Digital Technologies and Value Creation

Dr. Philippe Blaettchen
Bayes Business School (formerly Cass)

www.bayes.city.ac.uk

Overview – subject to change

Overarching theme	Week	
Introduction	1	Introduction to analytics applications and coding basics
Gathering data	2	Scraping web data
Gathering data / descriptive analytics	3	Data pre-processing and descriptive analytics
Gathering data / descriptive analytics	4	Descriptives in marketing analytics, and using social media APIs
Descriptive analytics	5	Descriptives in people analytics
NO LECTURE	6	NO LECTURE
Predictive analytics	7	Retaining employees and customers with classification
Predictive analytics	8	Wrapping up classification and a deep-dive into dimensionality reduction
Predictive analytics	9	Segmenting customers and positioning products
Prescriptive analytics	10	Optimizing products and organizations
Prescriptive analytics	11	A/B-testing in practice

Learning objectives of today

Goals:

- Understand how clustering works in more detail
- Apply clustering in a typical marketing context

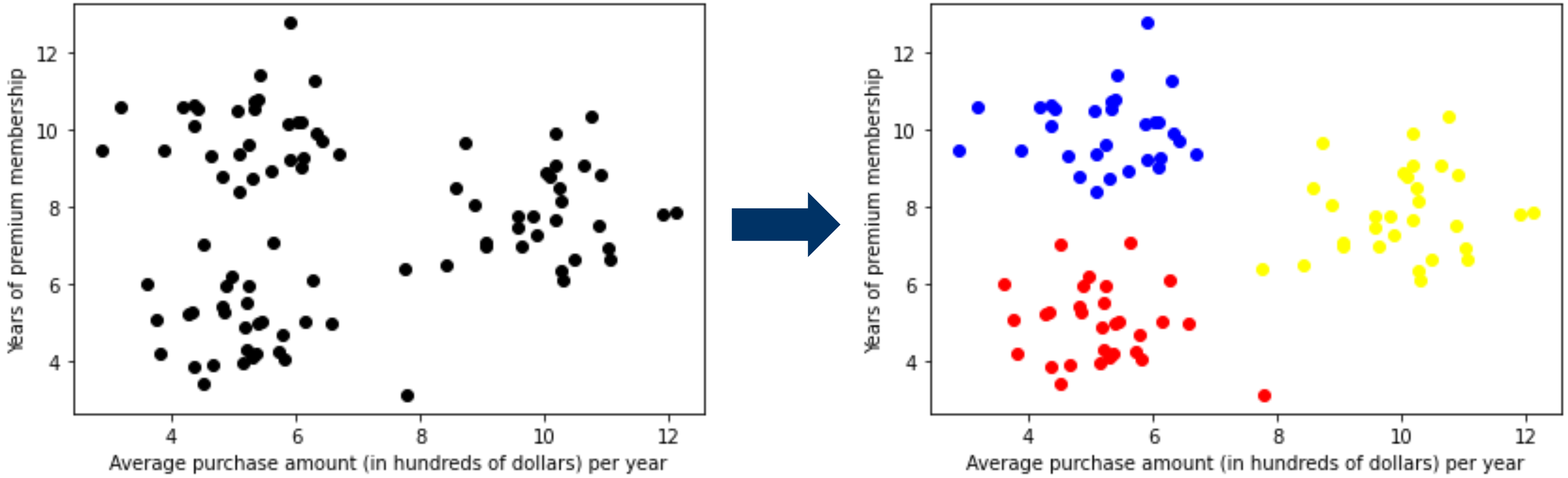
How will we do this?

- Discover how K-means and hierarchical clustering work by working on them “by hand”
- Applying the techniques in Python to cluster customers of a mall (Exercise notebook + *mall_customers.csv*)

BAYES
BUSINESS SCHOOL
CITY, UNIVERSITY OF LONDON

Clustering overview

What is clustering?



Each datapoint corresponds to a customer.

The goal of clustering is to **group together points that have "similar" characteristics**.

(Here, we could cluster to provide, e.g., specific promotional offers.)

