



## MACHINE LEARNING



### SUPERVISED LEARNING



### UNSUPERVISED LEARNING



#### CLASSIFICATION



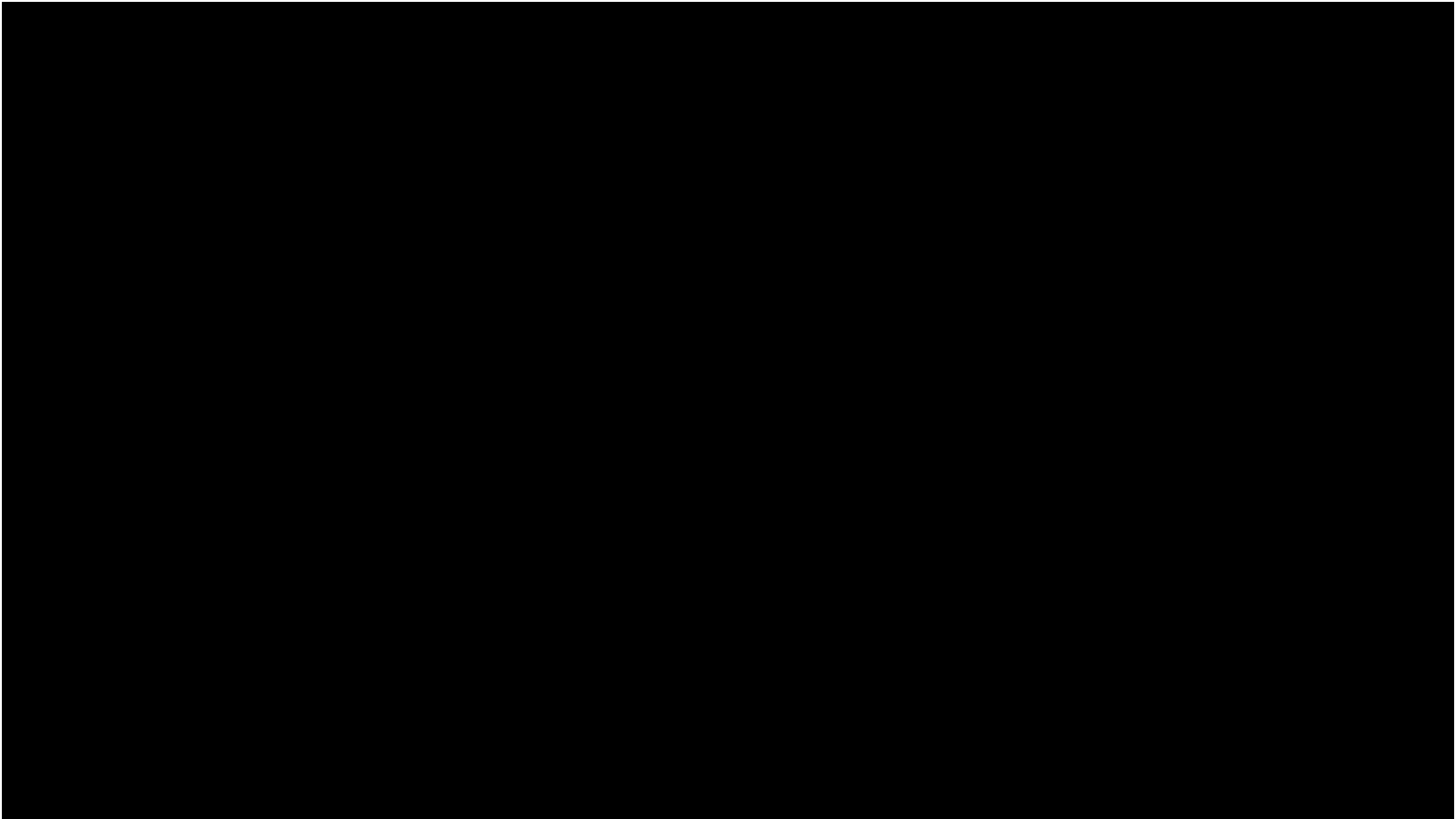
#### REGRESSION



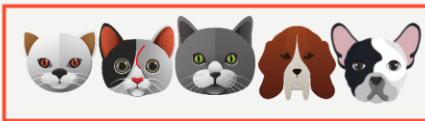
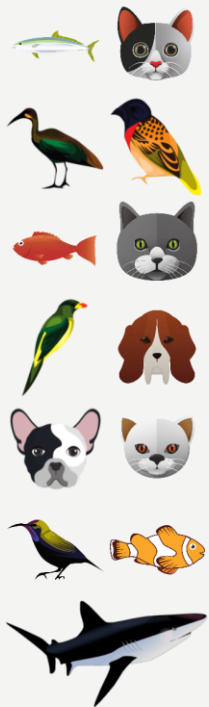
#### CLUSTERING



