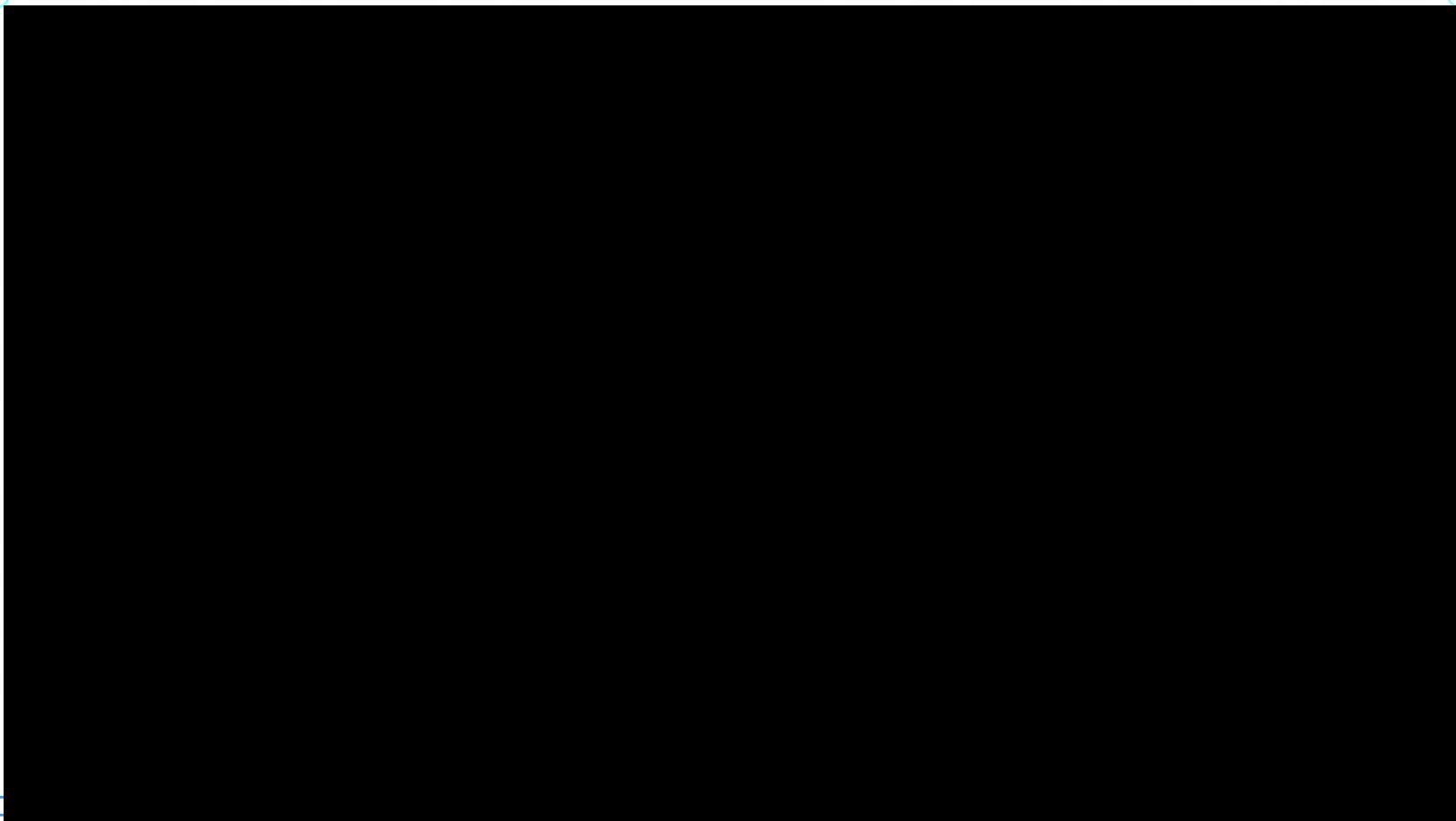


The background features a series of concentric, light gray circles centered on the slide. Overlaid on these are stylized, light blue circuit-like lines with small circles at the ends, resembling a network or data flow, located in the corners.