The underlying idea

- Clustering creates groups:
  - **Within the same group:** observations **similar**
  - **Across groups:** observations **dissimilar**

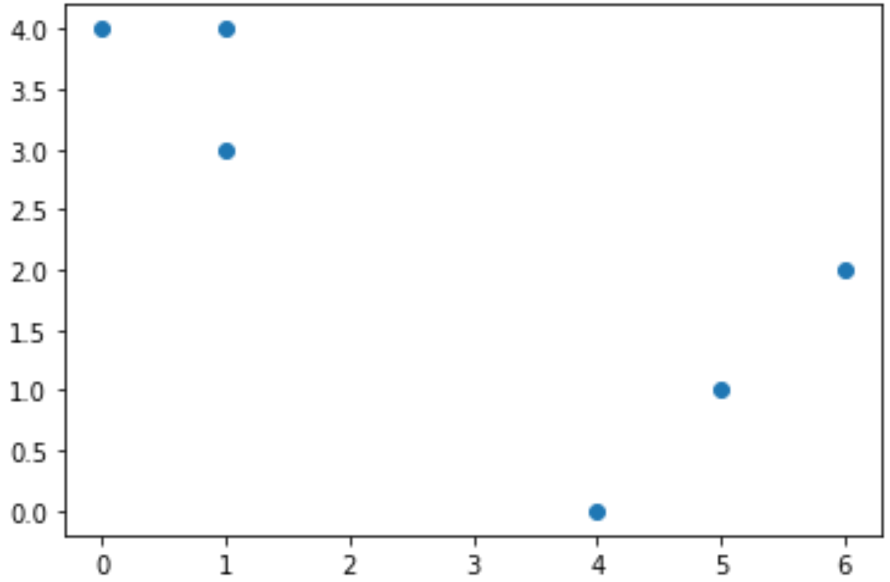
- **Similarity =** two observations are similar if the features of one observation are “close” to the features of the other

- What does “close” mean? → **Many methods for clustering:** we consider two very popular ones:
  - K-means
  - Hierarchical clustering

Easy "by hand" example

To better understand exactly what these algorithms do, we will work on a very easy example [James et al., 2013]

Goal: cluster this dataset into 2 clusters.



Obs	x_1	x_2
0	1	4
1	1	3
2	0	4
3	5	1
4	6	2
5	4	0

BAYES BUSINESS SCHOOL
CITY, UNIVERSITY OF LONDON

Examples of clustering applications

Image segmentation (e.g., background vs foreground)

Detection of cancer (finding lumps in scans)

