

Unsupervised ML



BANK OF UGANDA



Bank of Uganda

University of Cape Town

Short Course

November 2019



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

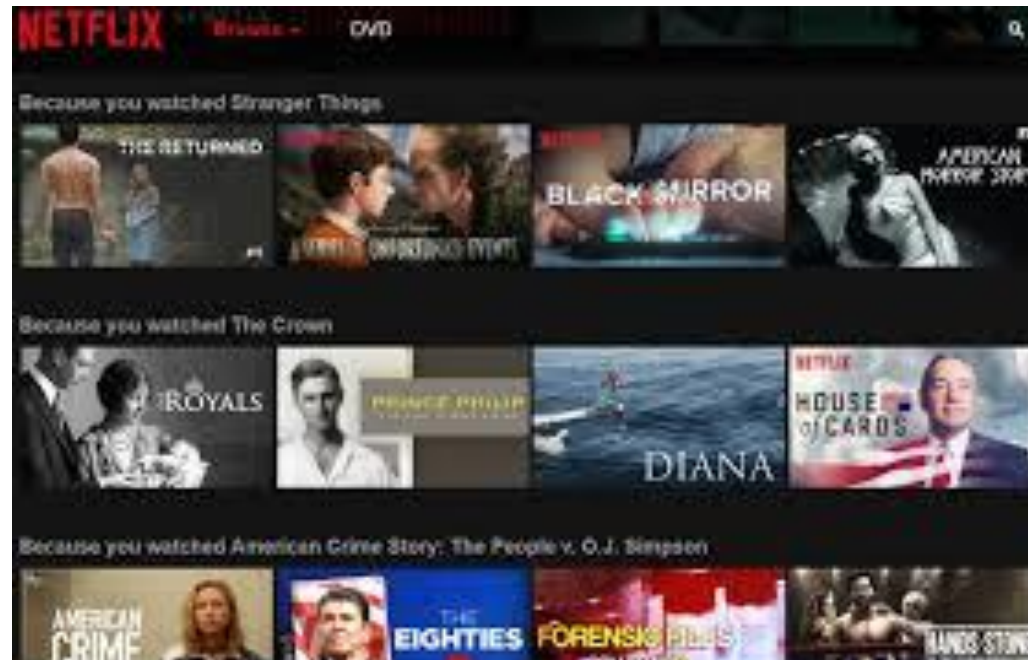
Learning outcomes

- Introduction to unsupervised machine learning
- Implement unsupervised ML



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

What is unsupervised ML?



From an interesting post on experimentation at Netflix ([link](#))



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

What is unsupervised ML?

- Data (observations/people/products/content) does not have labels
- Want to group 'similar' things
- Often these similarities are multi-dimensional and this might be difficult to pick up without ML



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Outline

- Similarity and distance
- Clustering using k-means
- Hierarchical clustering
- Practical application - can we find similar people in our South African data?