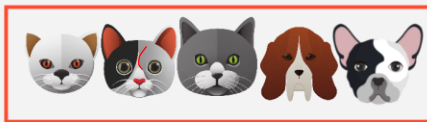
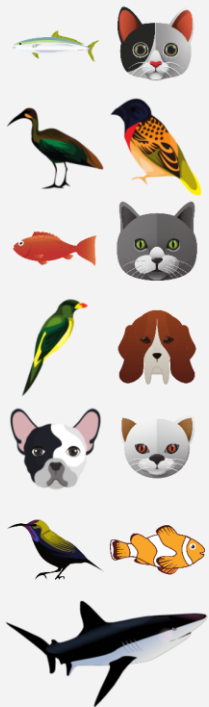
UNSUPERVISED LEARNING

JENS BAETENS



<https://www.youtube.com/watch?v=1qtfILYSDJY>

No labels



UNSUPERVISED LEARNING

Beschrijven van en zoeken naar structuur in niet-gelabelde data

Clustering

Outlier/Anomaly Detection

Dimensionality Reduction

Blind Signal separation

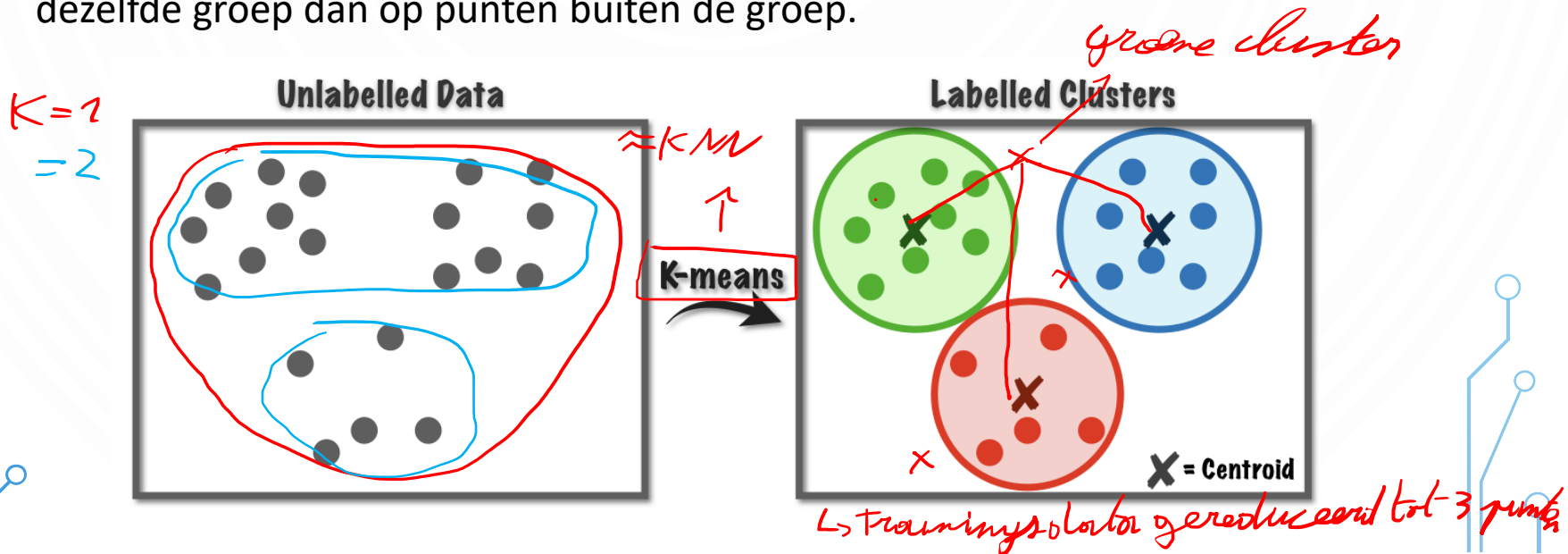
Technieken

- # clusters belangrijk
- Isolation Forest
- One-class SVM
- PCA principal component analysis
 - ↳ data reduce, compressie
 - ↳ curse of dimensionality