Segmentation of customers

Credit card or insurance claims fraud

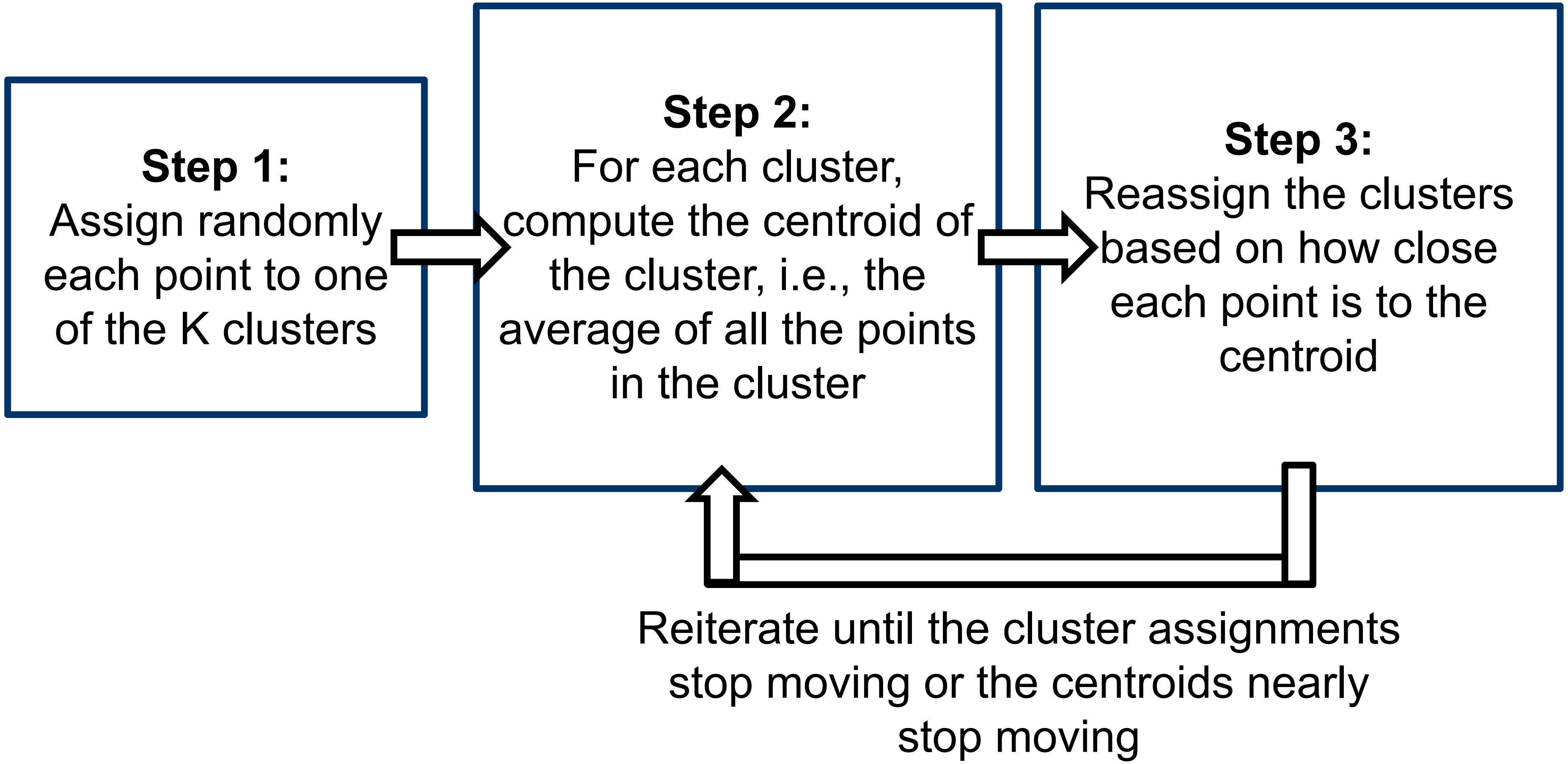
Communities in social networks

Phylogeny (clustering of species)

… and many more

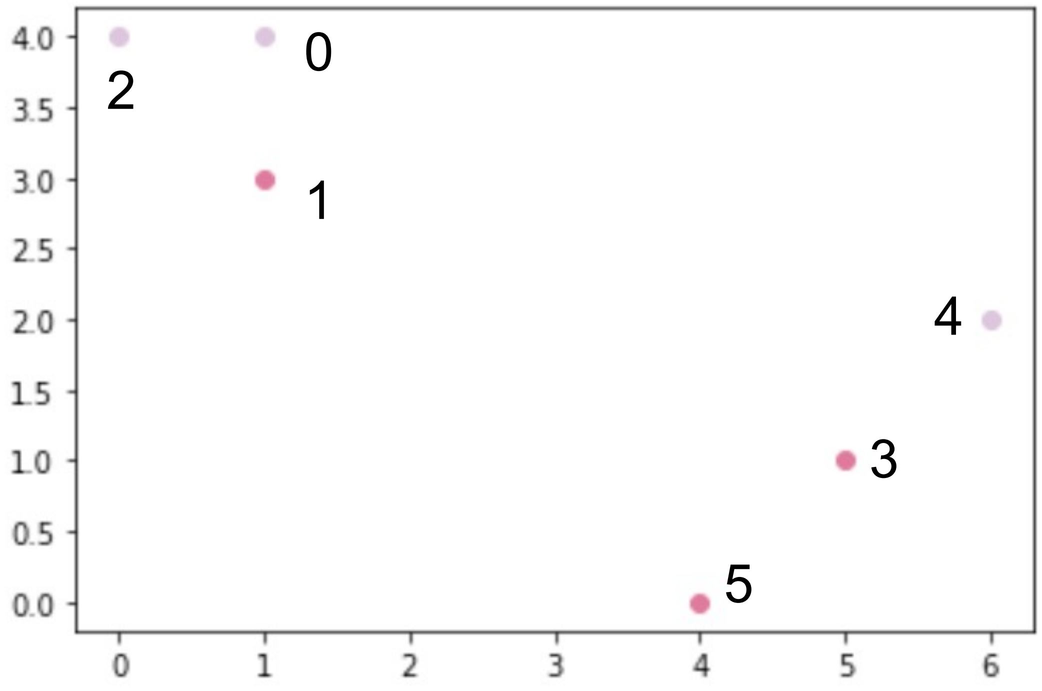
K-means clustering

K-means clustering overview



BAYES
BUSINESS SCHOOL
CITY, UNIVERSITY OF LONDON

K-means clustering



Obs	x_1	x_2
0	1	4
1	1	3
2	0	4
3	5	1
4	6	2
5	4	0

Cluster 1

Cluster 2

The points have been randomly assigned to each cluster: use steps 2 and 3 described in the lecture to find the clusters given by K-means.