Similarity and distance

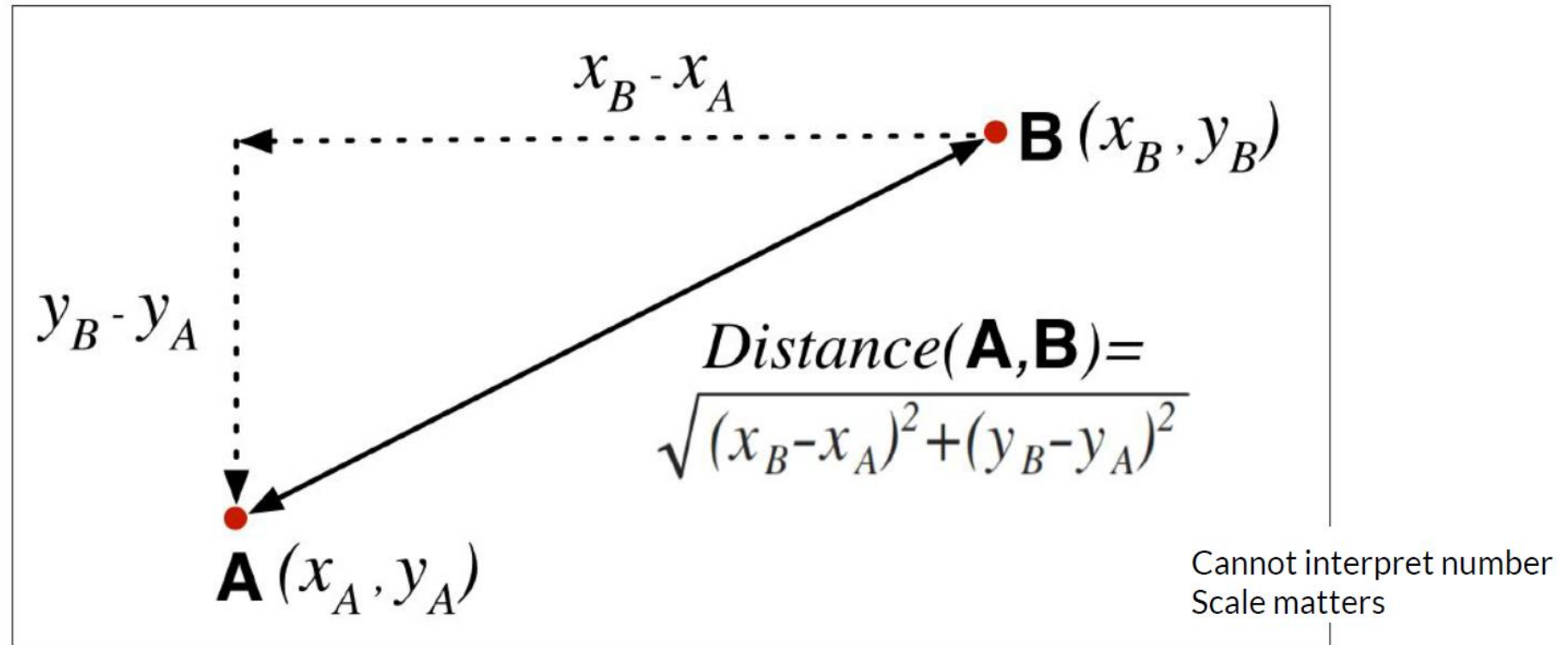


Figure 6-1. Euclidean distance.

Source: Provost and Fawcett, 2013



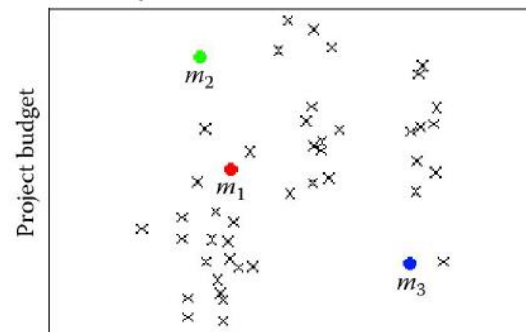
UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Normalising

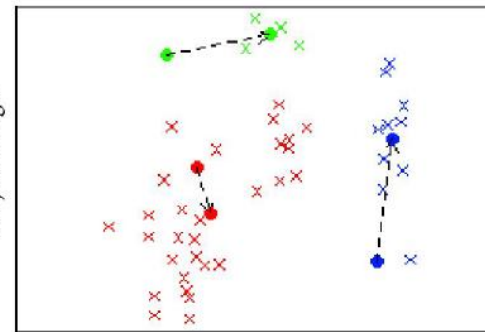
- Scale can matter a lot
- Not comparing like-with-like
- Normalise
 - Mean of 0 and standard deviation of 1
- In R: *scaled_x* <- *scale(x)*