#### ASSOCIATION



No labels



# Unsupervised learning

Beschrijven van en zoeken naar structuur in niet-gelabelde data

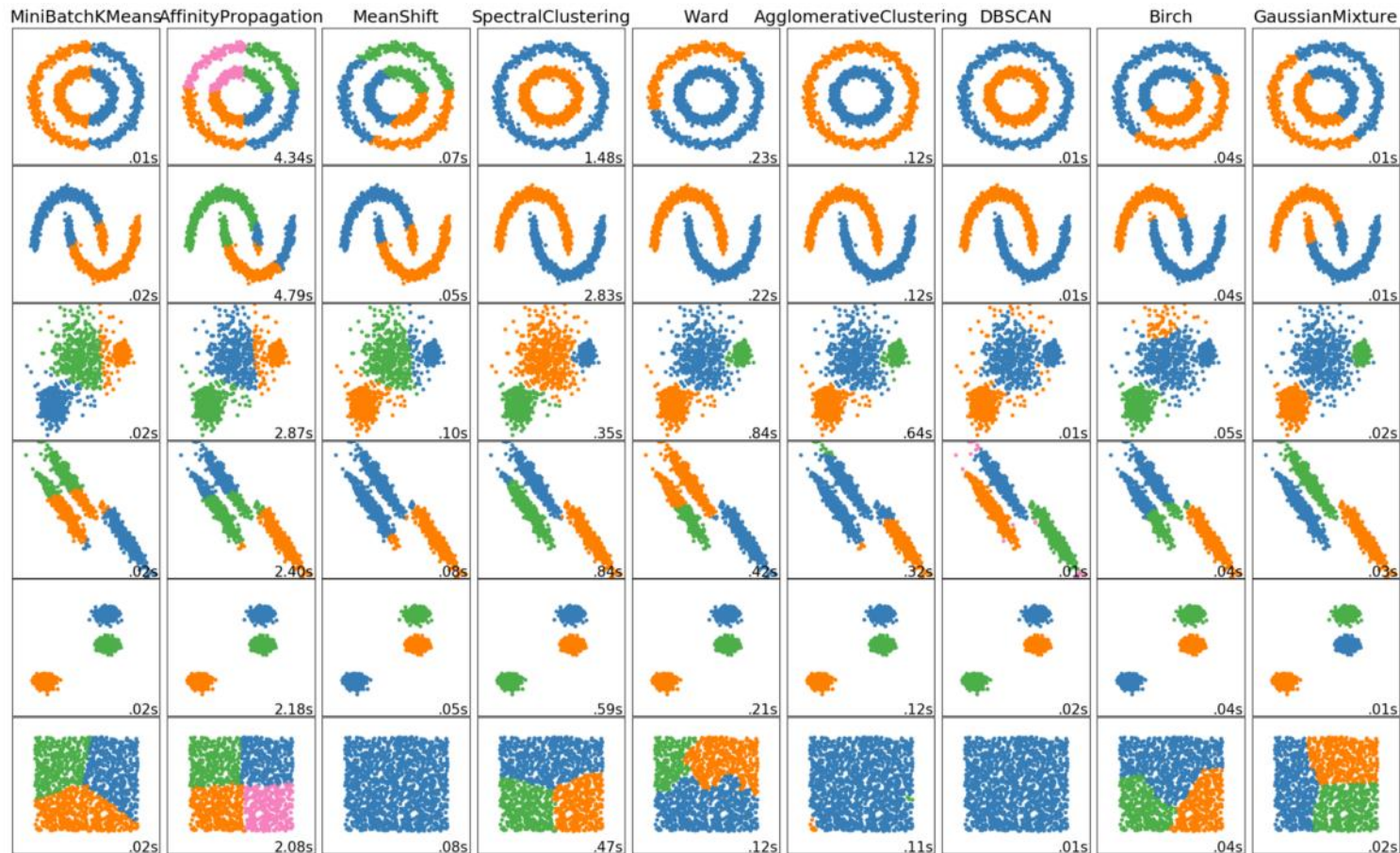
Clustering

Outlier/Anomaly Detection

Dimensionality Reduction