K-means clustering overview

Step 1:
Assign randomly each point to one of the K clusters

Step 2:
For each cluster, compute the centroid of the cluster, i.e., the average of all the points in the cluster

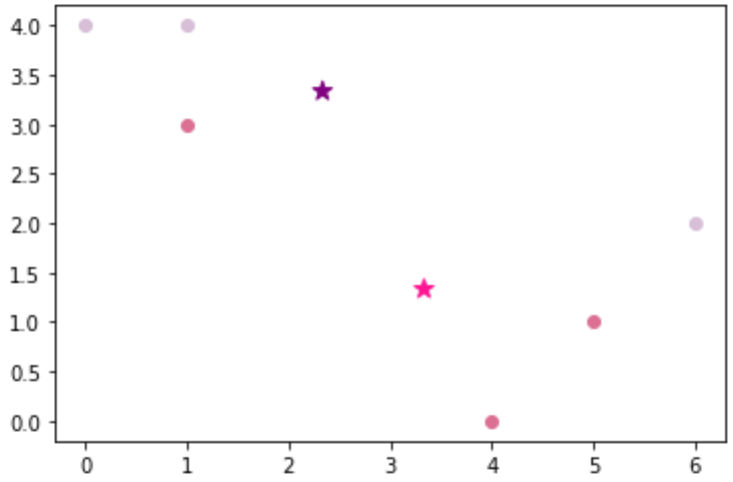
Step 3:
Reassign the clusters based on how close each point is to the centroid

Reiterate until the cluster assignments stop moving or the centroids nearly stop moving

Your turn! Open DTVC_Week 9_Activity.pdf on Moodle and complete the K-means part.

Activity recap

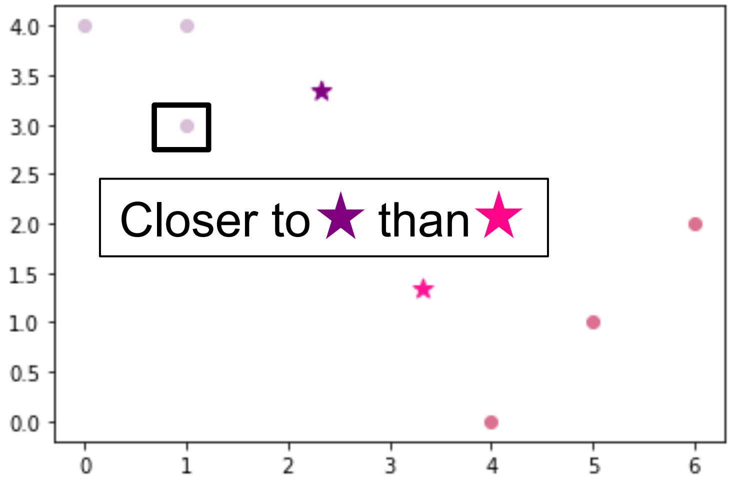
Step 2: Compute centroids



$$\color{purple}\bigstar\color{black} = \frac{\binom{1}{4}+\binom{0}{4}+\binom{6}{2}}{3} = \begin{pmatrix}7/3\\10/3\end{pmatrix}$$

$$\color{magenta}\bigstar\color{black} = \frac{\binom{1}{3}+\binom{5}{1}+\binom{4}{0}}{3} = \begin{pmatrix}10/3\\4/3\end{pmatrix}$$

Step 3: Reassign points



Distance from (1,3) to $\color{purple}\bigstar$:

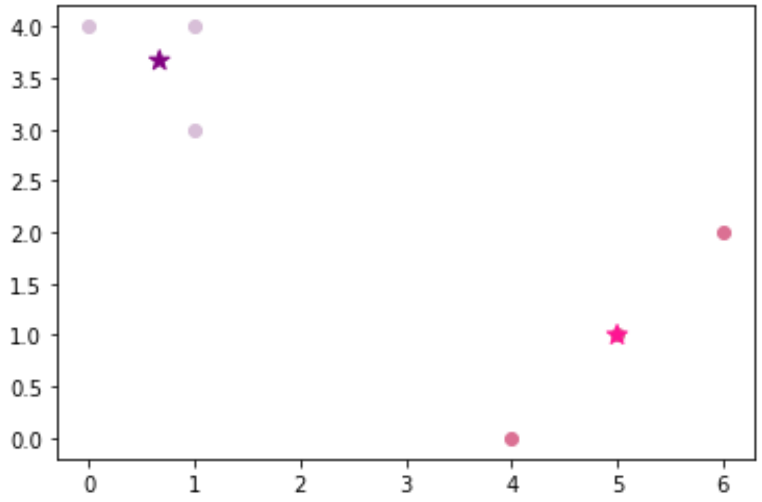
$$\sqrt{\left(1-\frac{7}{3}\right)^2+\left(3-\frac{10}{3}\right)^2} \approx 1.37$$