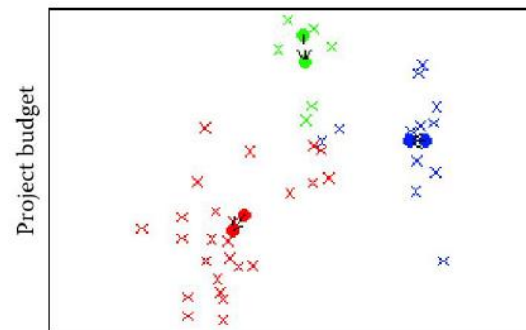
Clustering around a centroid (k means)



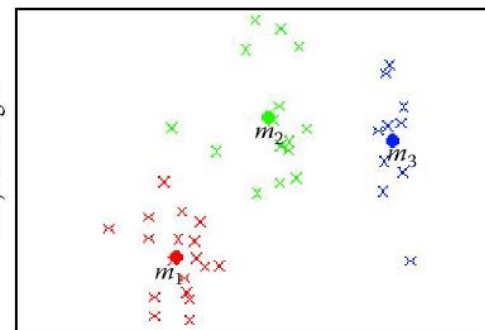
Number of people working on the project



Number of people working on the project



Number of people working on the project



Number of people working on the project

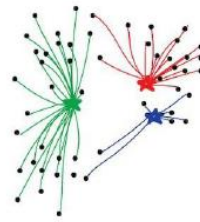


Clustering around a centroid (k means)

PUT KEBAB KIOSKS IN THE OPTIMAL WAY
(also illustrating the K-means method)



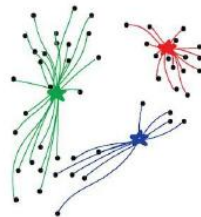
1. Put kebab kiosks in random places in city



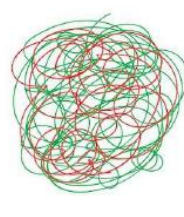
2. Watch how buyers choose the nearest one



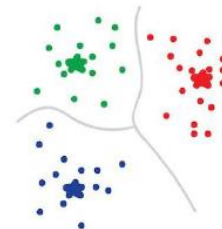
3. Move kiosks closer to the centers of their popularity