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

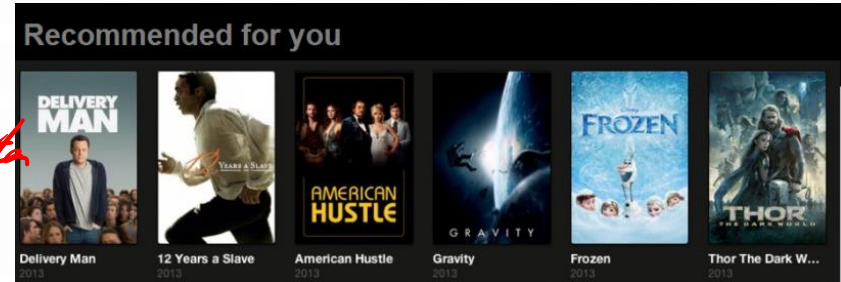


CLUSTERING - TOEPASSINGEN

Gezichtsherkenning

Gelijkaardige artikels of films | *Products*

↳ Recommendation Models

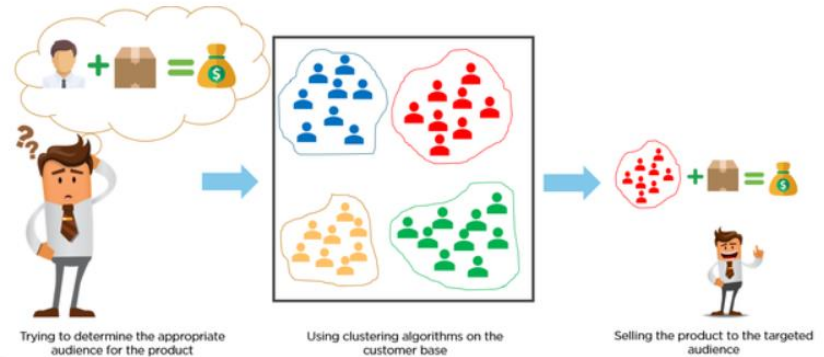


CLUSTERING - TOEPASSINGEN

Gezichtsherkenning

Gelijkaardige artikels of films

Customer clustering voor marketing



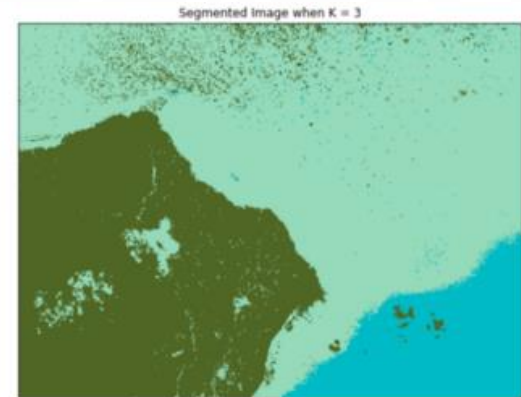