Distance from (1,3) to $\color{magenta}\bigstar$:

$$\sqrt{\left(1-\frac{10}{3}\right)^2+\left(3-\frac{4}{3}\right)^2} \approx 2.86$$

Activity recap

Repeat – Step 2: Compute centroids



$$\bigstar = \frac{\binom{1}{4} + \binom{1}{3} + \binom{0}{4}}{3} = \begin{pmatrix} 2/3 \\ 11/3 \end{pmatrix}$$

$$\bigstar = \frac{\binom{5}{1} + \binom{6}{2} + \binom{4}{0}}{3} = \begin{pmatrix} 5 \\ 1 \end{pmatrix}$$

Step 3: Reassign points

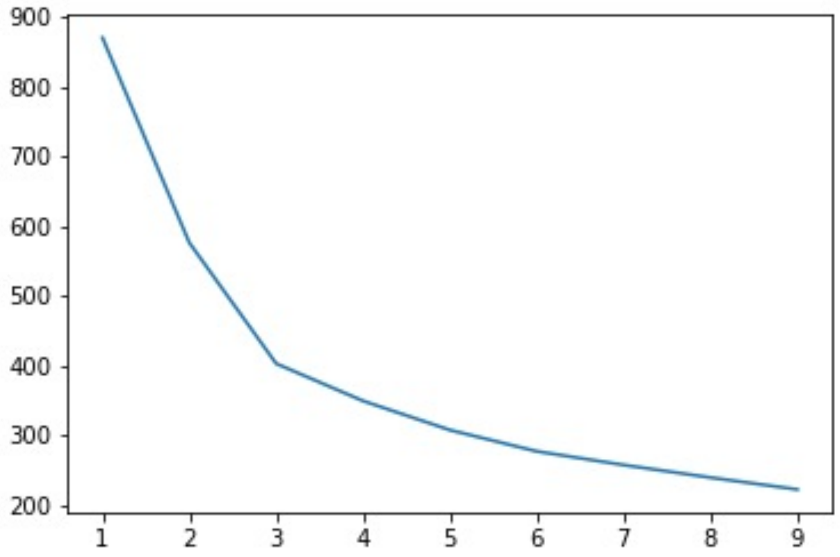
Nothing changes → We are done!

BAYES
BUSINESS SCHOOL
CITY, UNIVERSITY OF LONDON

Recall from the videos…

- Scaling can be very important!
- K-means starts from a random assignment, so we may not always get the same result

  → we can run the algorithm many times (default 10) and get the "best" model (the one with the smallest inertia – sum of squared distances between points and their centroid)
- We can pick K, also based on inertia. For this, we draw an "elbow" plot:



Hierarchical clustering

Hierarchical clustering overview

Step 1:
Make a table of the distances between all pairs of points. Each point is considered as its own cluster.

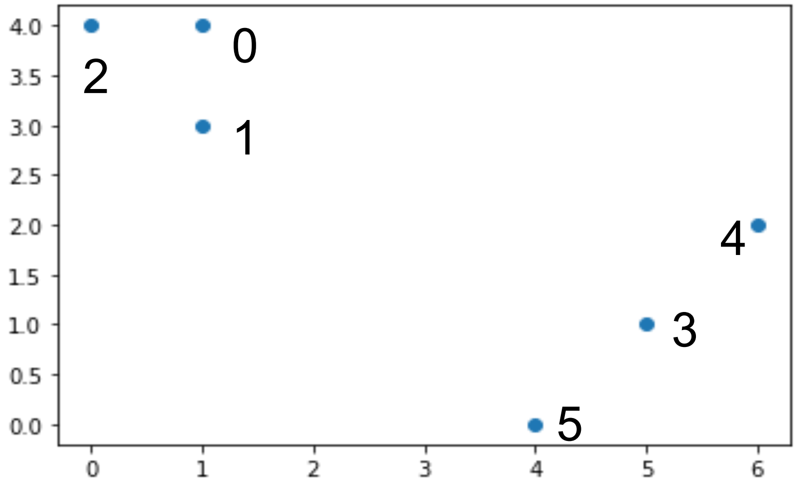
Step 2:
Fuse the two closest clusters into a new cluster.
Draw the corresponding dendrogram.

Step 3:
Recompute the pairwise distance between clusters.
Here, we use average linkage.