4. Watch and move again



5. Repeat a million times

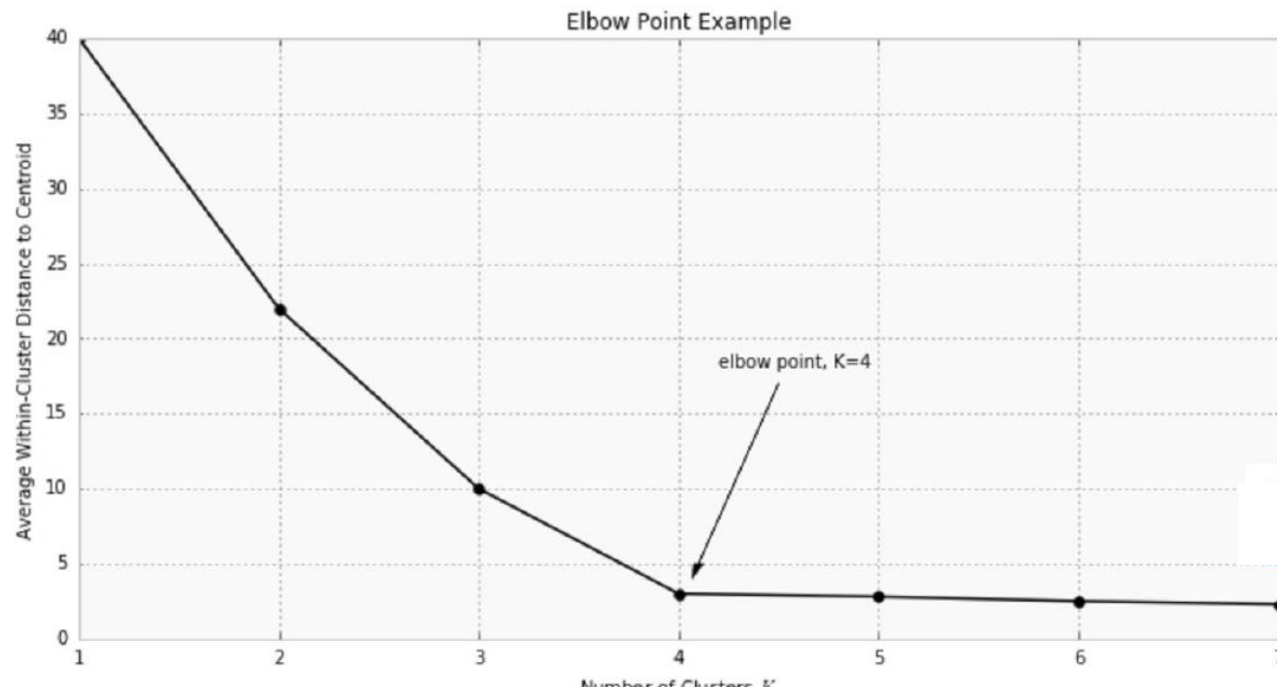


6. Done!
You're god of kebabs!



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

How to choose k? Elbow plots



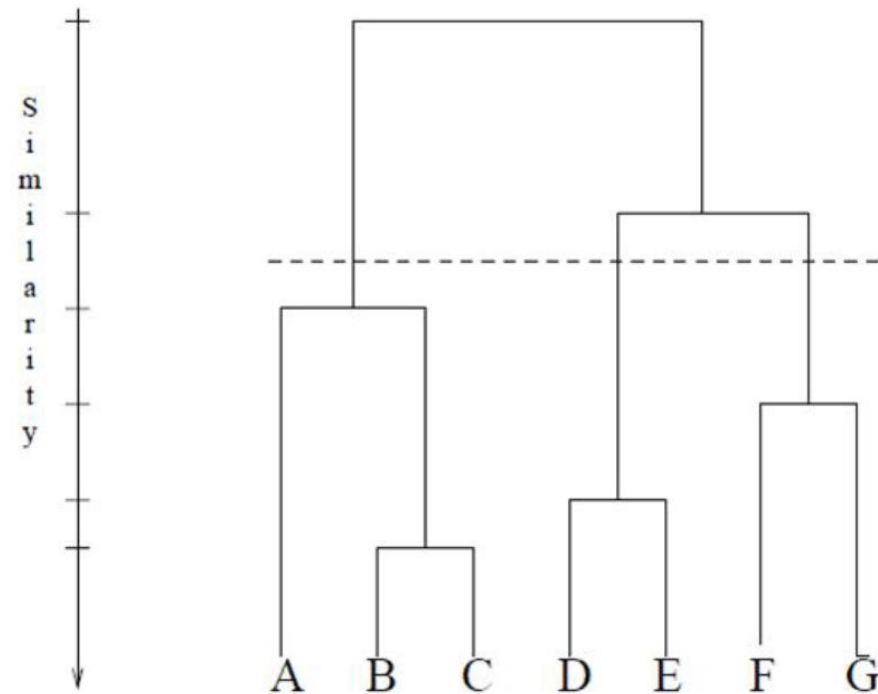
UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Hierarchical clustering

- Bottom-up approach
- Focus on similarities and how these link instances together
- Represented as a dendrogram



Hierarchical clustering



Pick number of clusters where the 'branches' are longest (most dissimilar)

[From a brief explanation of hierarchical clustering](#)



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Practical application

Let's find similar people



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Resources

- K-means explanation

https://uc-r.github.io/kmeans_clustering

- Unsupervised ML using R

<https://lgatto.github.io/IntroMachineLearningWithR/unsupervised-learning.html>



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD