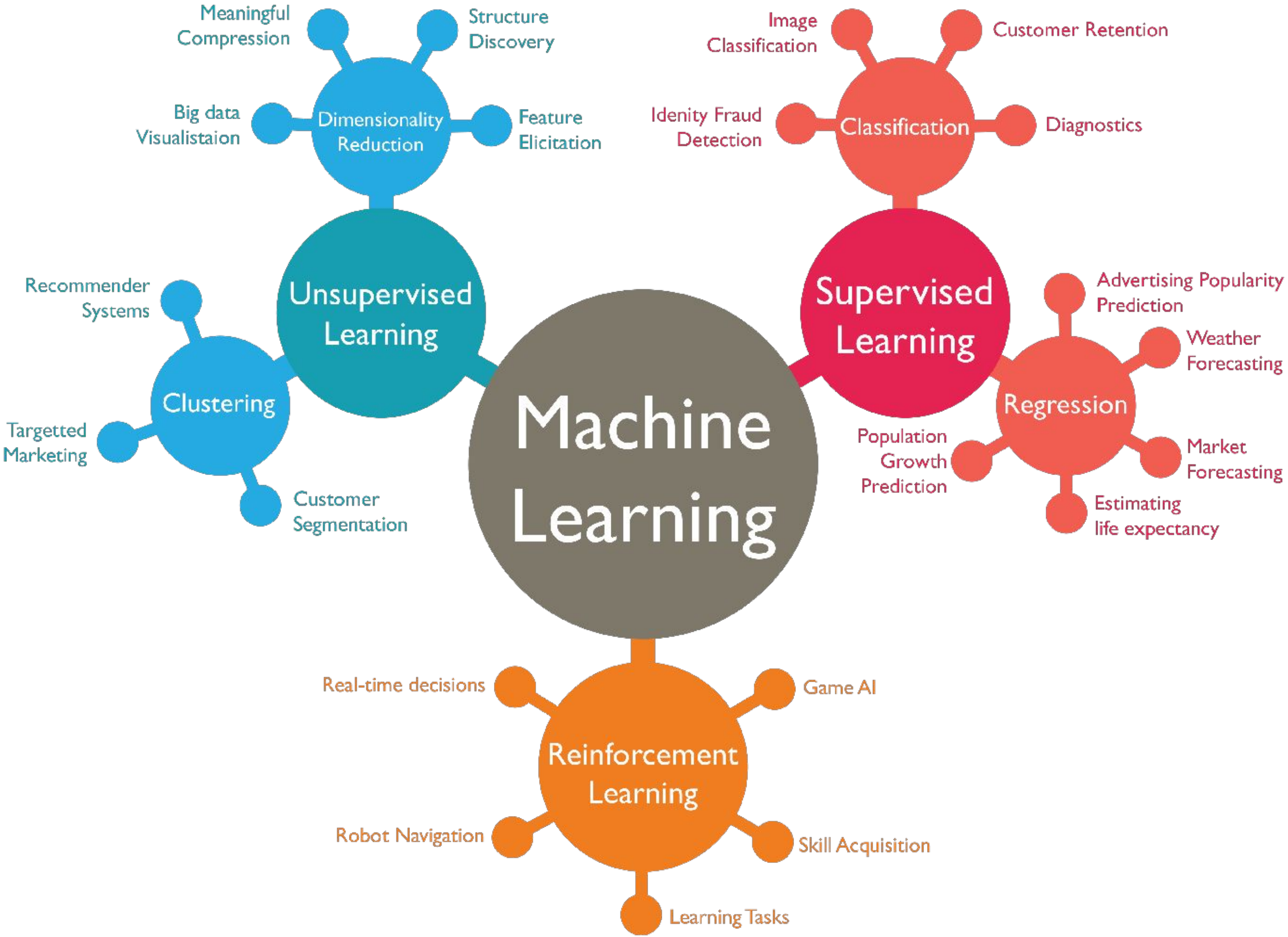
Goals (1/2)

The **junior-colleague**
- can explain Unsupervised Learning in their own words
- can describe the general flow of an unsupervised learning pipeline
- can list at least 3 examples of unsupervised learning
- can explain the purpose of clustering
- can explain the difference between clustering and classification
- can explain how to prepare data for clustering
- can explain how the k-means algorithm works in their own words
- can describe the advantages and disadvantages of k-means
- can explain how k-means++ optimizes the centroid choice problem of k-means
- can explain how the Elbow method in the context of k-means works
- can implement k-means and DBSCAN for a given clustering problem

Goals (2/2)

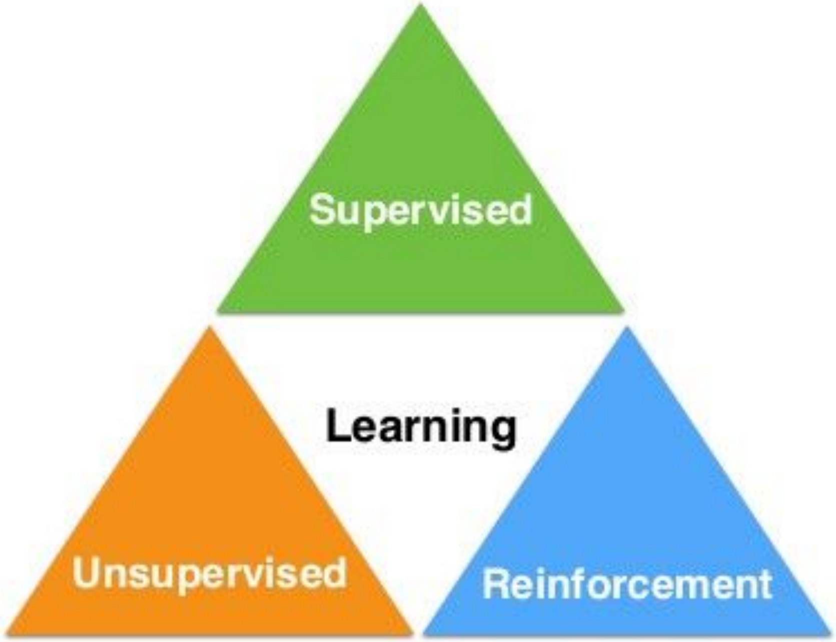
The **junior-colleague**

- can explain how the DBSCAN clustering algorithm works in their own words
- can describe the advantages and disadvantages of DBSCAN
- can explain how hierarchical clustering works on the basis of a visual representation of a dataset
- can describe the different linkage forms for hierarchical clustering
- can describe the advantages and disadvantages of hierarchical clustering


Machine Learning
Unsupervised Learning
Dimensionality Reduction
Meaningful Compression
Structure Discovery
Big data Visualistaion
Feature Elicitation
Clustering
Recommender Systems
Targetted Marketing
Customer Segmentation
Supervised Learning
Classification
Image Classification
Customer Retention
Idenity Fraud Detection
Diagnostics
Regression
Advertising Popularity Prediction
Weather Forecasting
Population Growth Prediction
Market Forecasting
Estimating life expectancy
Reinforcement Learning
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks

Machine Learning

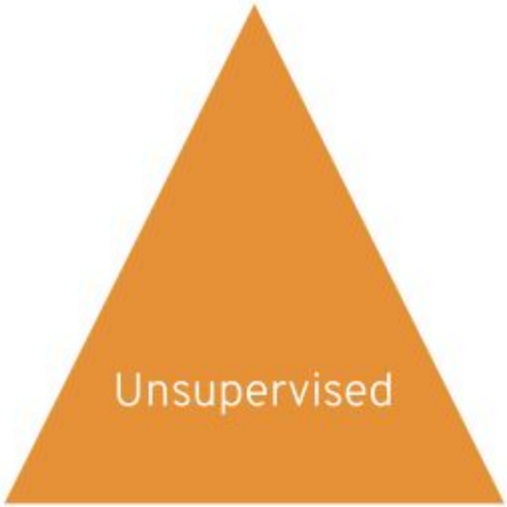
- Labeled data
- Direct feedback
- Predict outcome/future



- No labels
- No feedback
- “Find hidden structure”

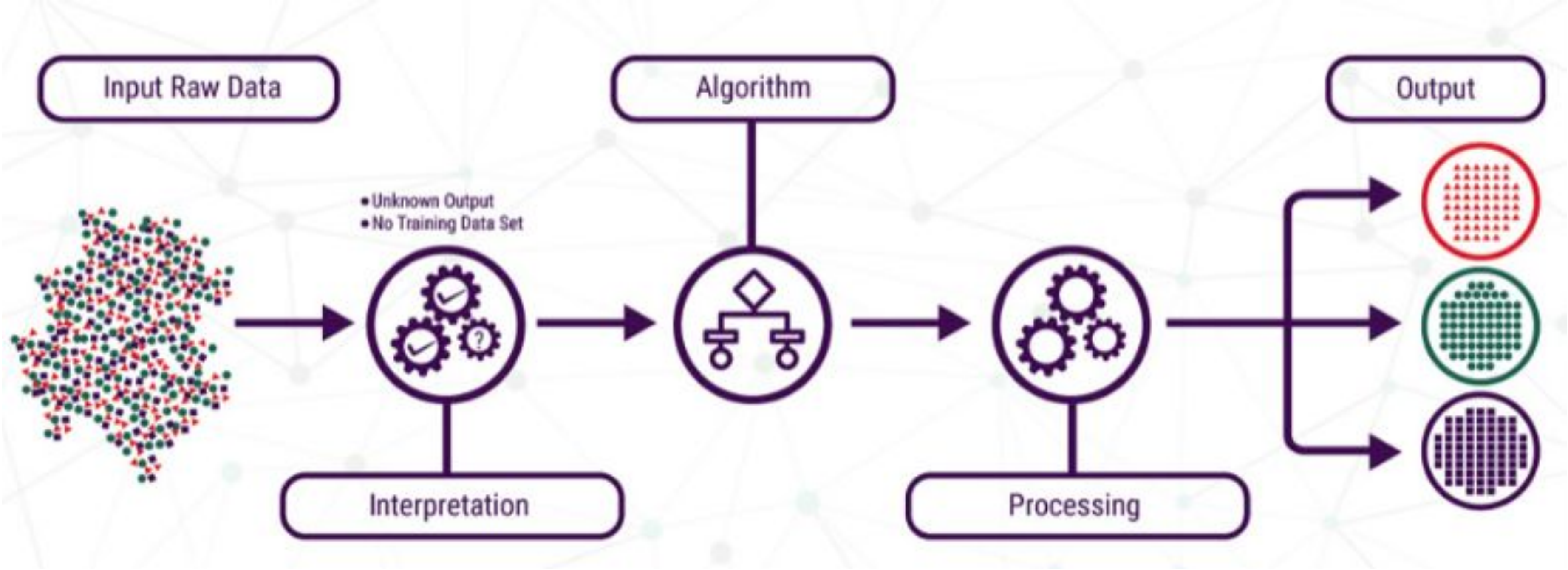
- Decision process
- Reward system
- Learn series of actions

Unsupervised Learning

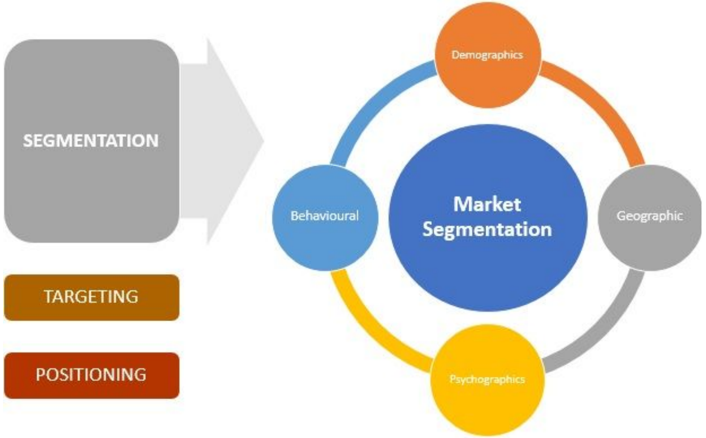


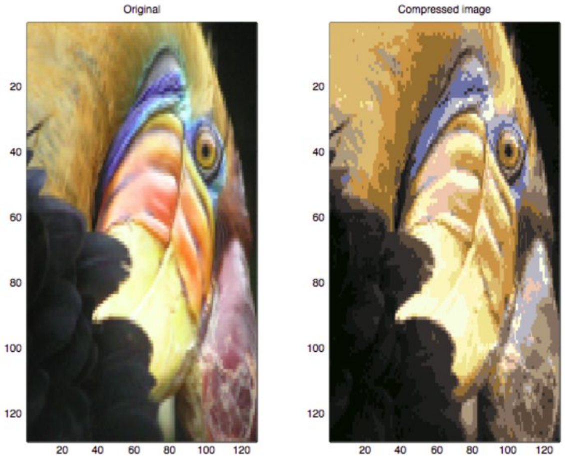
- No labels
- No feedback
- “Find hidden structure”
- Meaningful patterns in unlabeled data
- Usage is on the rise

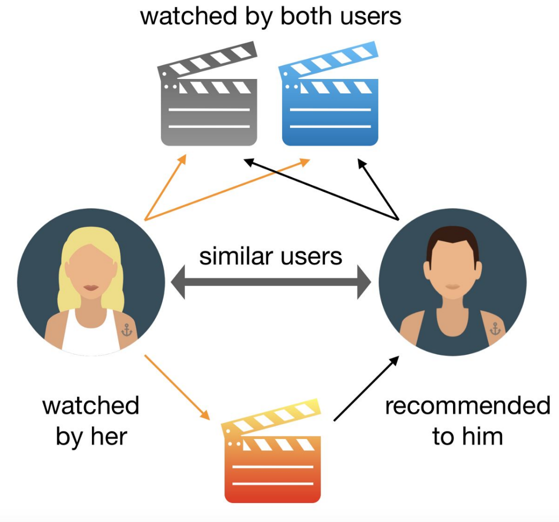
Unsupervised Learning



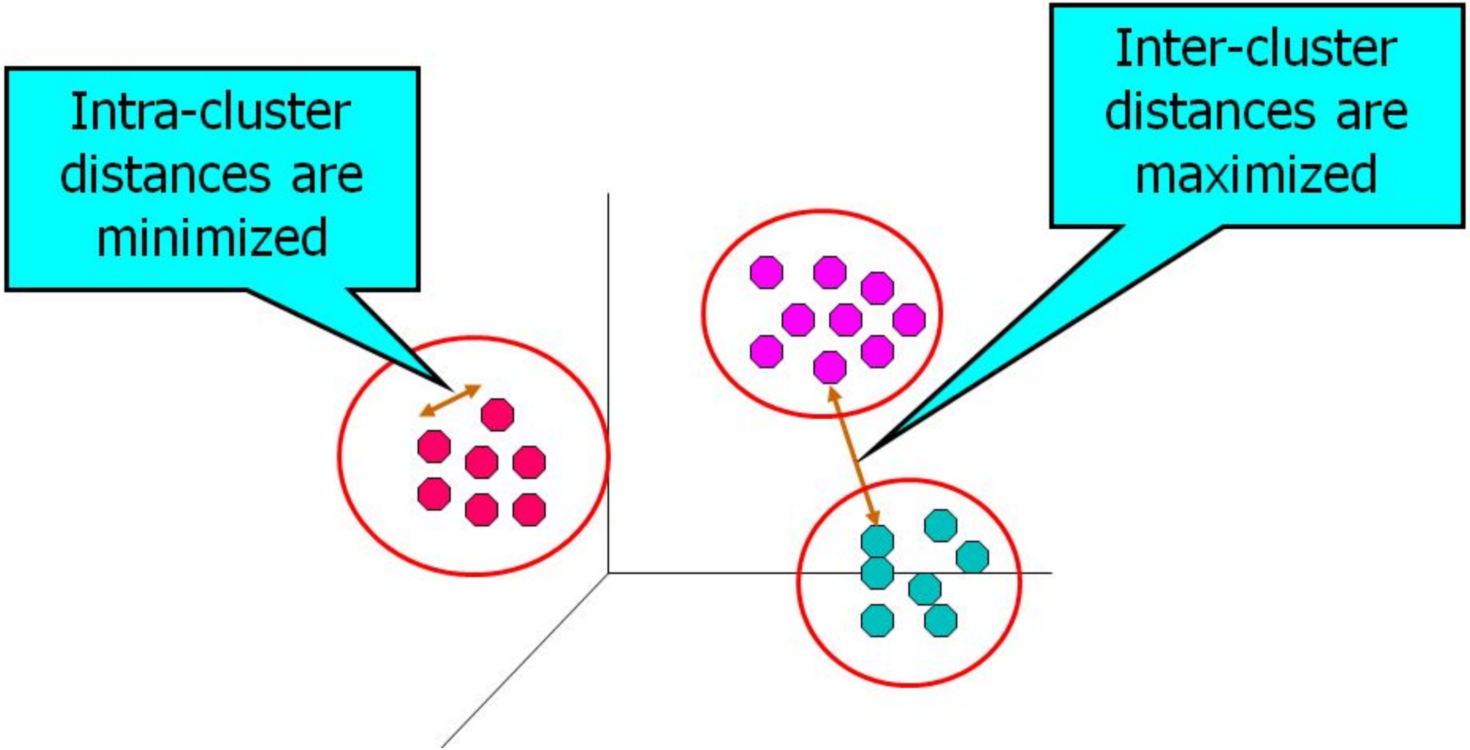
Unsupervised Learning







Clustering


Intra-cluster distances are minimized
Inter-cluster distances are maximized

Clustering

- Centroid-based clustering
  - **K-Means**
  - Mean-shift
- Density-based Clustering
  - **DBSCAN**
  - Mean-shift
- Connectivity-based clustering
  - **Hierarchical (agglomerative) Clustering**

Preparing data

- Categorical variables to numerical data
- Handle missing values
- Clustering uses distance metrics

  => Scale the data

  - Not necessary for some types of well defined data

    => i.e. clustering geolocation data (longitudes and latitudes)
  - Necessary for data that is of different physical measurements or units

K-Means Clustering

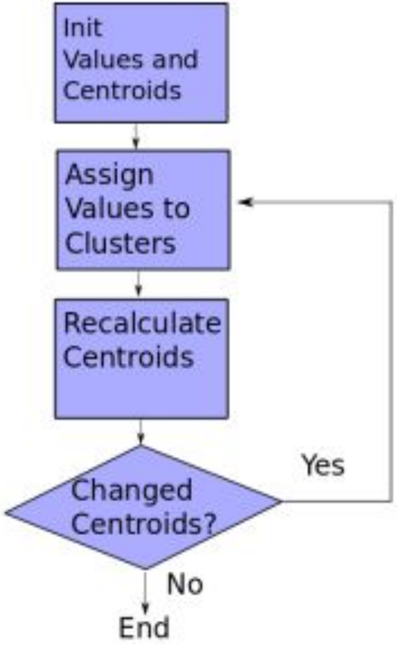
K-Means Clustering

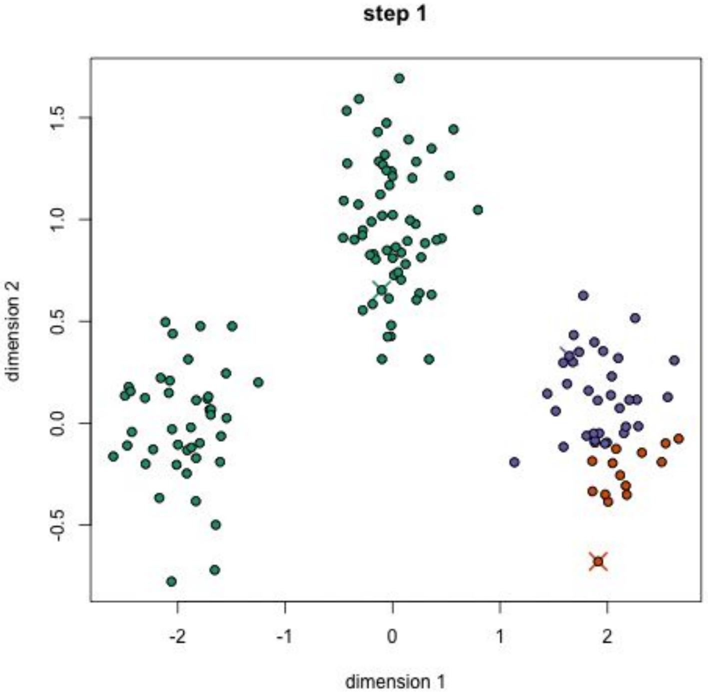
- Very popular clustering algorithm
- Easy to implement
- Naive method

  => Doesn’t know what the number of clusters is

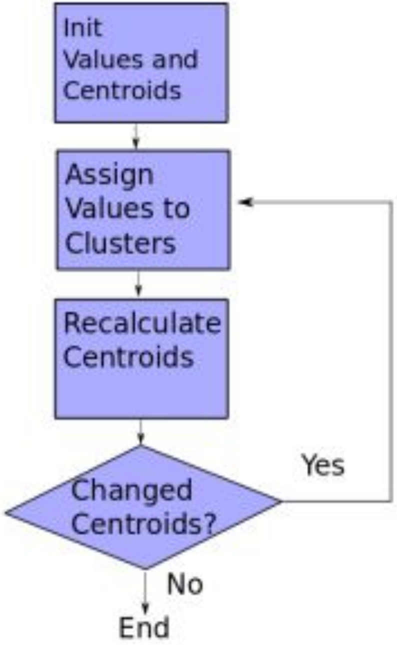
  => K (# of clusters) is supplied as a parameter
- Finds a local optimum

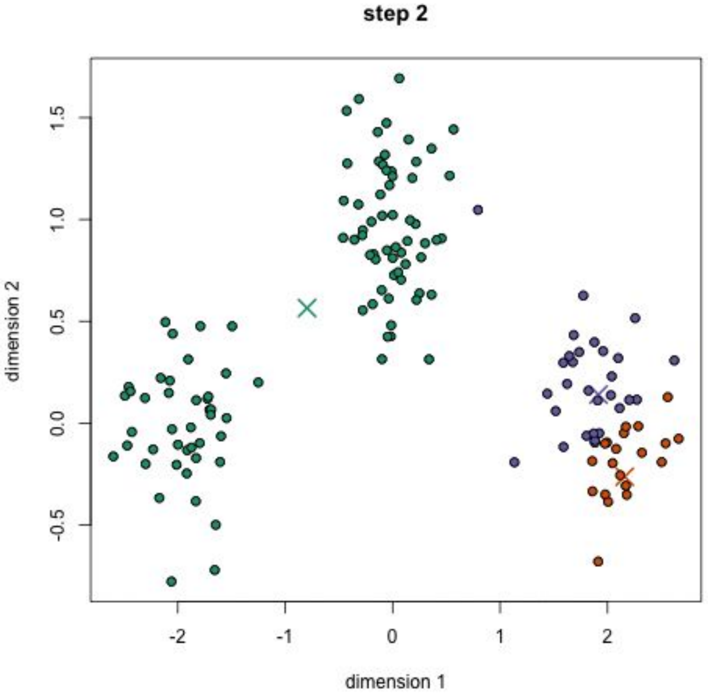
K-Means Clustering



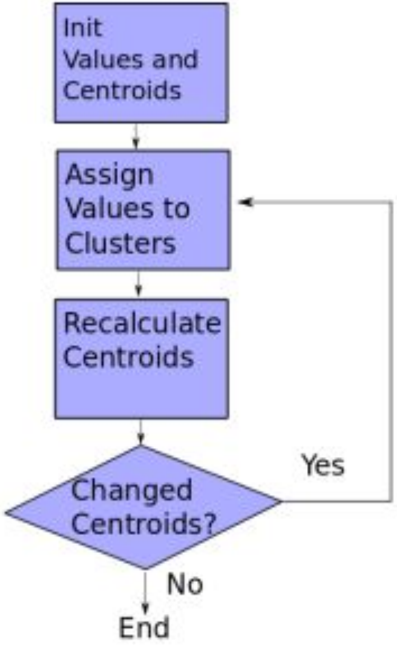


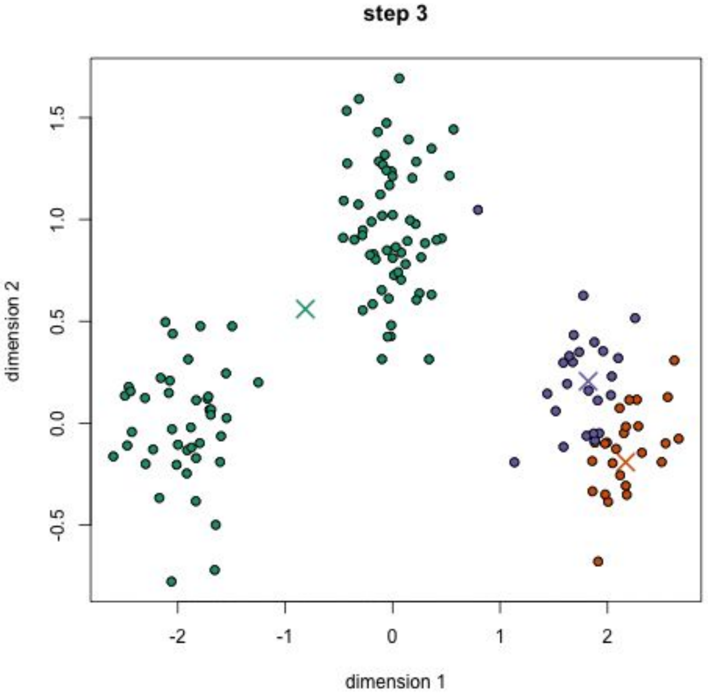
K-Means Clustering



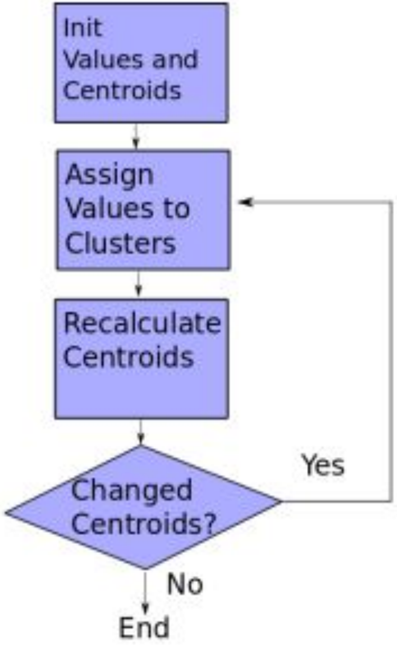


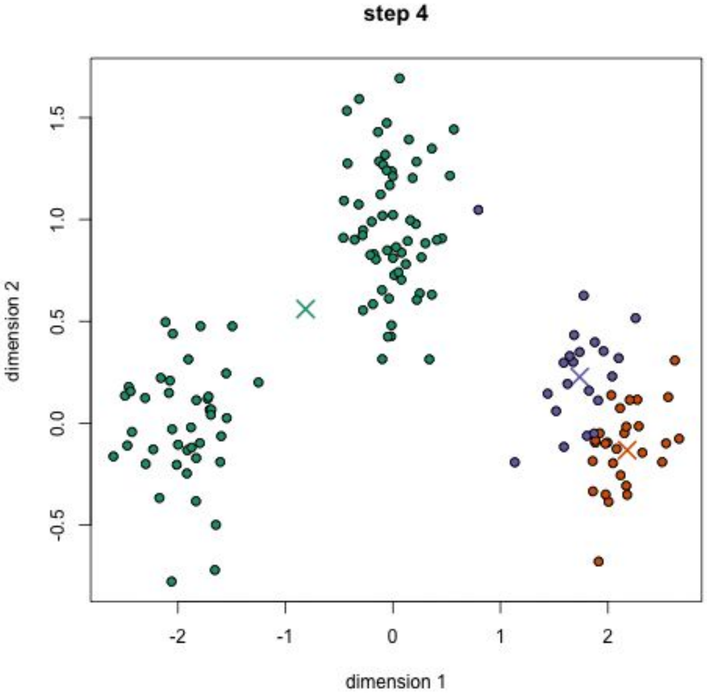
K-Means Clustering



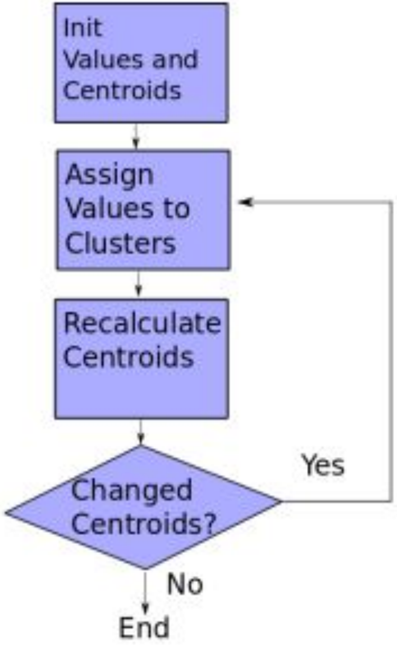


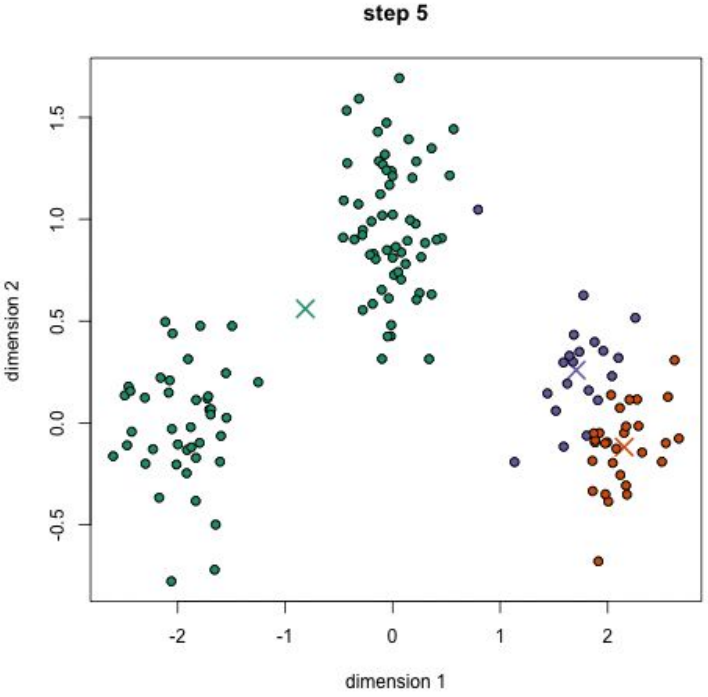
K-Means Clustering



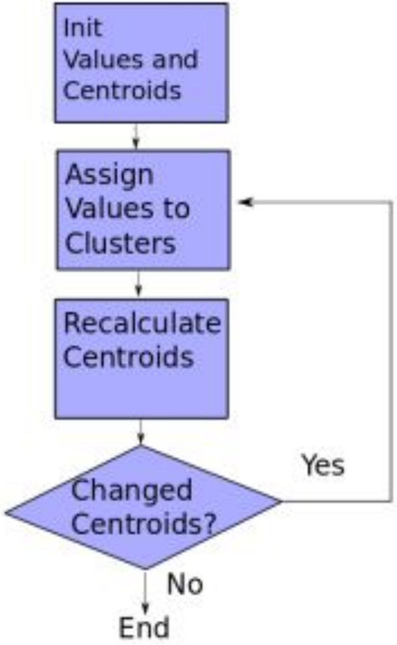


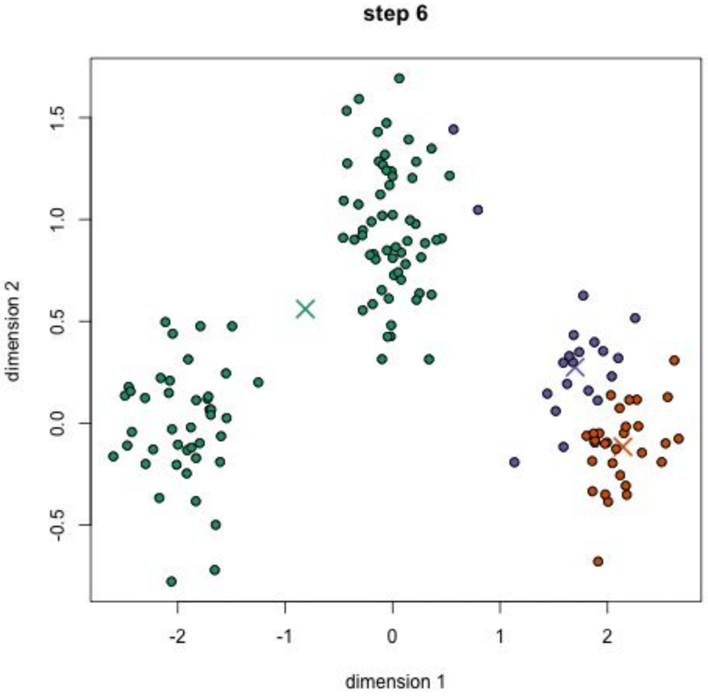
K-Means Clustering



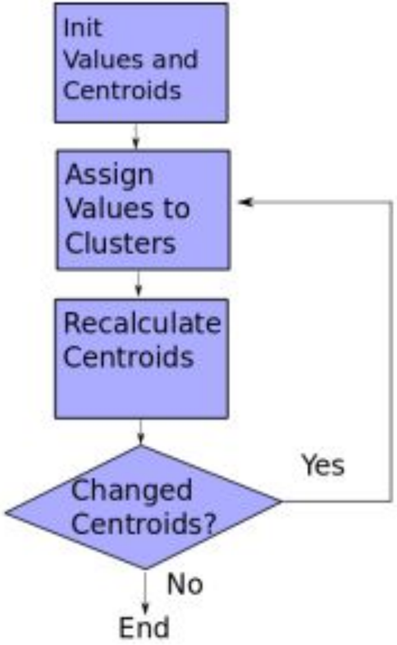


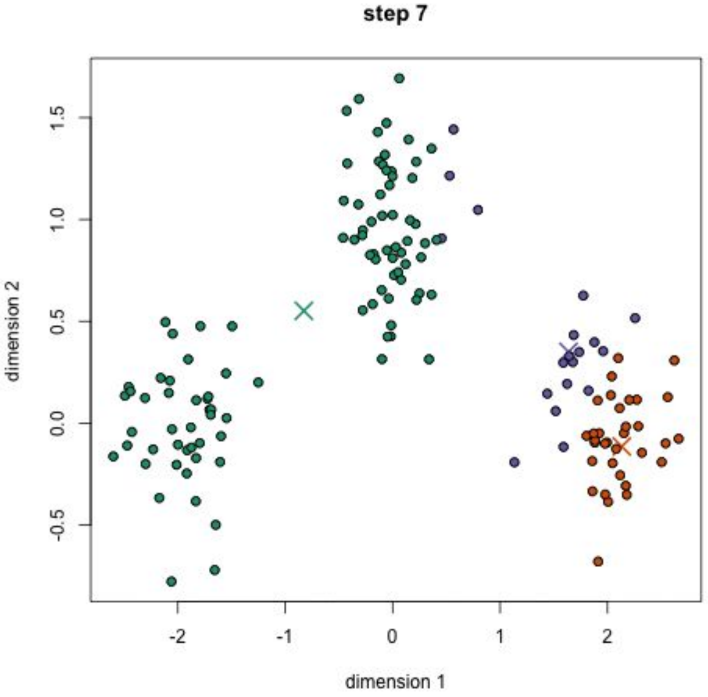
K-Means Clustering



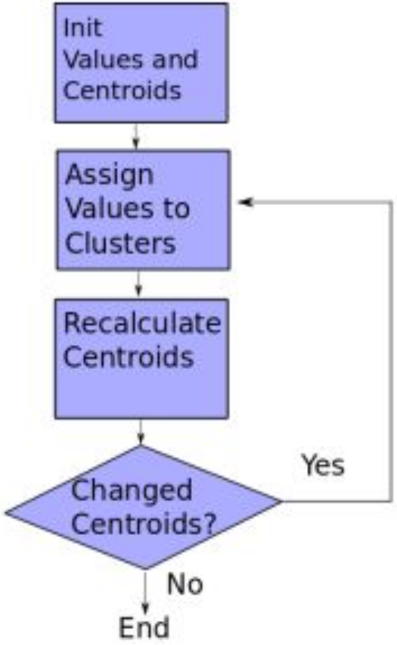


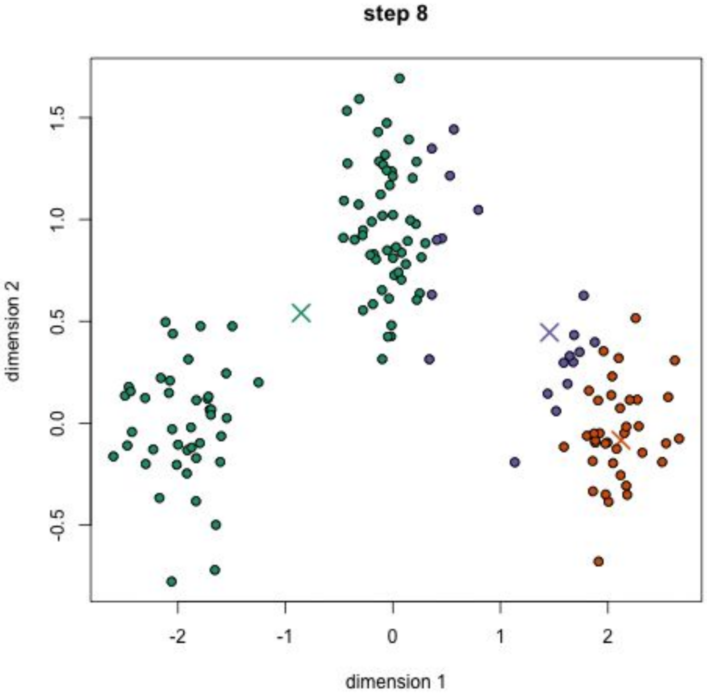
K-Means Clustering



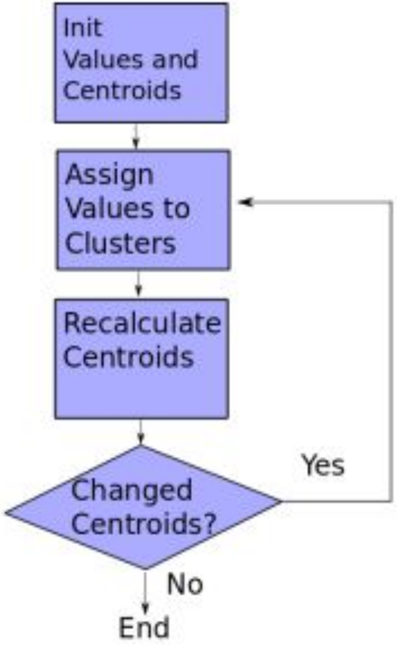


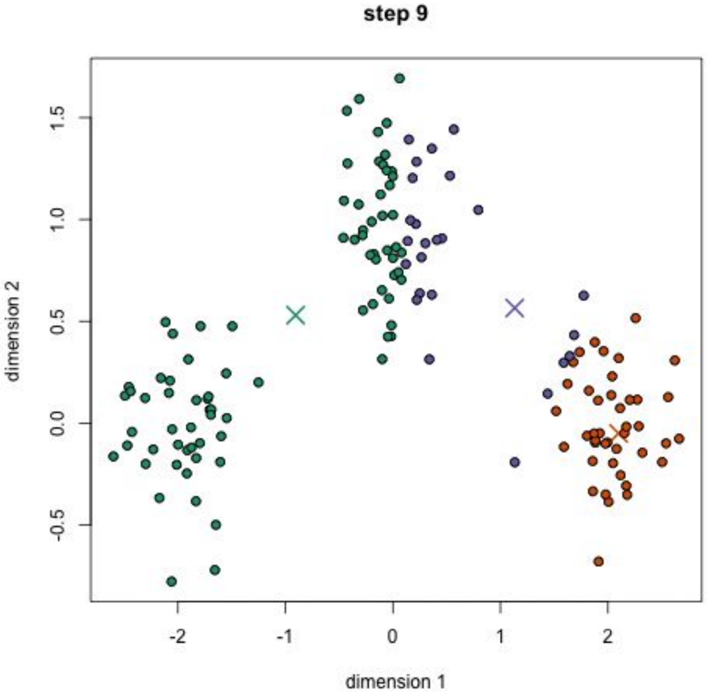
K-Means Clustering



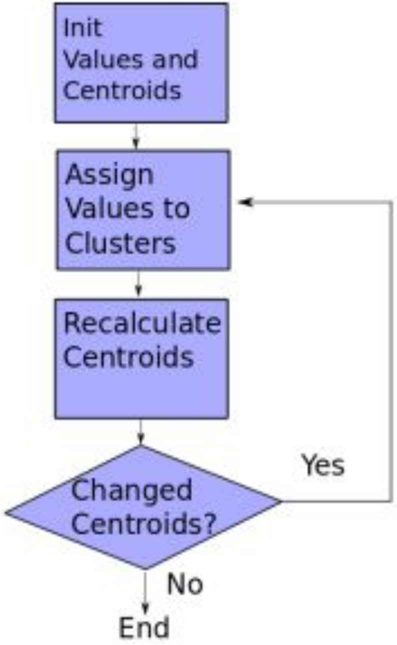


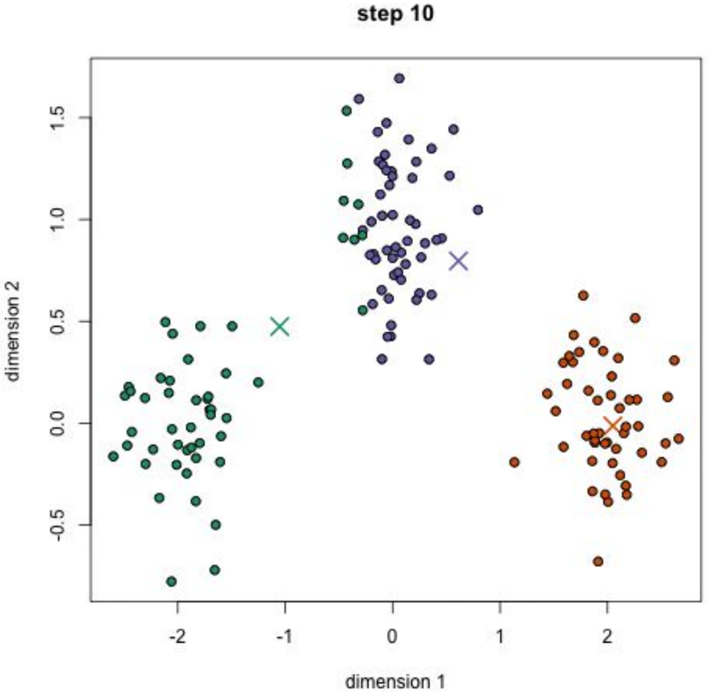
K-Means Clustering



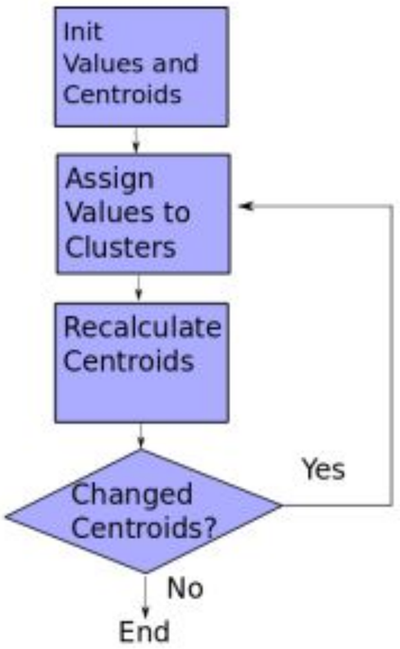


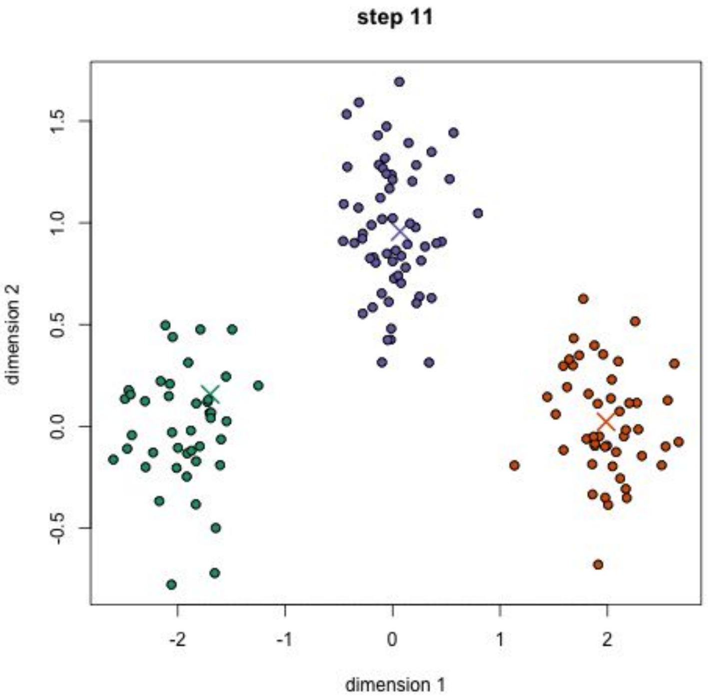
K-Means Clustering



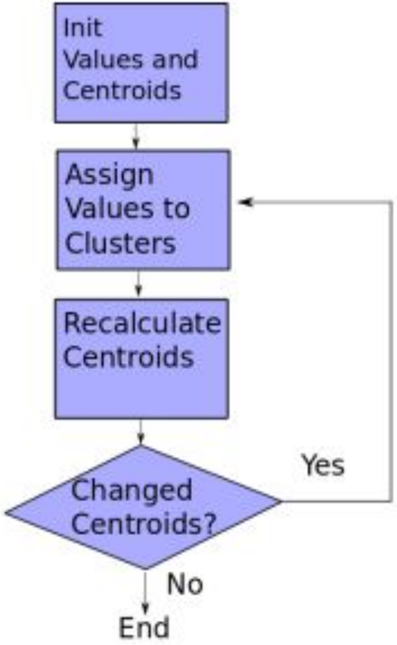


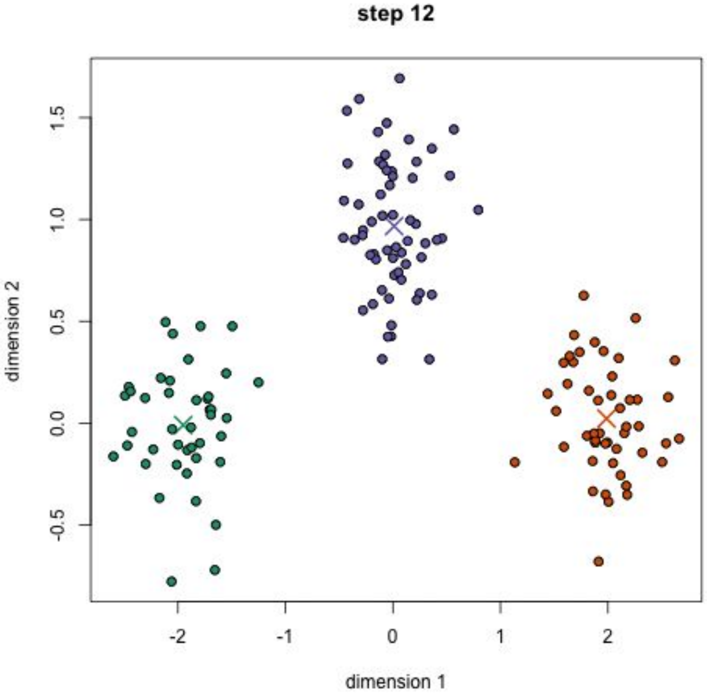
K-Means Clustering





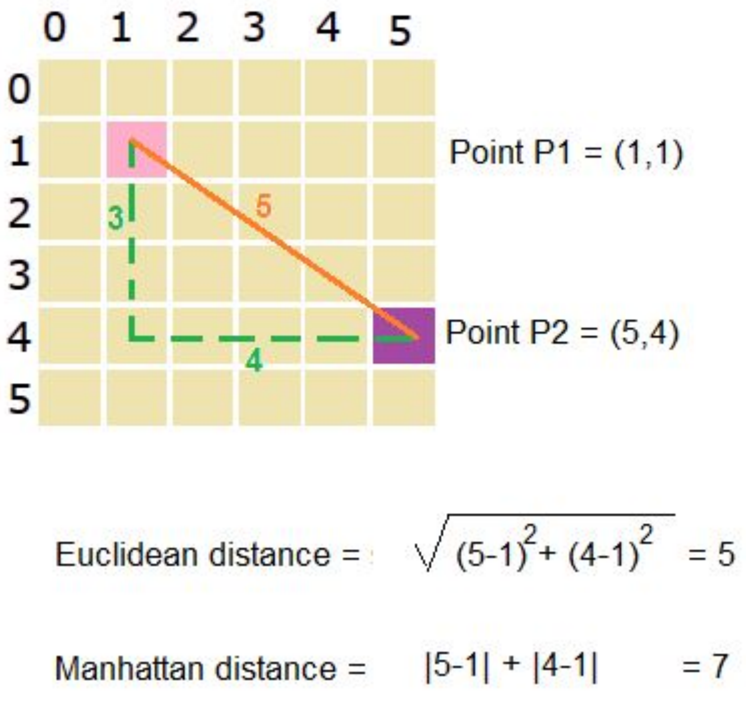
K-Means Clustering

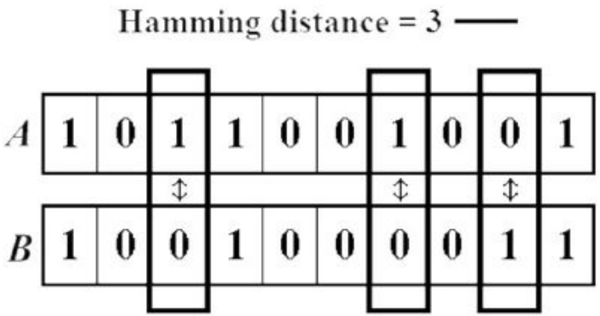




K-Means Clustering: Distance metrics

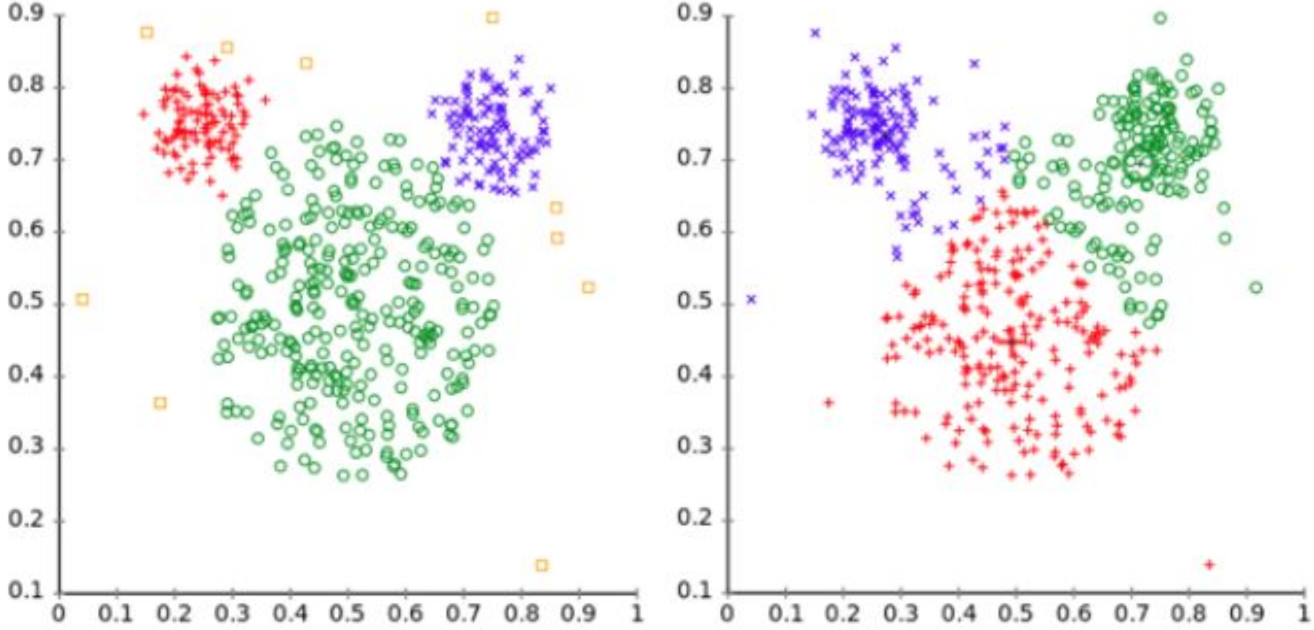
- Euclidean distance (most used)
- Manhattan distance
- Hamming distance





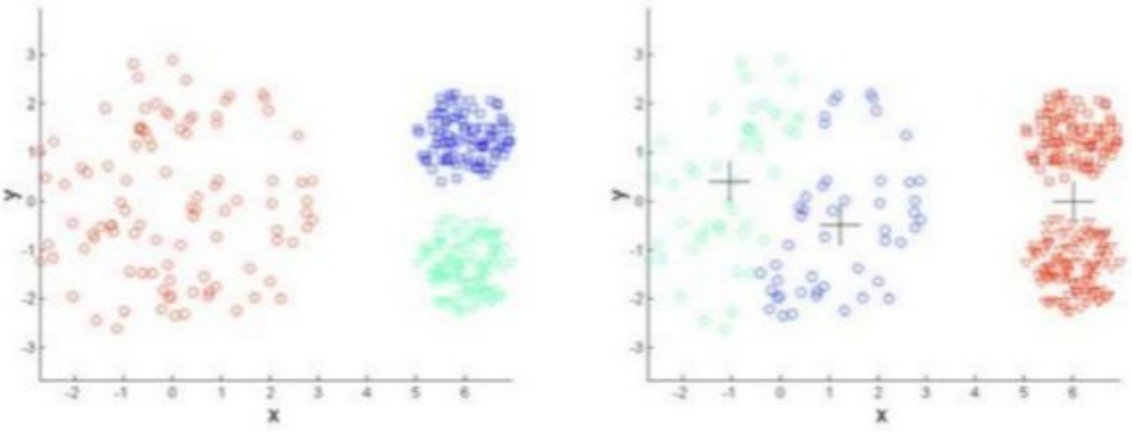
K-Means Clustering: Limitations

Clusters of different size



K-Means Clustering: Limitations

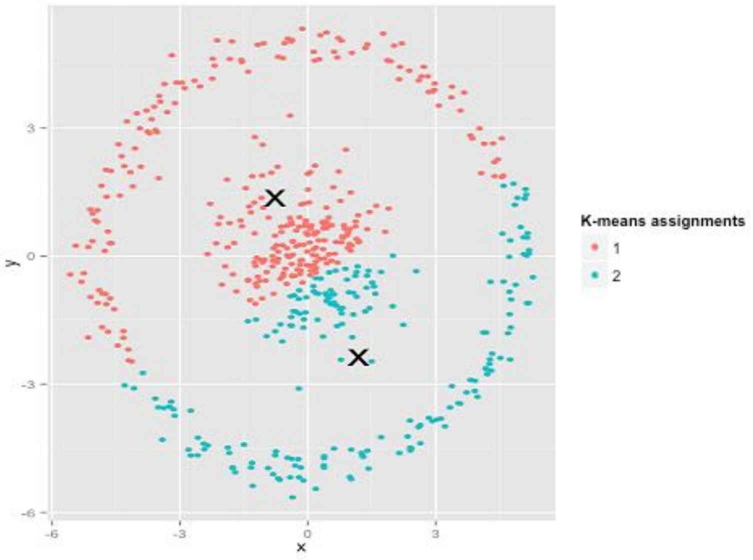
Clusters of different density

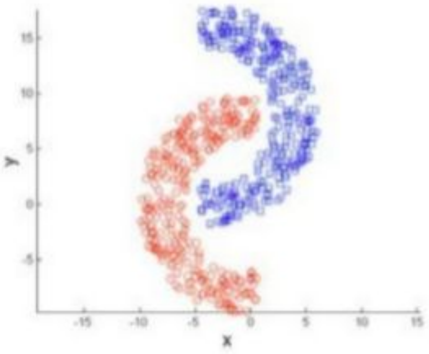


Original Points **K-means (3 Clusters)**

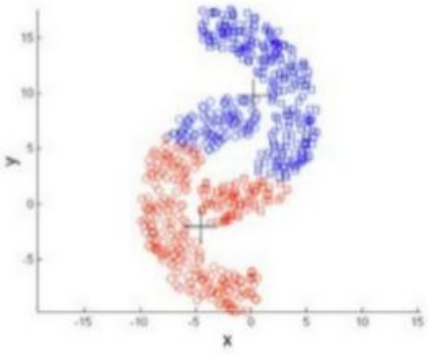
K-Means Clustering: Limitations

Clusters of non-spherical shape





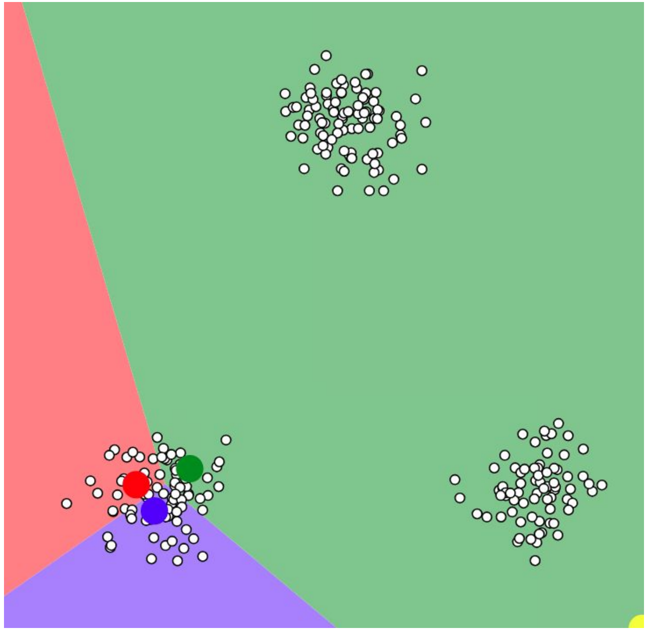
Original Points

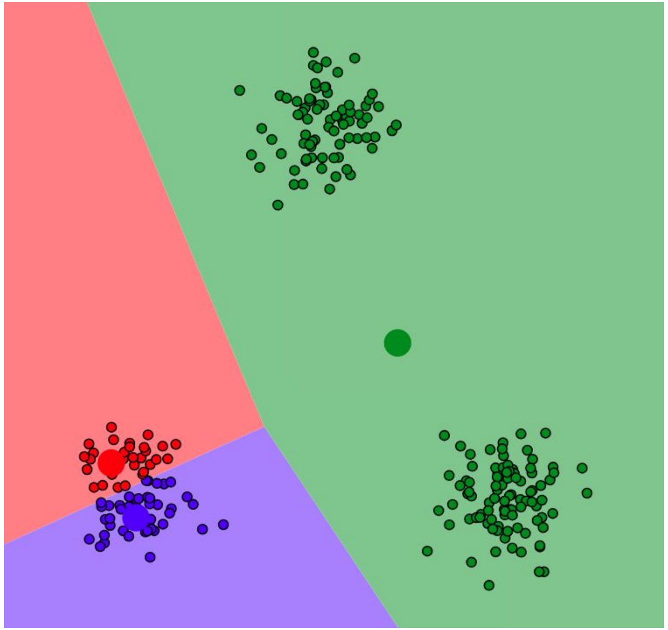


K-means (2 Clusters)

K-Means Clustering: Limitations

Choosing the initial centroids





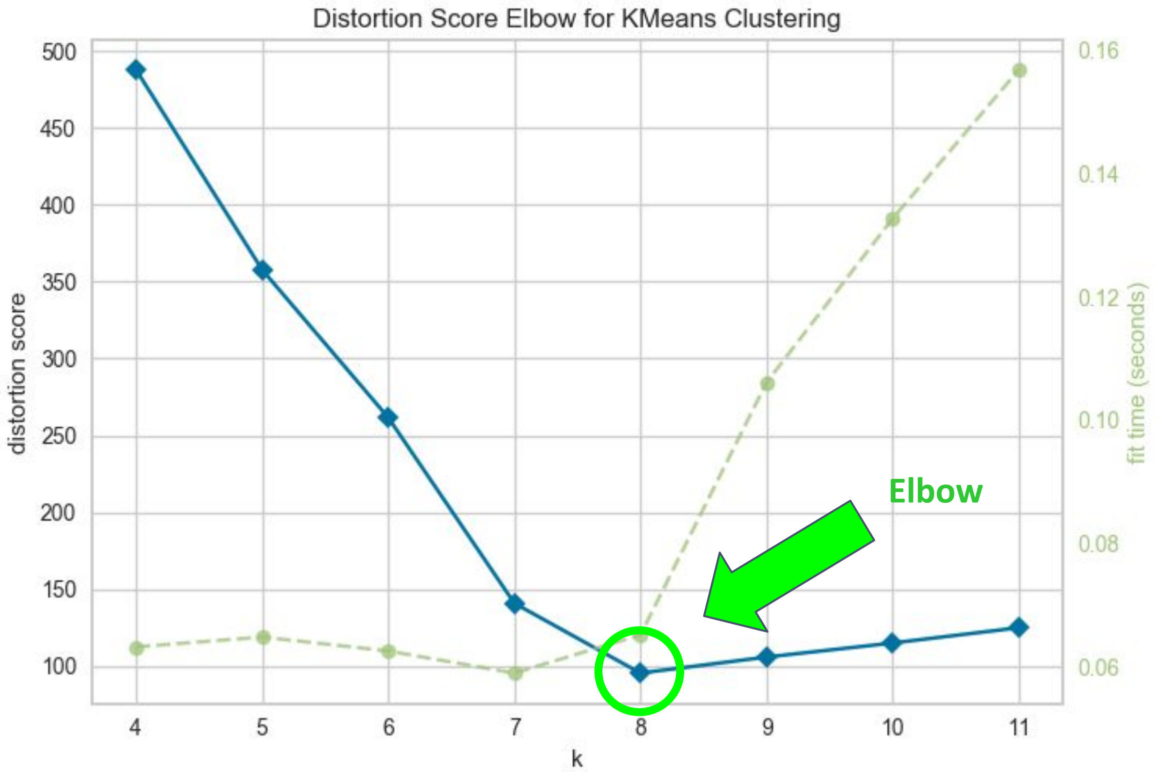
K-Means Clustering

Advantages	Disadvantages
Fast and efficient	How to choose k?
Simple and easy to understand	How to choose the initial centroids?
	Sensitive to outliers
	Bad at handling clusters of different size
	Bad at handling clusters of different density
	Bad at handling clusters of non-spherical shape

K-Means Clustering: Optimizations

- Choosing centroids
  - **K-means++**
    - Don’t choose randomly
    - Choose the next centroid with a probability proportional to a distance function
    - Default in scikit-learn
- Estimating k
  - **Elbow method**
  - Silhouette score
  - Calinski-Harabasz index
  - Cluster instability

K-Means Clustering: Elbow Method

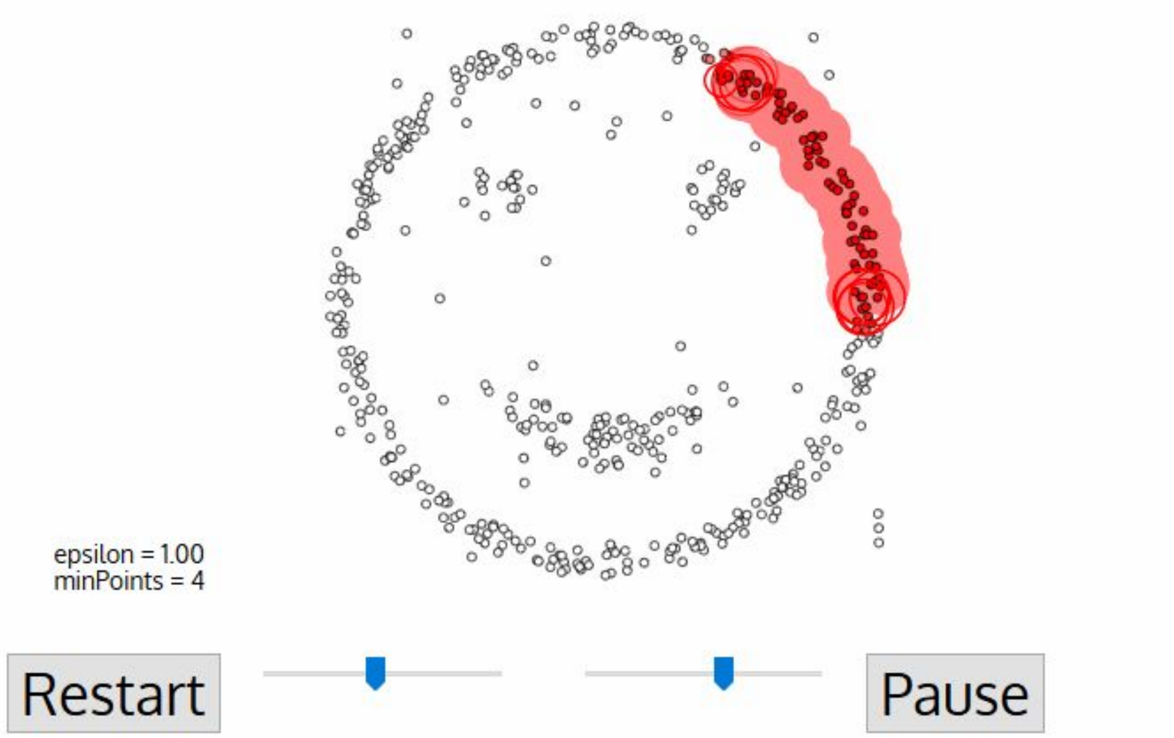


DBSCAN

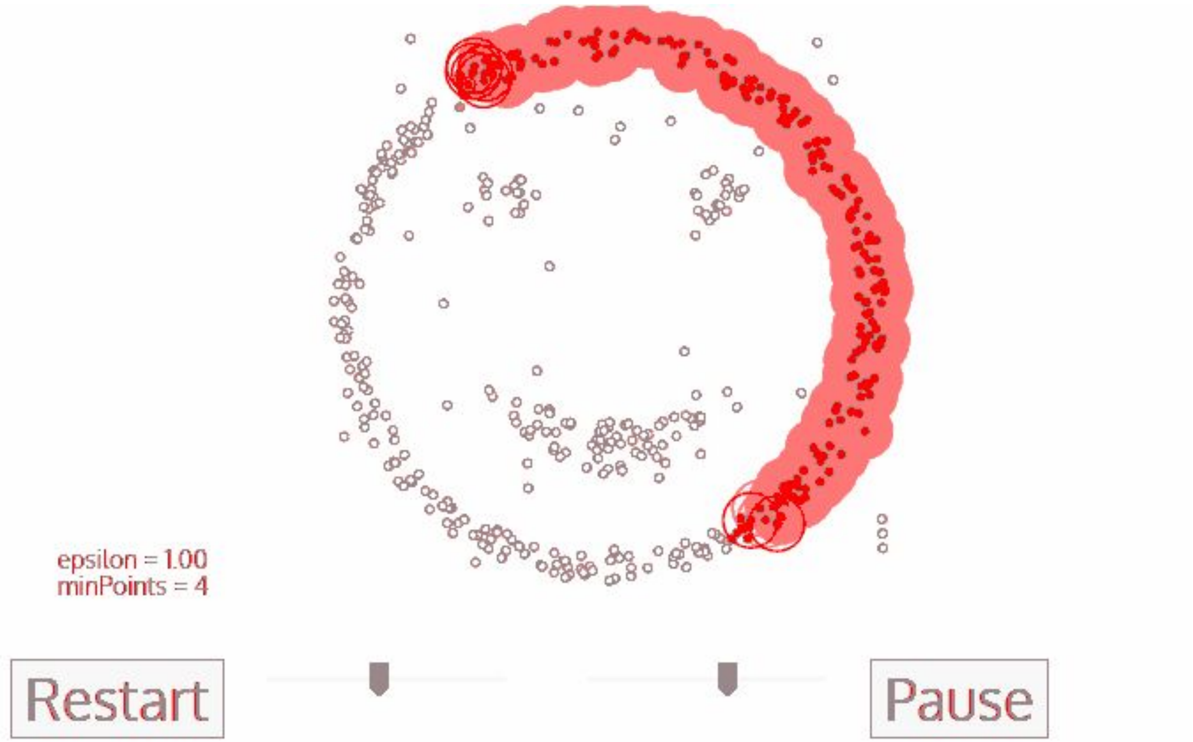
DBSCAN

- Density Based Clustering
- A cluster is a high-density area (there are no restrictions on its shape) surrounded by a low-density one
- Better at dealing with non-convex problems
- Doesn't need an initial estimate about the number of expected clusters
- More robust against noise

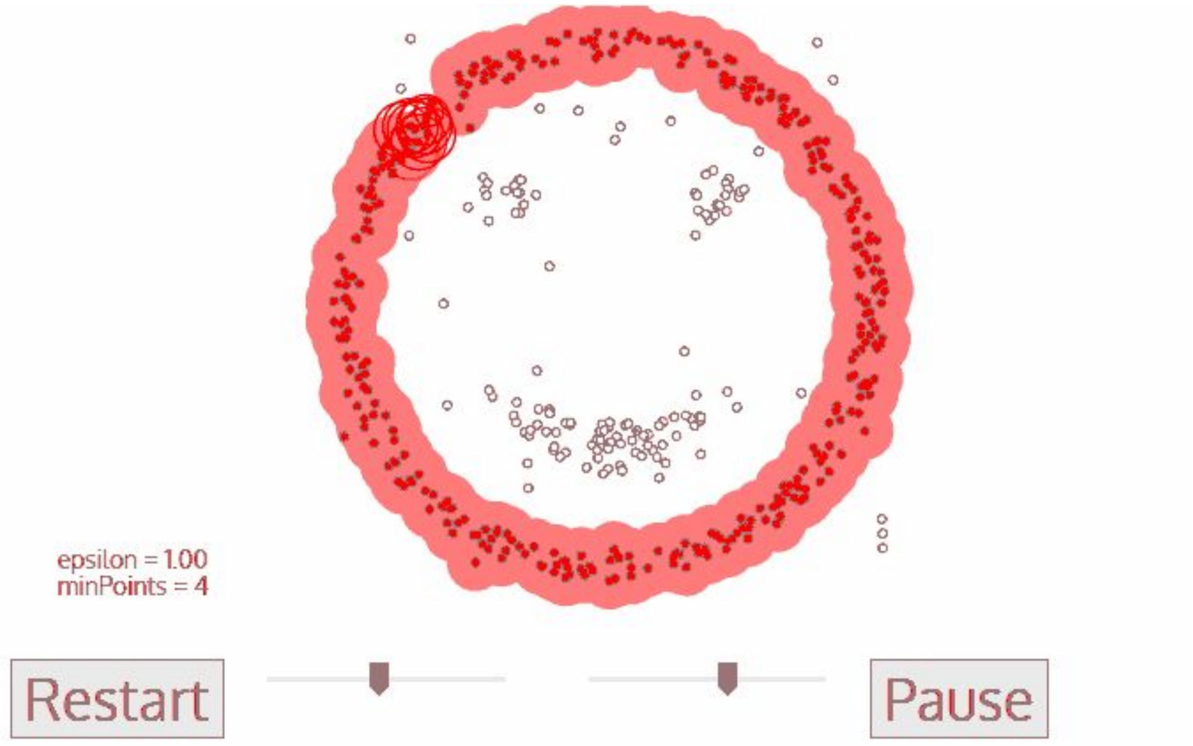
DBSCAN



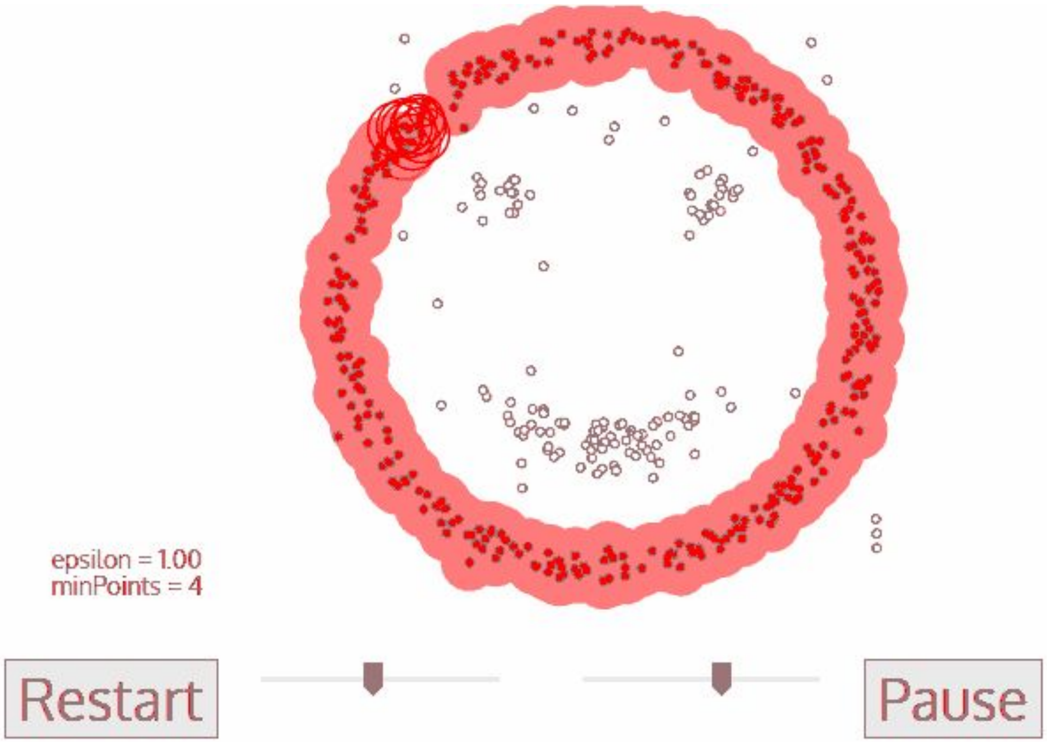
DBSCAN



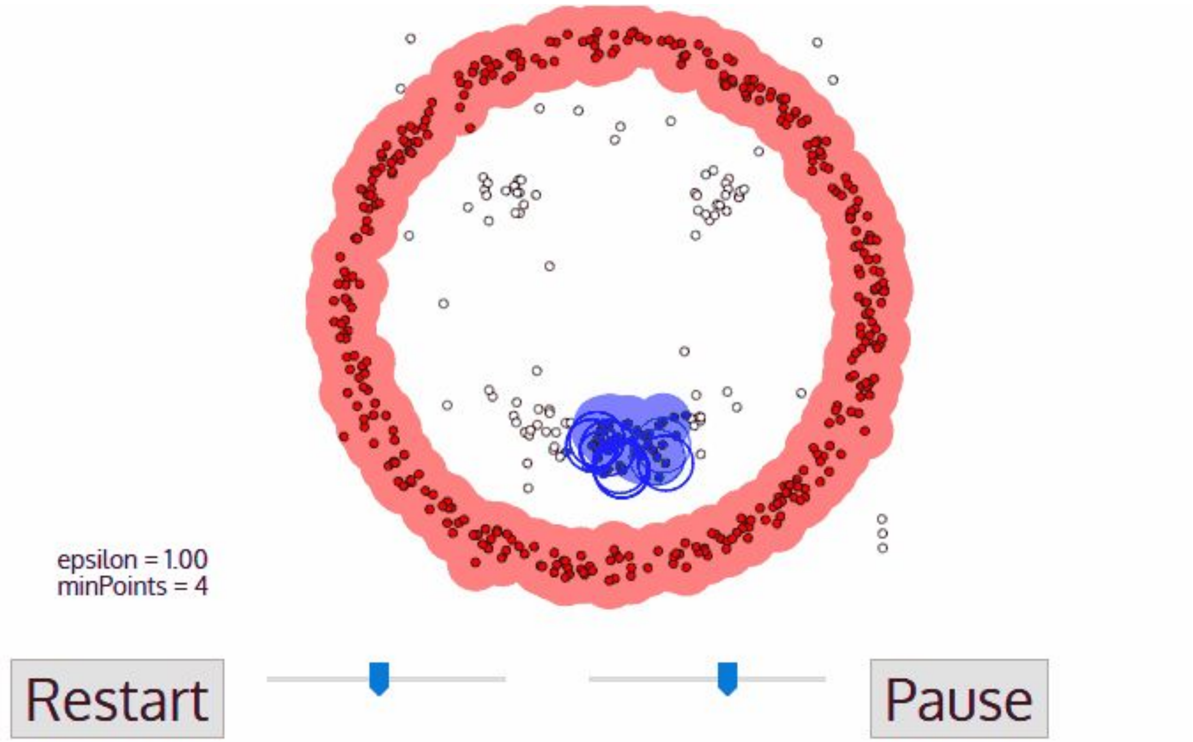
DBSCAN



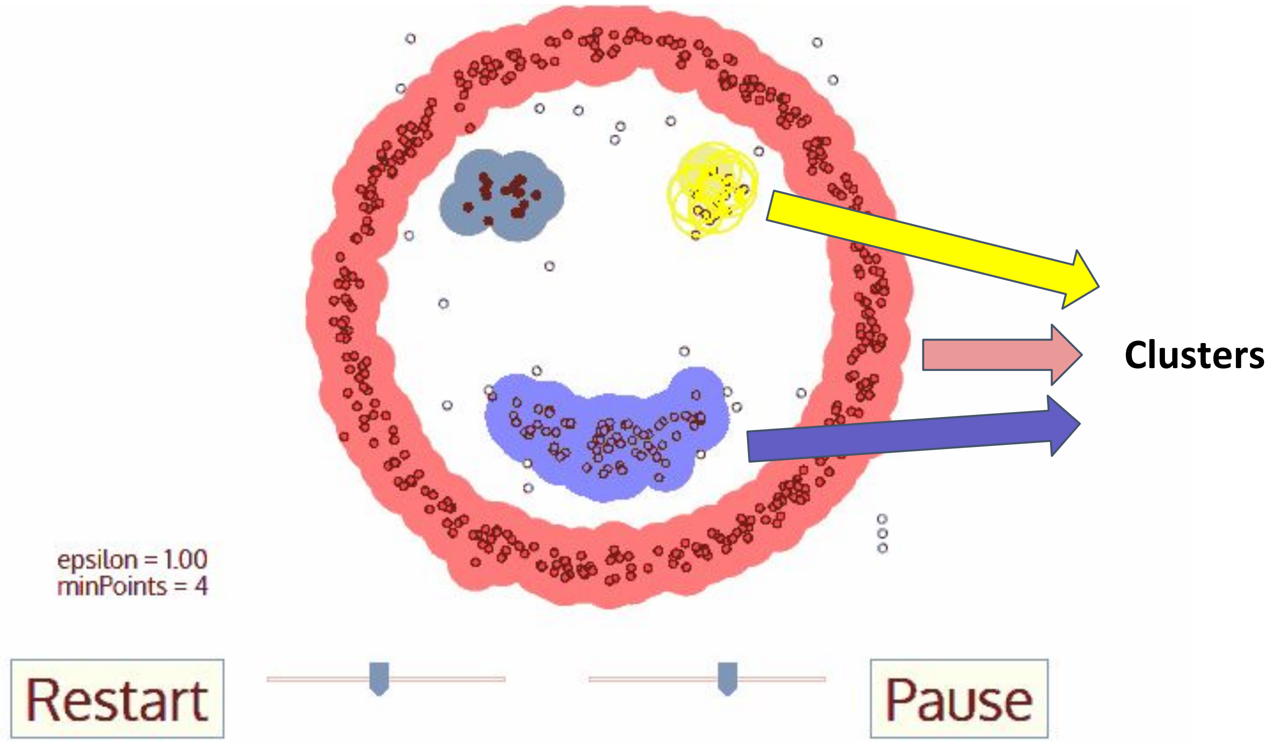
DBSCAN



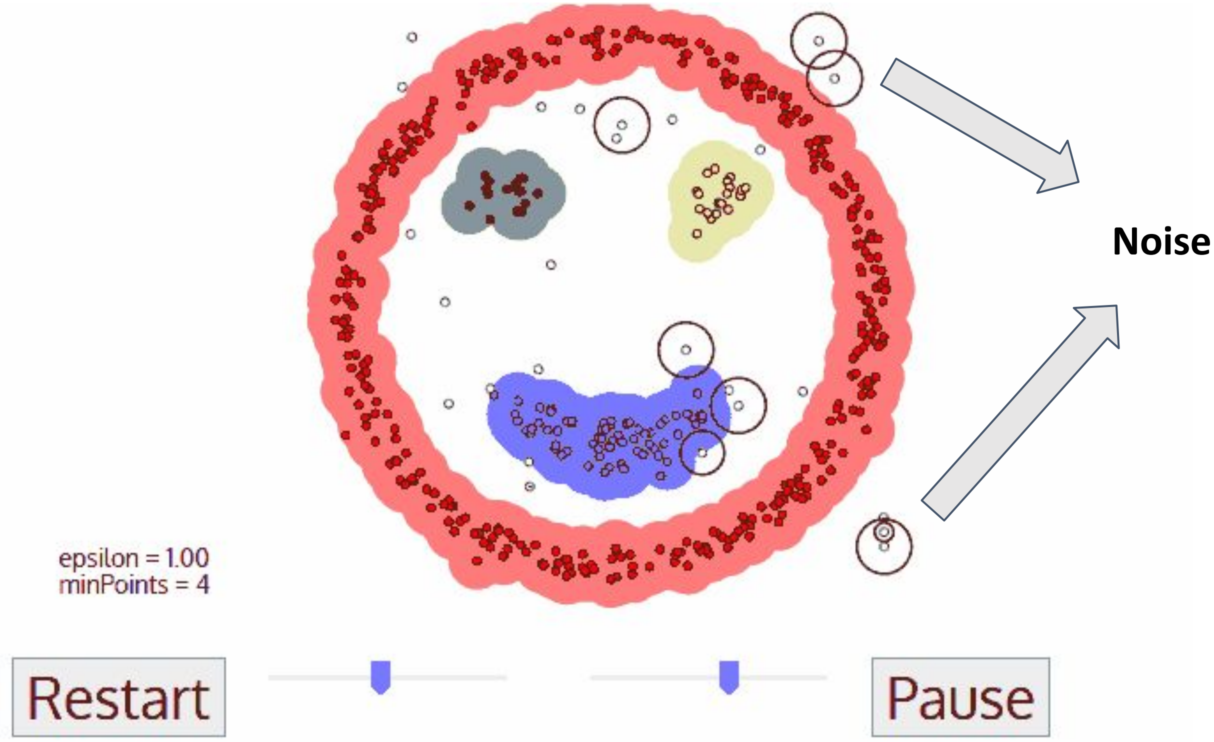
DBSCAN



DBSCAN



DBSCAN



Noise

DBSCAN

1. Start with an unvisited point
2. Check neighbourhood (based on parameter epsilon ε)
   a. If surrounded by a minimum number of other samples (parameter min_samples)

      => Part of cluster

   b. Else:

      => Noise

   => In both cases this point is marked as “visited”
3. Check neighbours

   => If they also have a high density, they are merged with the first area
4. If point is not in a cluster, continue with a new unvisited data point

DBSCAN

Parameters:

- eps : maximum distance between two neighbors.

  => Higher values will aggregate more points

  => Smaller values will create more clusters.
- min_samples : how many surrounding points are necessary to define an area

DBSCAN

Advantages	**Disadvantages**
No pre-set number of clusters required	Choosing ϵ
Robust against noise	Choosing minPoints
	Bad at handling clusters of different density

DBSCAN: OPTICS

Ordering Points to Identify Cluster Structure

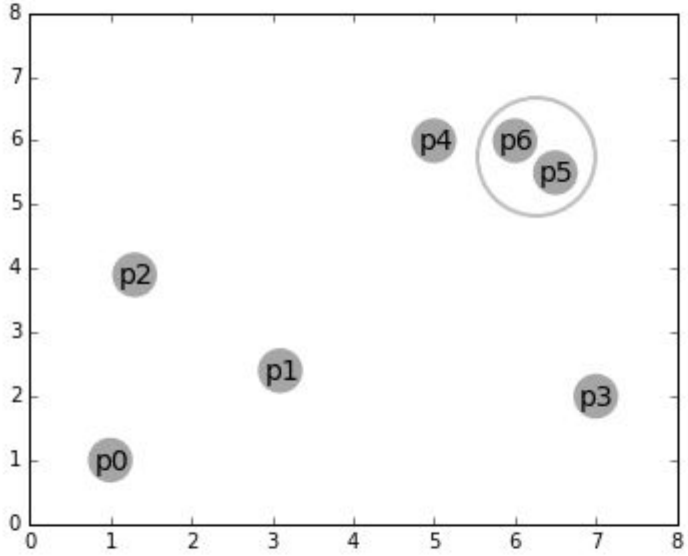
- Generalizes DBSCAN to clusters of different densities
- Epsilon becomes an optional maximum

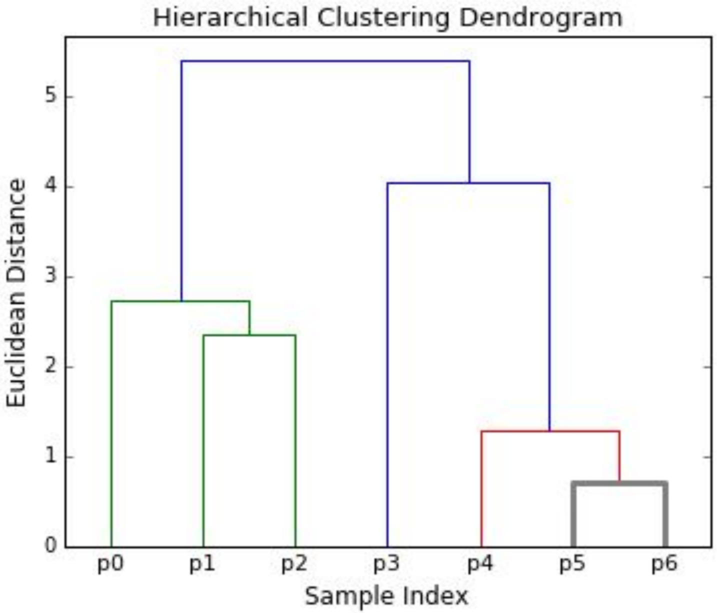
=> Exact implementation falls outside the extent of this course

=> Important to know there is an optimization possible (check extra material on github)

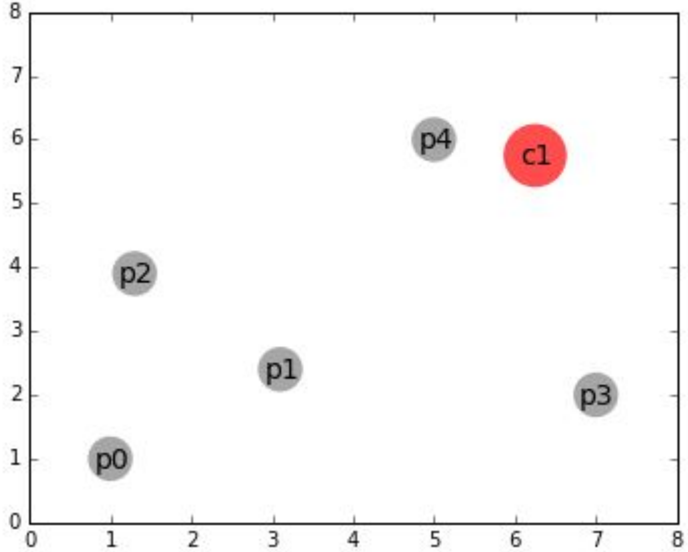
Hierarchical clustering

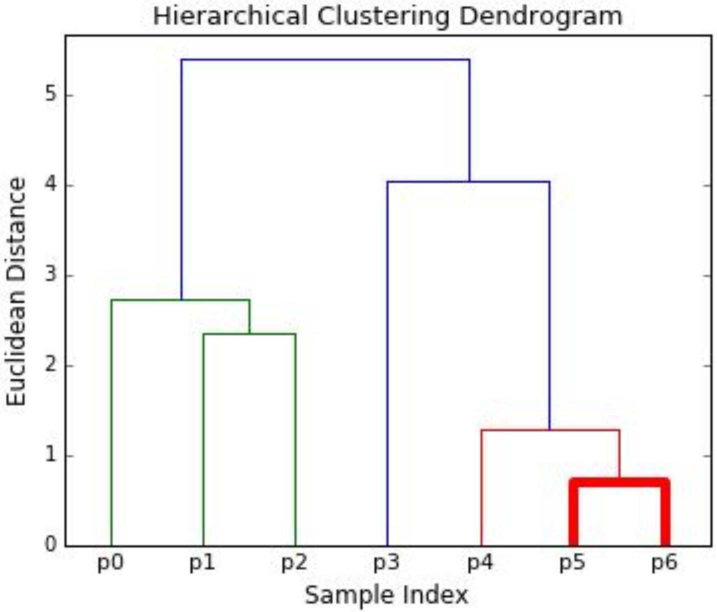
Hierarchical clustering



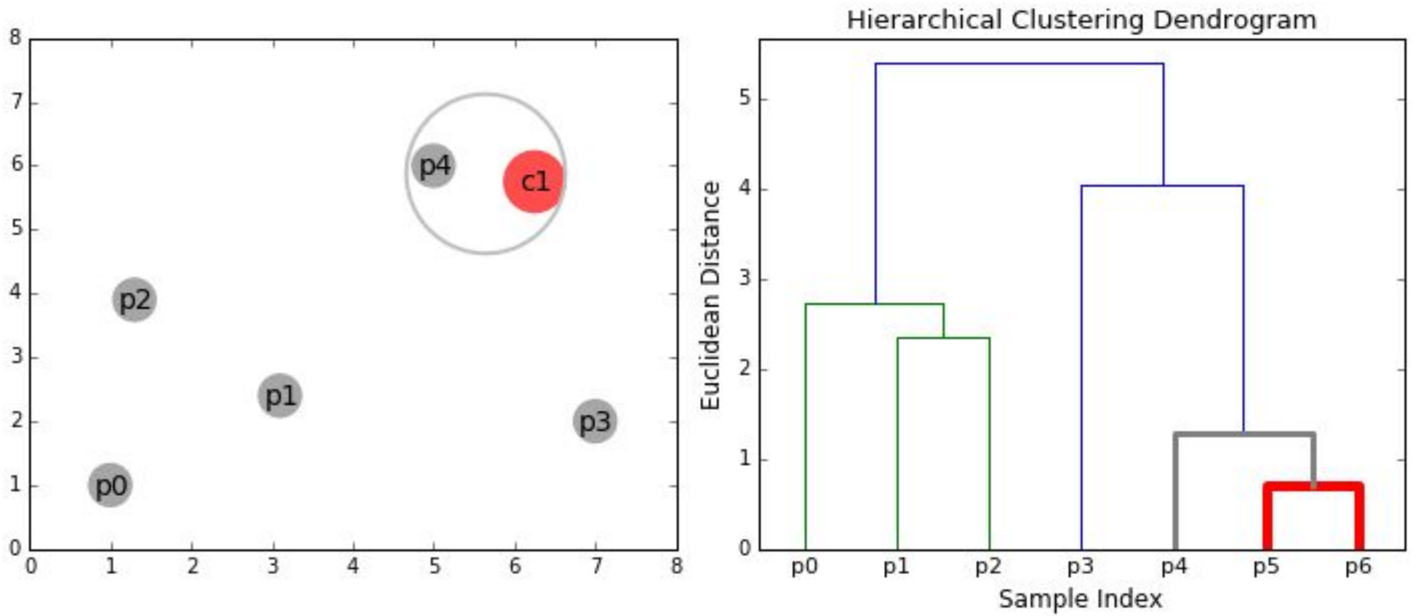


Hierarchical clustering

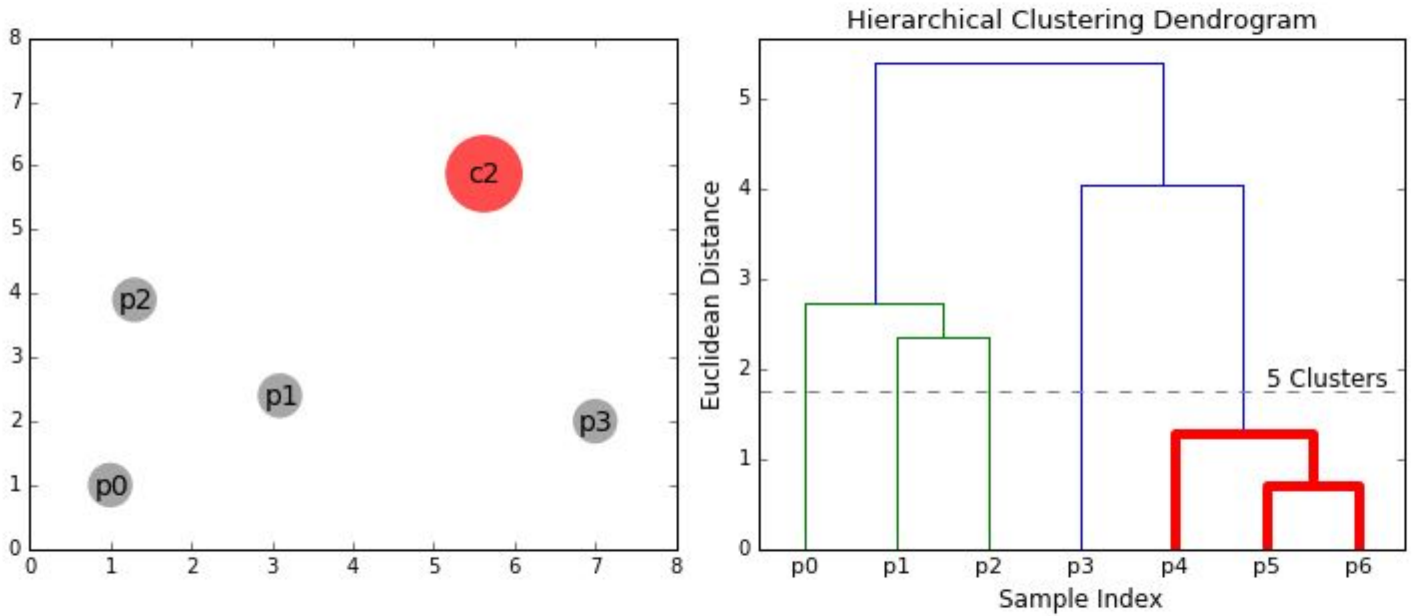




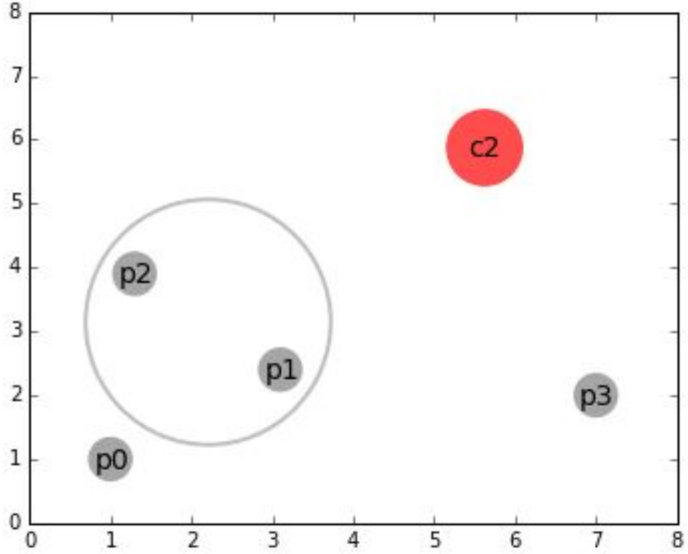
Hierarchical clustering

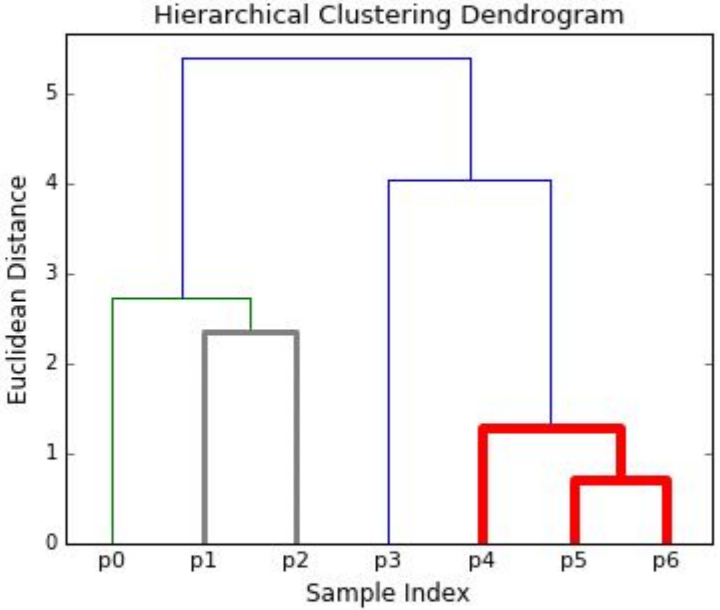


Hierarchical clustering

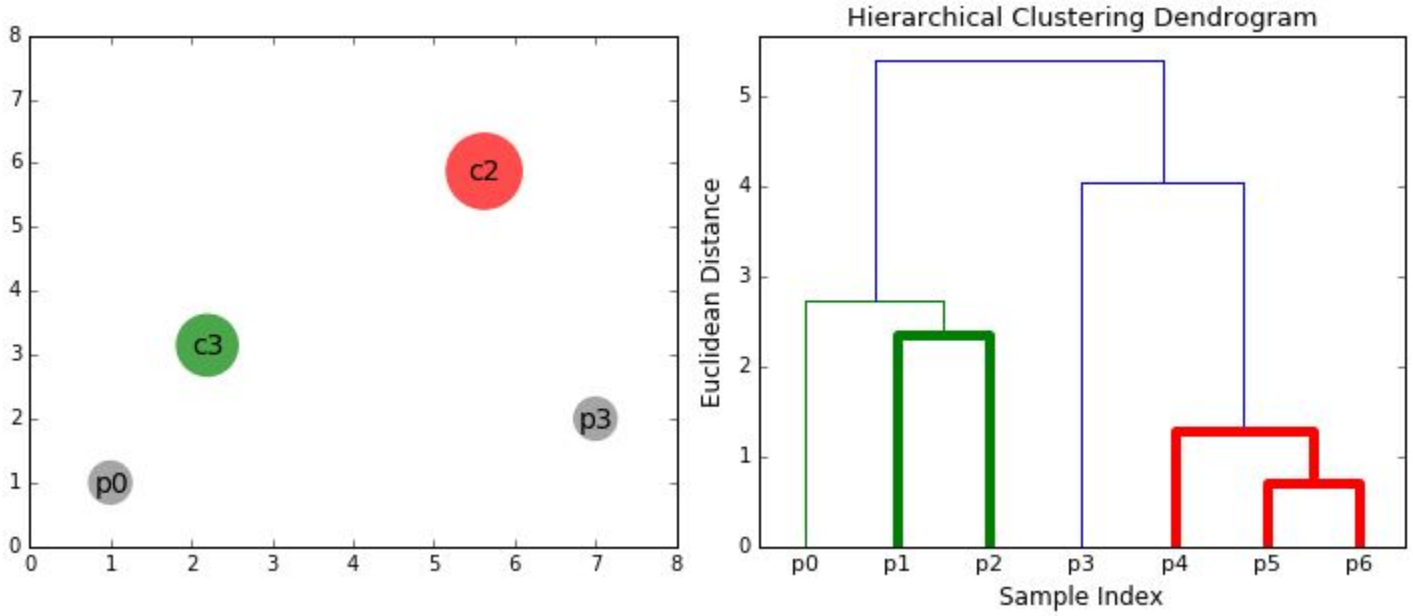


Hierarchical clustering

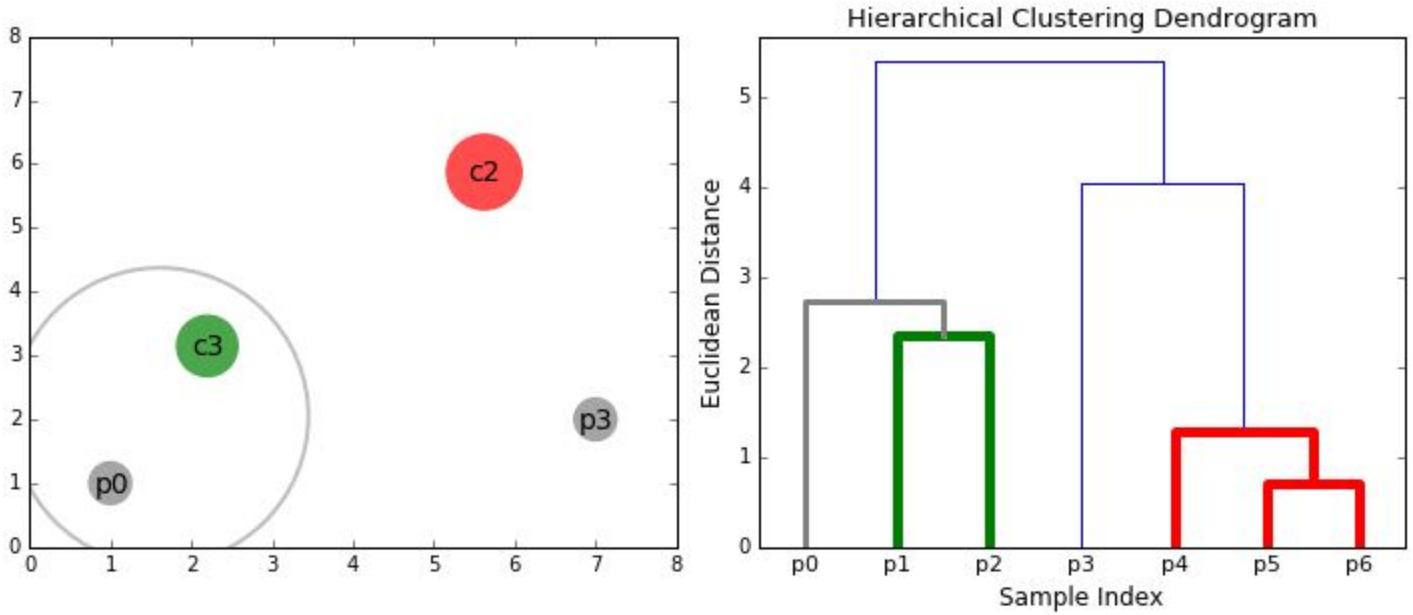




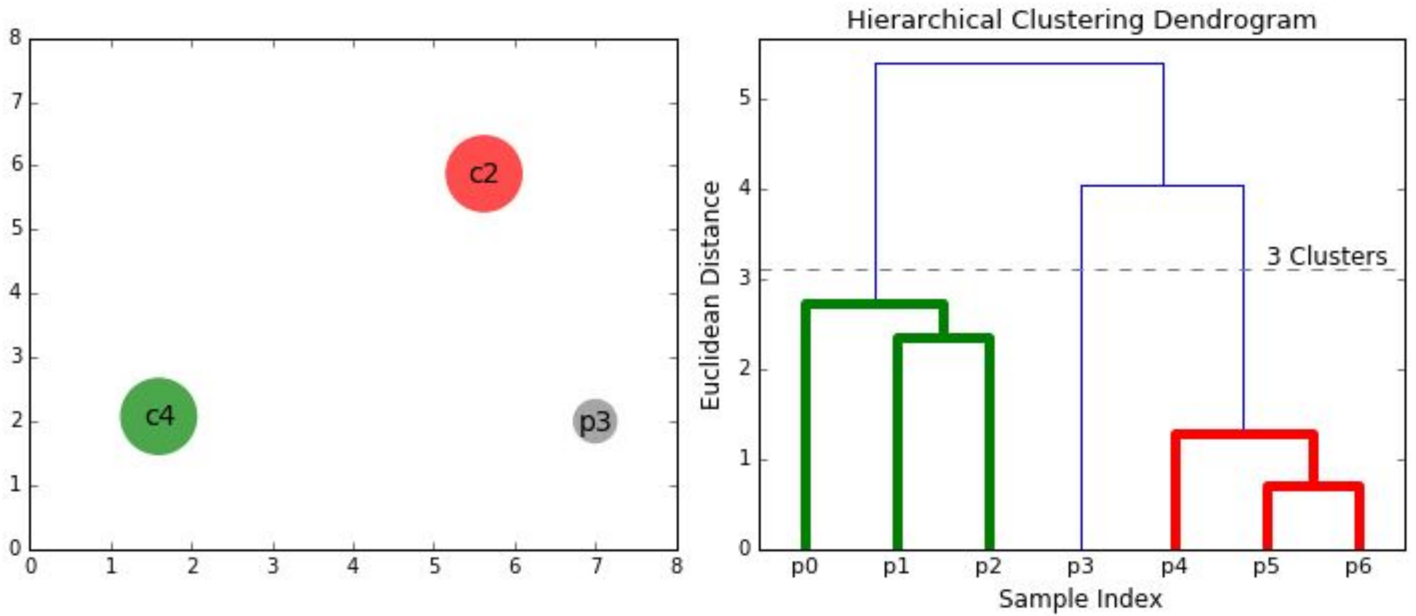
Hierarchical clustering



Hierarchical clustering



Hierarchical clustering



Hierarchical clustering

1. Start with each data point as a single cluster

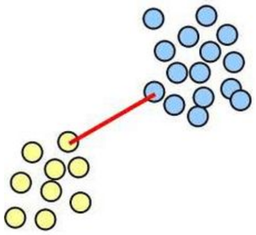
   X data points => X clusters

2. Each iteration combines two clusters into one
   => Choose clusters with smallest average linkage according to distance metric
3. Repeat until one cluster contains all data points

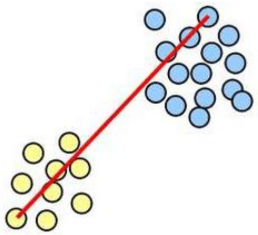
=> Select how many clusters we want in the end

Hierarchical clustering: Linkage

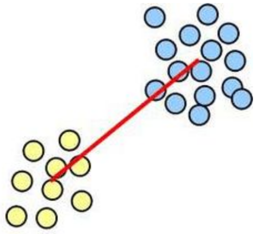
- **Single linkage:** minimizes the distance between the closest observations of pairs of clusters.
- **Maximum or complete linkage:** minimizes the maximum distance between observations of pairs of clusters.
- **Average linkage:** minimizes the average of the distances between all observations of pairs of clusters.
- **Ward:** minimizes the sum of squared differences within all clusters. It is a variance-minimizing approach (similar to the k-means). If you join 2 clusters, minimize the total distance to the new centroid



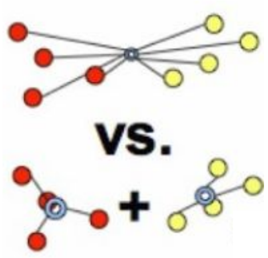
single-link



complete-link

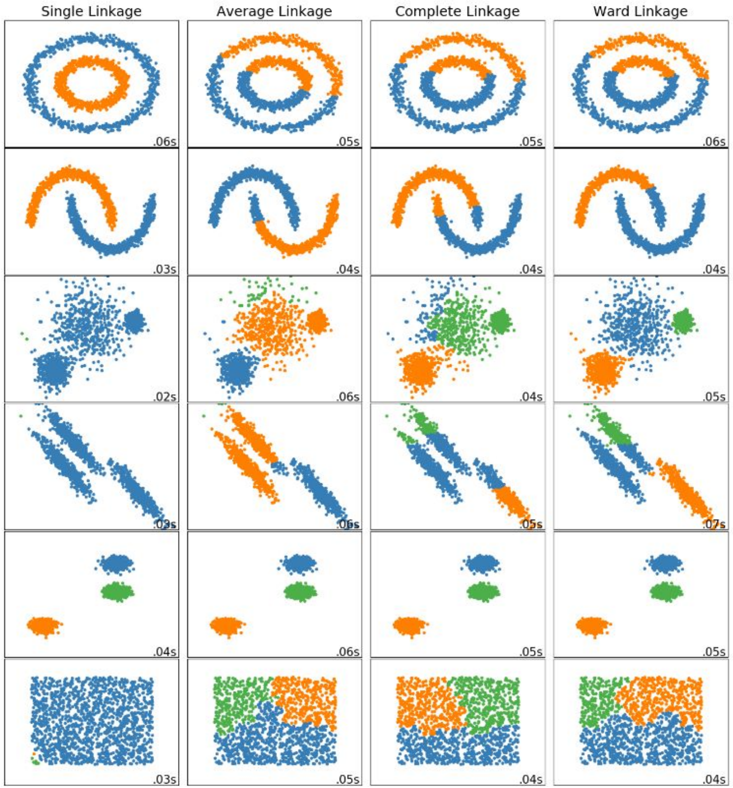


average-link



Ward

Hierarchical clustering: Linkage



Hierarchical clustering