Reiterate until all clusters are merged

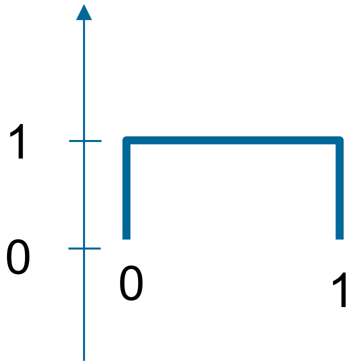
Activity

Step 1: Compute pairwise distances



Dist.	0	1	2	3	4	5
0	0	$\boxed{1}$	1	5	5.39	5
1		0	1.41	4.47	5.09	4.24
2			0	5.83	6.32	5.65
3				0	1.41	1.41
4					0	2.82
5						0

Step 2: Build dendrogram



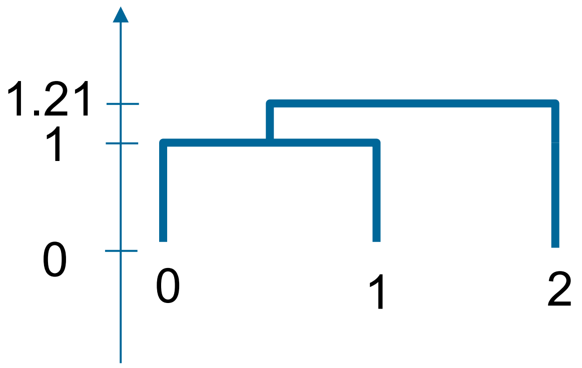
BAYES BUSINESS SCHOOL
CITY, UNIVERSITY OF LONDON

Activity

Step 3: Recompute distance matrix

Dist.	0	1	2	3	4	5
0	0	$\boxed{1}$	1	5	5.39	5
1		0	1.41	4.47	5.09	4.24
2			0	5.83	6.32	5.65
3				0	1.41	1.41
4					0	2.82
5						0

Step 2': Build dendrogram



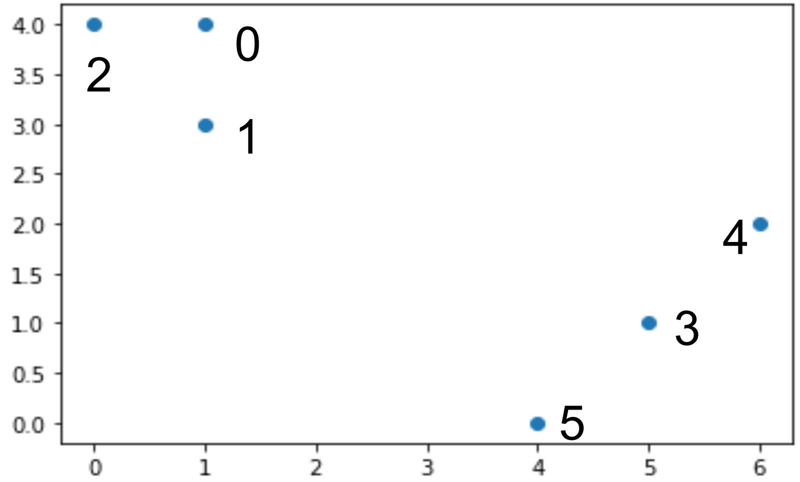
Your turn! Open DTVC_Week 9_Activity.pdf on Moodle and finish the dendrogram!

Dist.	{0,1}	2	3	4	5
{0,1}	0	$\boxed{=\dfrac{d(0,2)+d(1,2)}{2}=1.21}$	4.74	5.24	4.62
2		0	5.83	6.32	5.65
3			0	1.41	1.41
4				0	2.82
5					0

Average linkage

BAYES BUSINESS SCHOOL
CITY, UNIVERSITY OF LONDON

Hierarchical clustering



Obs	x_1	x_2
0	1	4
1	1	3
2	0	4
3	5	1
4	6	2
5	4	0

Dist.	0	1	2	3	4	5
0	0	1	1	5	5.39	5
1		0	1.41	4.47	5.09	4.24
2			0	5.83	6.32	5.65
3				0	1.41	1.41
4					0	2.82
5						0

Dist.	{0,1}	2	3	4	5
{0,1}	0	1.21	4.74	5.24	4.62
2		0	5.83	6.32	5.65
3			0	1.41	1.41
4				0	2.82
5					0