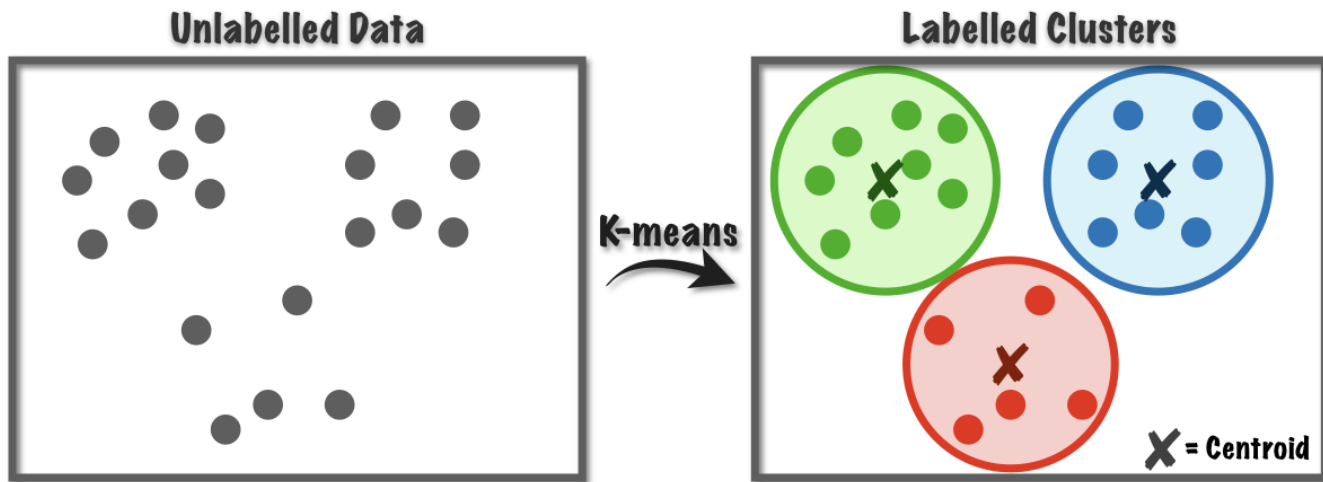
Blind Signal separation

# Clustering



# Clustering

Groeperen van datapunten zodat een punt sterker lijkt op een punt binnen dezelfde groep dan op punten buiten de groep.