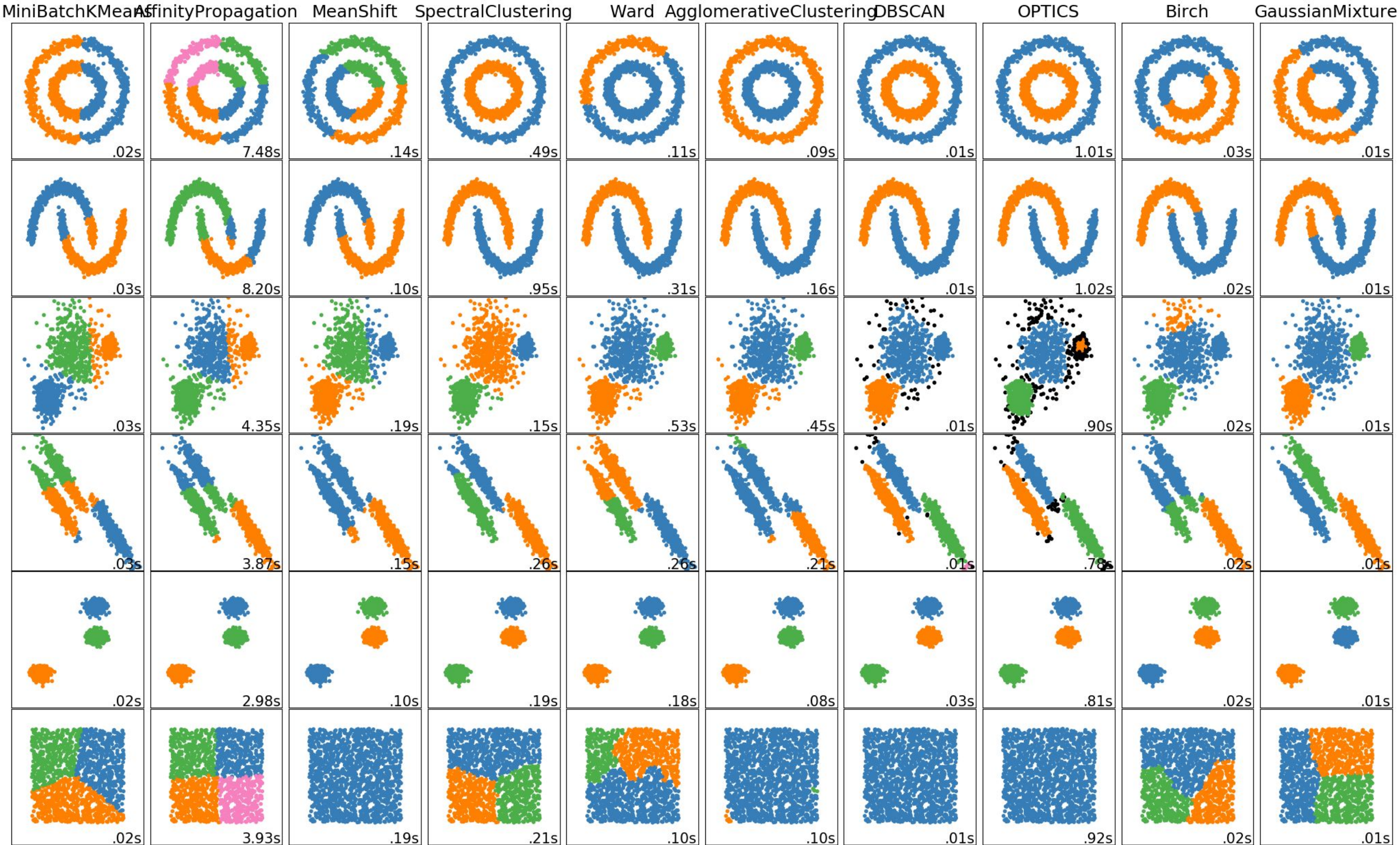
Advantages	Disadvantages
Easy to visualize with the dendrogram	Difficult to identify the correct number of clusters from the dendrogram
Provides hierarchical relations between clusters	Can be sensitive to noise and outliers based on linkage
No a priori information about the number of clusters required	Data points may be incorrectly grouped at an early stage
Less influenced by cluster shapes and densities	Slow

Clustering: Evaluation

- **Distortion/Inertia score:** Computes the sum of squared distances from each point to its assigned center

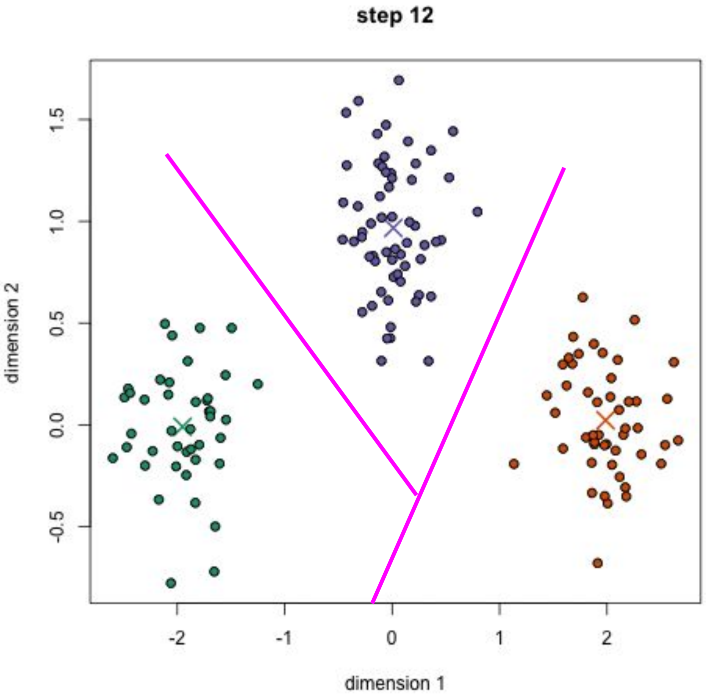
  => Assumes convex clusters (only really useful for k-Means)
- **Silhouette score:**
  - Measures the distance between each data point, the centroid of the cluster it was assigned to and the closest centroid belonging to another cluster
  - A measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation)
  - [-1, 1]

RANGE OF SC	*INTERPRETATION*
0.71-1.0	A strong structure has been found
0.51-0.70	A reasonable structure has been found
0.26-0.50	The structure is weak and could be artificial. Try additional methods of data analysis.
£ 0.25	No substantial structure has been found



MiniBatchKMeans
AffinityPropagation
MeanShift
SpectralClustering
Ward
AgglomerativeClustering
DBSCAN
OPTICS
Birch
GaussianMixture
.02s
7.48s
.14s
.49s
.11s
.09s
.01s
1.01s
.03s
.01s
.03s
8.20s
.10s
.95s
.31s
.16s
.01s
1.02s
.02s
.01s
.03s
4.35s
.19s
.15s
.53s
.45s
.01s
.90s
.02s
.01s
.03s
3.87s
.15s
.26s
.26s
.21s
.01s
.78s
.02s
.01s
.02s
2.98s
.10s
.19s
.18s
.08s
.03s
.81s
.02s
.01s
.02s
3.93s
.19s
.21s
.10s
.10s
.01s
.92s
.02s
.01s

From unsupervised to supervised



Once clusters are identified:

- Label the datapoints in the same cluster with the same label
- Apply supervised learning to new datapoints