CLUSTERING - TOEPASSINGEN

Gezichtsherkenning

Gelijkaardige artikels of films

Customer clustering voor marketing

Image Segmentation



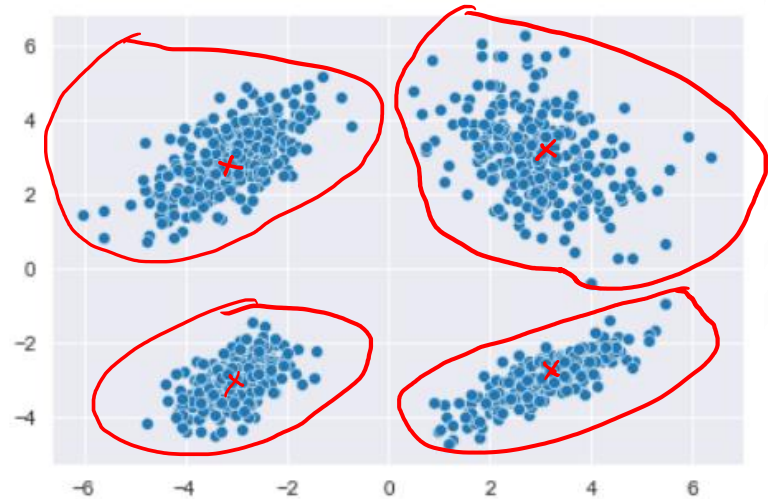
K-MEANS CLUSTERING

↳ gem cluster

Eenvoudige maar populaire techniek voor clustering

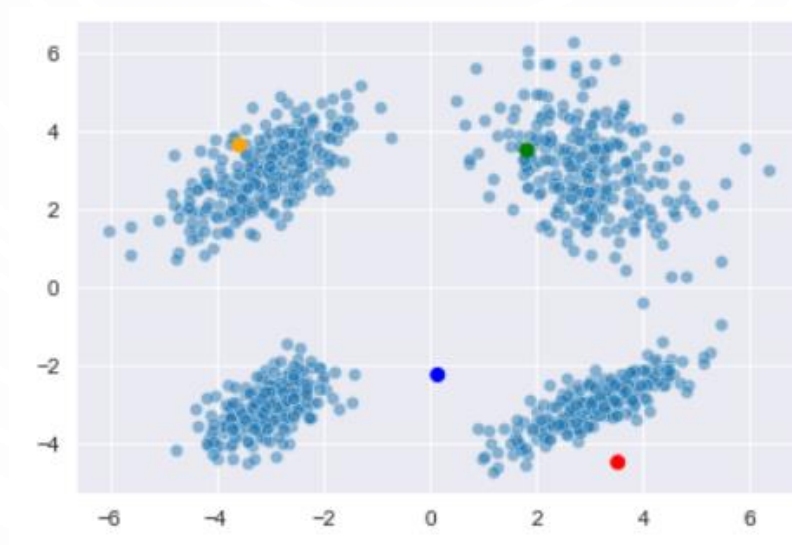
Sterk gerelateerd aan K-Nearest Neighbours

Hoeveel en welke clusters zie je hier?