# Clustering - Toepassingen

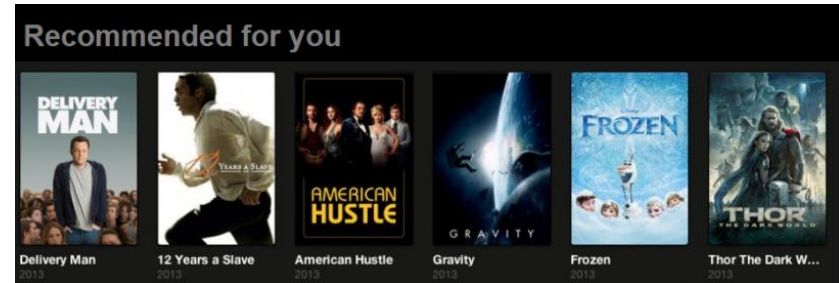
Gezichtsherkenning



# Clustering - Toepassingen

Gezichtsherkenning

Gelijkaardige artikels of films



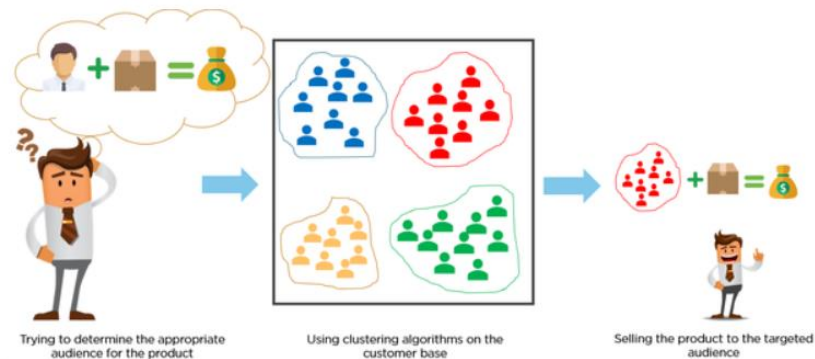


# Clustering - Toepassingen

Gezichtsherkenning

Gelijkaardige artikels of films

Customer clustering voor marketing



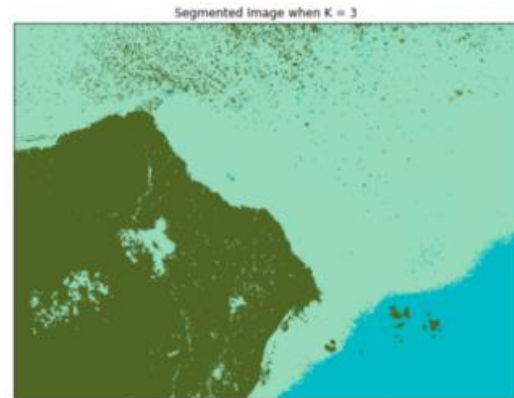
## Clustering - Toepassingen

Gezichtsherkenning

Gelijkaardige artikels of films

Customer clustering voor marketing

Image Segmentation



### Undeclared pools in France uncovered by AI technology

3 days ago



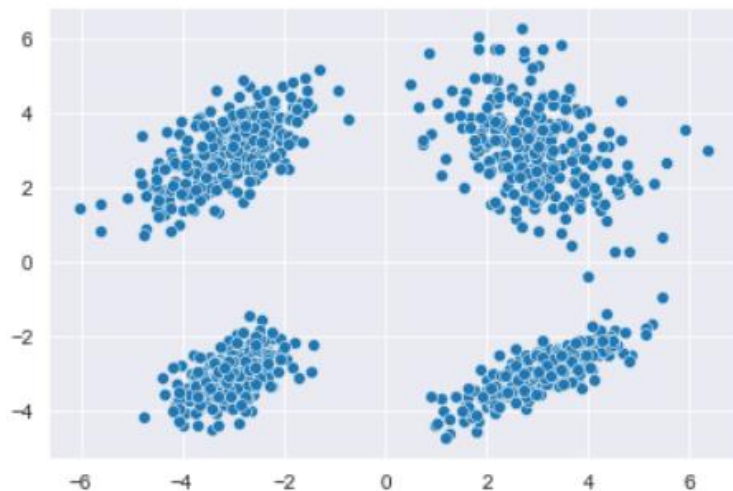
<https://www.bbc.com/news/world-europe-62717599>

## K-Means clustering

Eenvoudige maar populaire techniek voor clustering

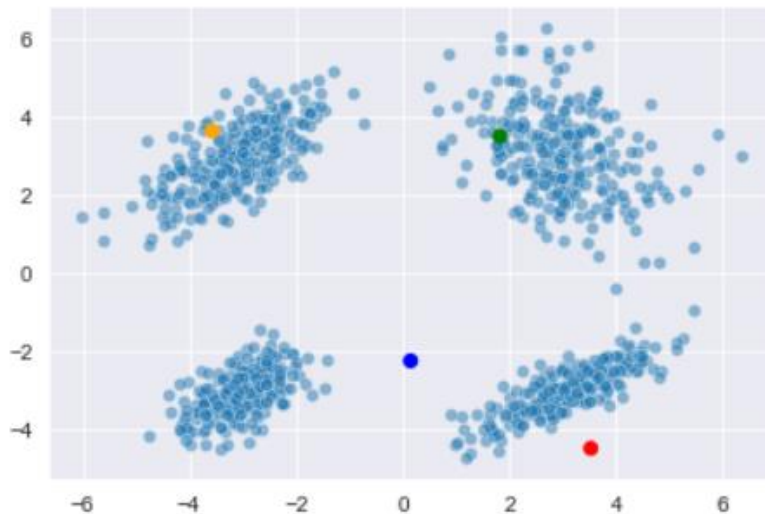
Sterk gerelateerd aan K-Nearest Neighbours

Hoeveel en welke clusters zie je hier?



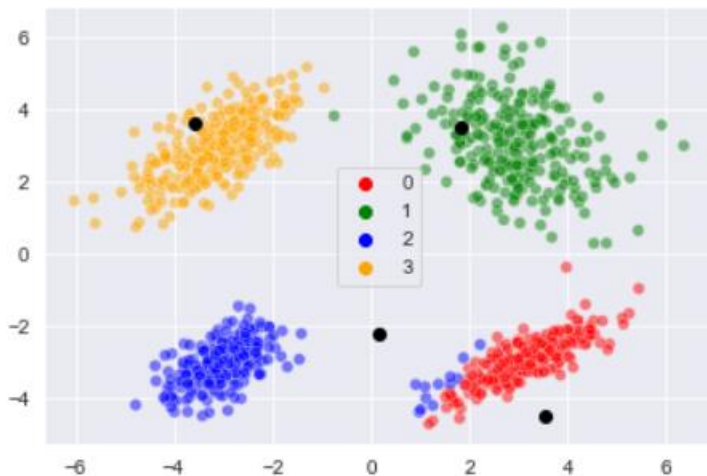
# K-Means clustering

1. Kies K willekeurige punten ( K is het aantal clusters dat je zoekt)



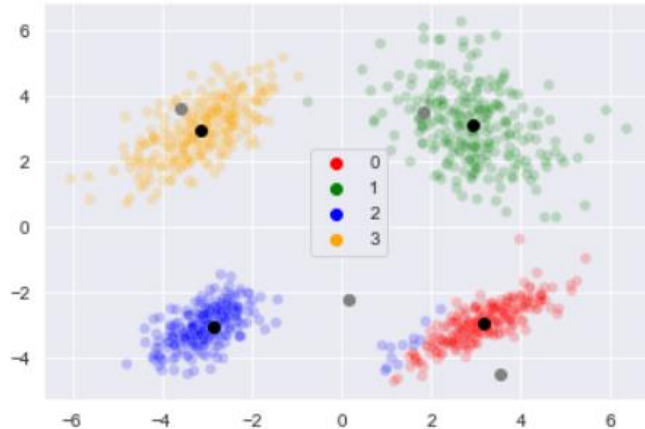
## K-Means clustering

1. Kies K willekeurige punten/centroids ( K is het aantal clusters dat je zoekt)
2. Ken elk punt toe aan het dichtste centroid



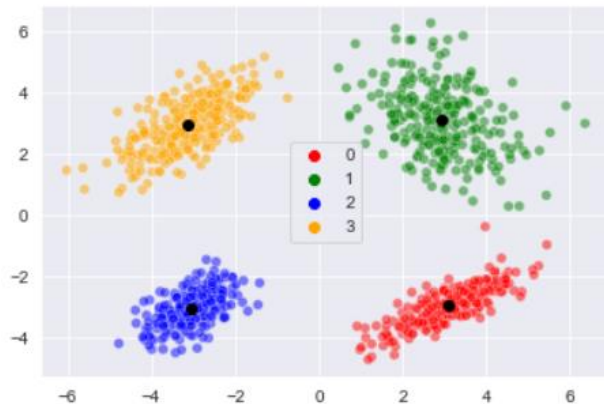
## K-Means clustering

1. Kies K willekeurige punten/centroids ( K is het aantal clusters dat je zoekt)
2. Ken elk punt toe aan het dichtste centroid
3. Verplaats elke centroid naar het gemiddelde van de punten toegekend aan het punt



## K-Means clustering

1. Kies K willekeurige punten/centroids ( K is het aantal clusters dat je zoekt)
2. Ken elk punt toe aan het dichtste centroid
3. Verplaats elke centroid naar het gemiddelde van de punten toegekend aan het punt
4. Herhaal stap 2 en 3 tot er convergentie is



## Kenmerken

Eenvoudig algoritme en resultaten gemakkelijk te interpreteren

Resultaat hangt af van origineel gekozen punten

- Probeer verschillende initialisatie
- Gebruik datapunten als centroids
- Verspreid de centroids bij initialisatie

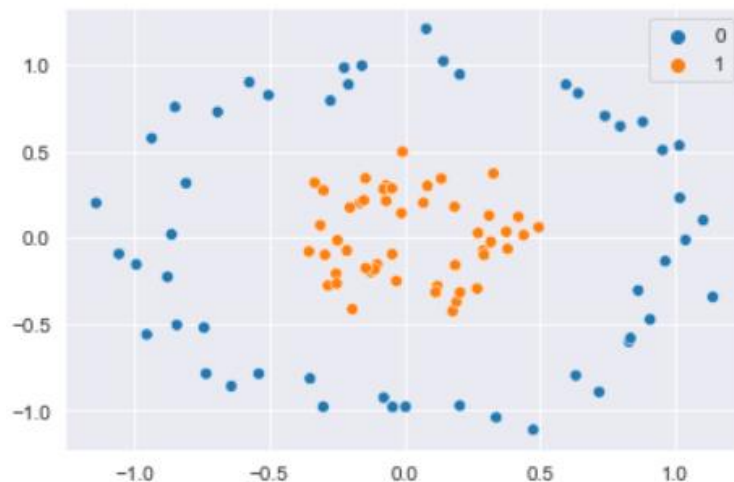
Globaal optimum niet altijd gevonden



## Kenmerken

Gevoelig aan outliers

Probleem bij niet sferische clusters -> Gebruik een kernel zoals bij SVM

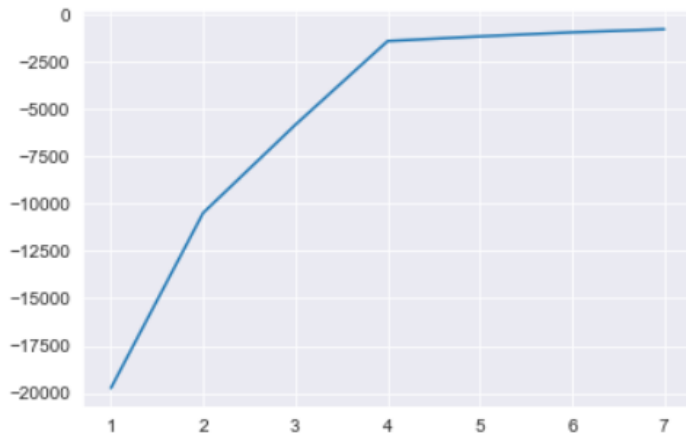


## Hyperparameter K

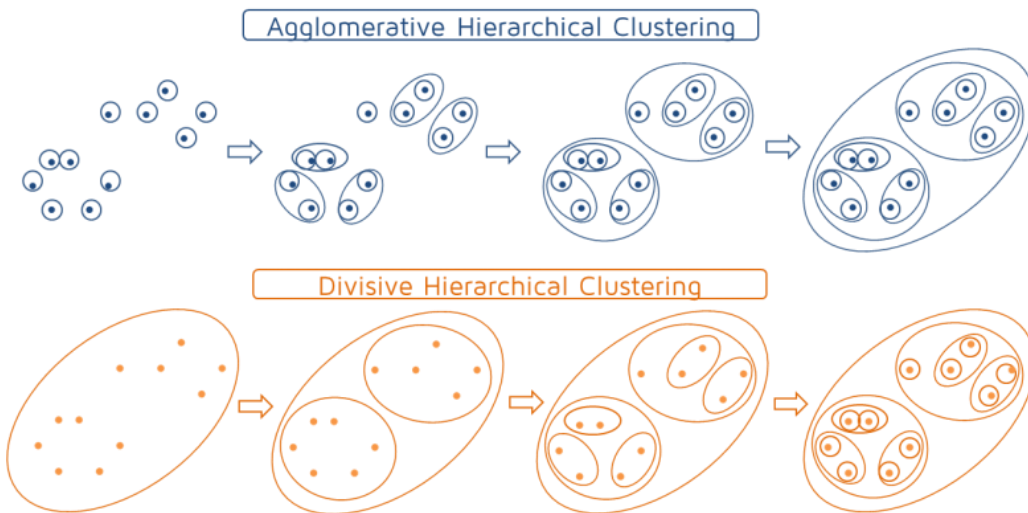
Hoe aantal clusters bepalen?

Elbow method:

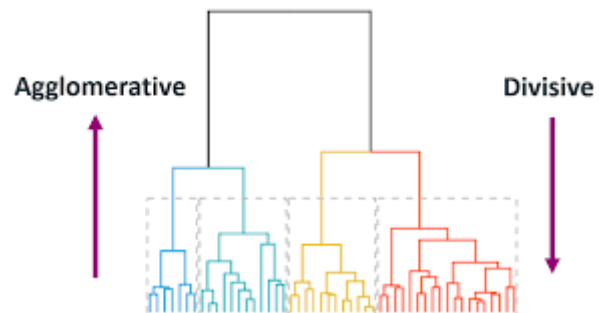
- ▣ Probeer verschillende waarden en kies de K waar de score niet sterk veranderd



# Hierarchical clustering



# Hierarchical clustering - dendrogram



# Hierarchical clustering

Voordelen:

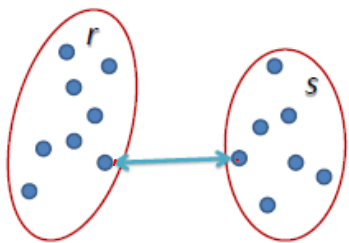
- Geen zelf-gekozen aantal clusters nodig
- Clusterstructuur kan handig zijn

Nadelen:

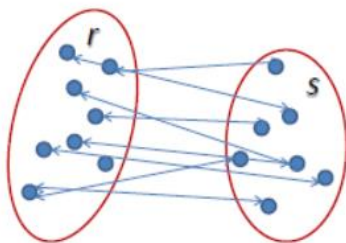
- Snel zeer rekenintensief door afstand tussen clusters te bepalen

## Hierarchical clustering

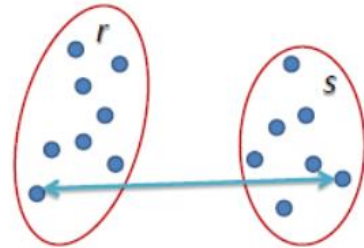
Afstand tussen clusters -> steeds elk punt van een cluster vergelijken met elk punt uit de andere cluster



$$L(r,s) = \min(D(x_{ri}, x_{sj}))$$



$$L(r,s) = \frac{1}{n_r n_s} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} D(x_{ri}, x_{sj})$$



$$L(r,s) = \max(D(x_{ri}, x_{sj}))$$

## Mean-Shift Clustering

Sliding window techniek dat de punten met de hoogste densiteit probeert te zoeken.

Aantal clusters wordt zelf gezocht

Grootte van de sliding window kan een grote impact hebben op het resultaat

Animatie: <https://towardsdatascience.com/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68>