K-MEANS CLUSTERING

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



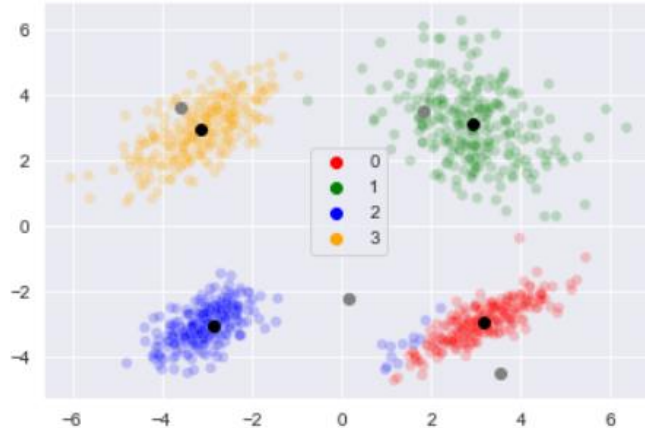
K-MEANS CLUSTERING

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



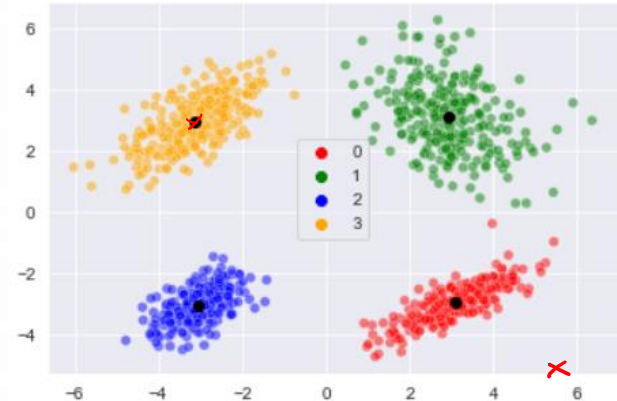
K-MEANS CLUSTERING

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

Evaluëren?

↳ Train

↳ Cluster to
Oepalen

- Test

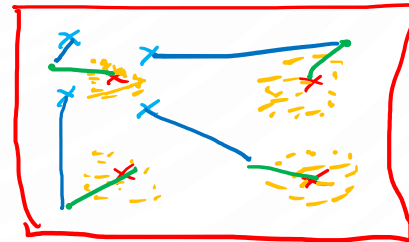
↳ Kortste afstand
→ optellen

Eenvoudig algoritme en resultaten gemakkelijk te interpreteren

Resultaat hangt af van origineel gekozen punten

- Probeer verschillende initialisatie
- Gebruik datapunten als centroids (*beginpunten*)
- Verspreid de centroids bij initialisatie

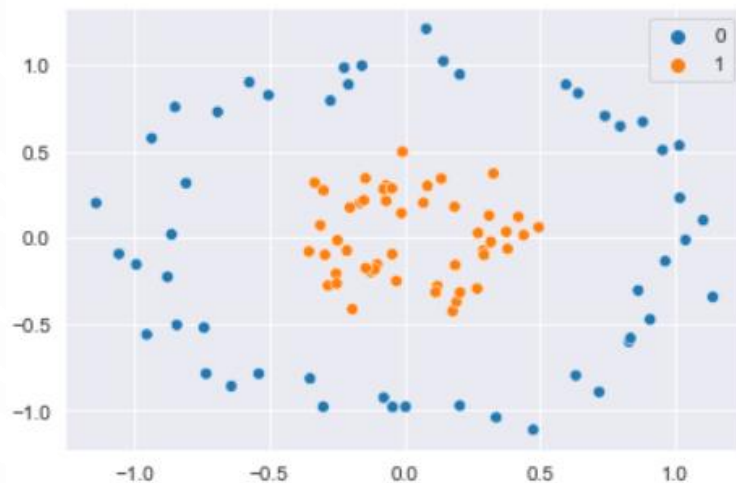
Globaal optimum niet altijd gevonden



KENMERKEN

Gevoelig aan outliers

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

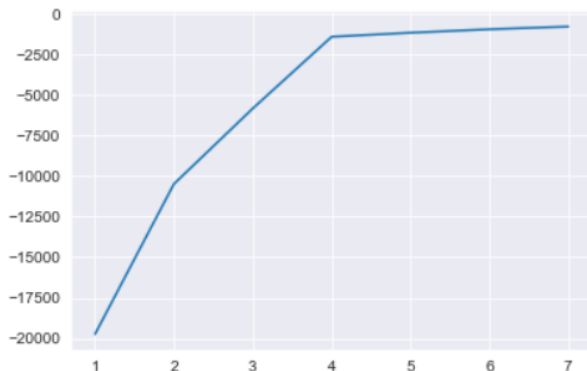


HYPERPARAMETER K

Hoe aantal clusters bepalen?

Elbow method:

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



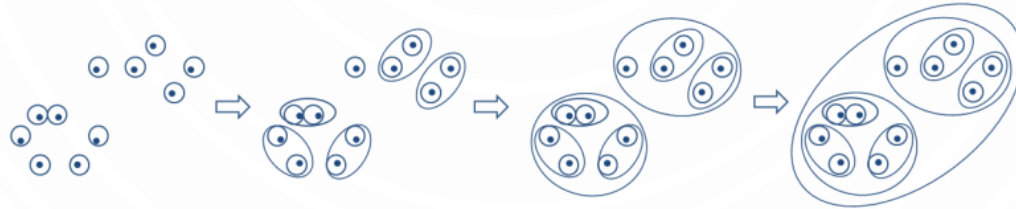
HIERARCHICAL CLUSTERING

→ zeer rekenintensief

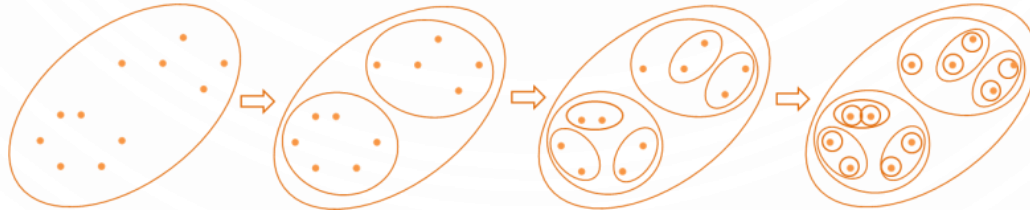
$K=2$



Agglomerative Hierarchical Clustering



Divisive Hierarchical Clustering



HIERARCHICAL CLUSTERING - DENDROGRAM

*↳ Datapunta combinare met datapunten
Clusters*

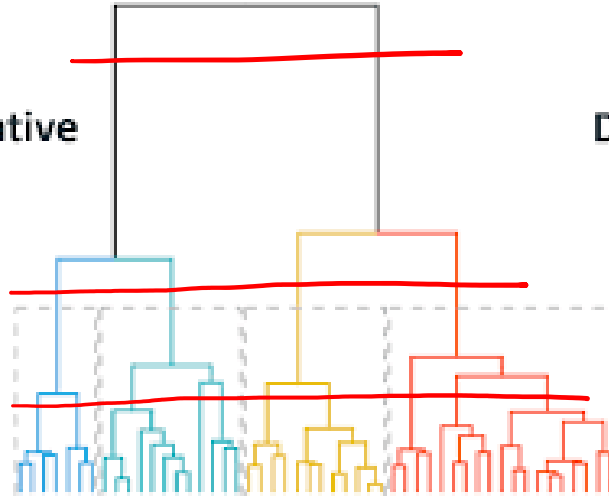
$K=2$

$K=4$

$K=10$

Agglomerative

Divisive



HIERARCHICAL CLUSTERING

Voordelen:

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

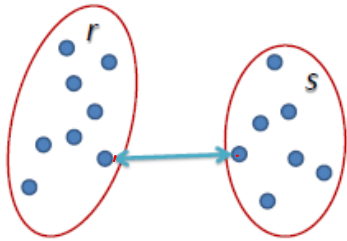
Nadelen:

- Snel zeer rekenintensief door afstand tussen clusters te bepalen

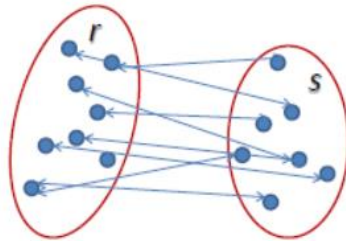
↳ Verbreiden / Specificeren aanbevelingen

HIERARCHICAL CLUSTERING

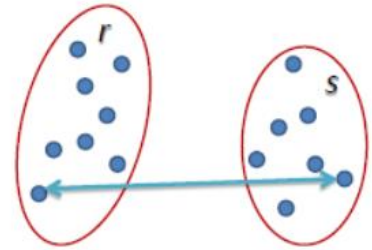
Afstand tussen clusters -> steeds elk punt van een cluster vergelijken met elk punt uit de andere cluster



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



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



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

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>