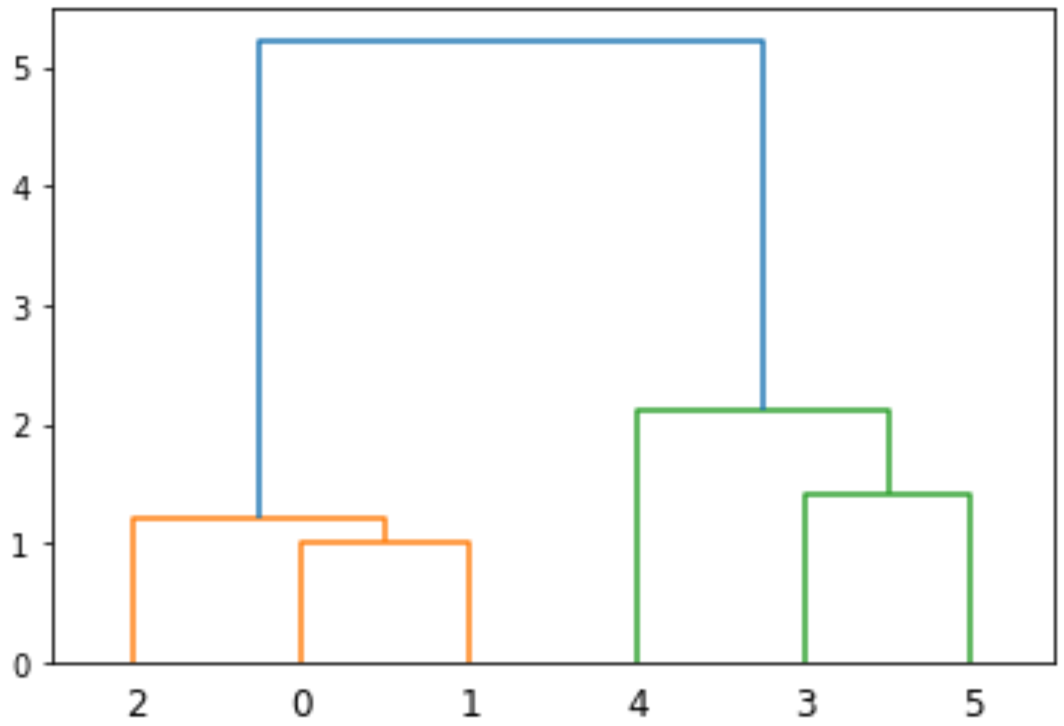
Dist.	

Dist.	

Dist.	

Activity recap

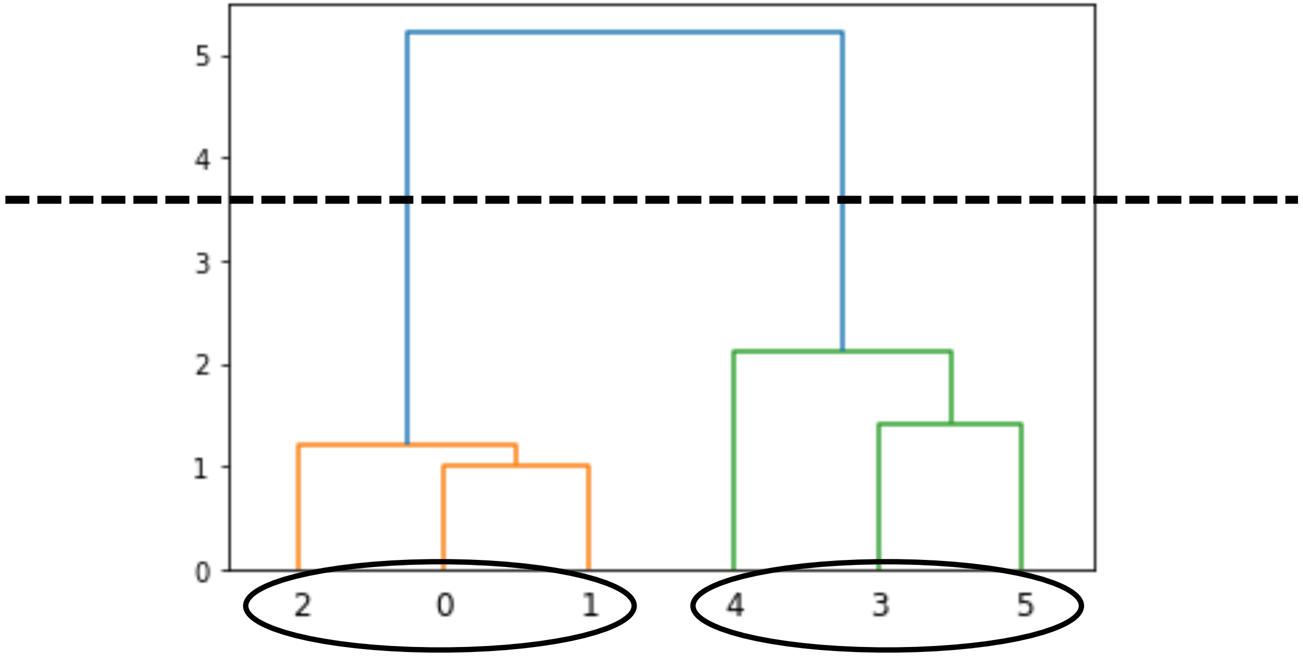
The final dendrogram:



Obtained via a **bottom-up strategy** with **average linkage**.

How to get clusters from dendograms?

Cut the final dendrogram at some level:



In general: try and cut where there's a "jump" on the dendrogram.

Different types of linkages

Dist.	0	1	2	3	4	5
0	0	1	1	5	5.39	5
1		0	1.41	4.47	5.09	4.24
2			0	5.83	6.32	5.65
3				0	1.41	1.41
4					0	2.82
5						0

Dist.	{0,1}	2	3	4	5
{0,1}	0	$= \frac{d(0,2) + d(1,2)}{2} = 1.21$	4.74	5.24	4.62
2		0	5.83	6.32	5.65
3			0	1.41	1.41
4				0	2.82
5					0

Average linkage: average distance between instances (seen)

Example: $d(\{0,1\}, 2) = \frac{d(0,2)+d(1,2)}{2} = 1.21$

Single linkage: smallest distance between instances

Example: $d(\{0,1\}, 2) = \min\big(d(0,2), d(1,2)\big) = 1$

BAYES BUSINESS SCHOOL
CITY, UNIVERSITY OF LONDON

Different types of linkages

Dist.	0	1	2	3	4	5
0	0	1	1	5	5.39	5
1		0	1.41	4.47	5.09	4.24
2			0	5.83	6.32	5.65
3				0	1.41	1.41
4					0	2.82
5						0

Dist.	{0,1}	2	3	4	5
{0,1}	0	$= \frac{d(0,2)+d(1,2)}{2} = 1.21$	4.74	5.24	4.62
2		0	5.83	6.32	5.65
3			0	1.41	1.41
4				0	2.82
5					0

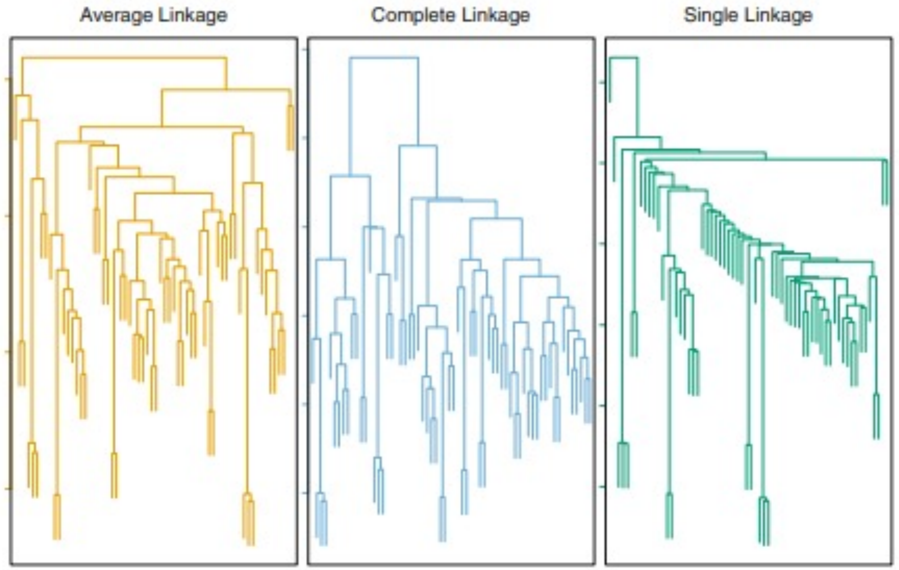
Complete linkage: largest distance between instances
Example: $d(\{0,1\},2) = \max(d(0,2), d(1,2)) = 1.41$

Ward linkage: a bit more complex.
To obtain distance between two clusters: (1) compute the center of each cluster, (2) compute distance between the two centers, (3) weight it by the number of points in each.
Example: center of 2: (0,4); center of {0,1}: (1,3.5)
Distance between the two centers: $\sqrt{(1-0)^2 + (4-3.5)^2} \approx 1.12$
Weighting: $\frac{2 \cdot 1}{1+2} \cdot 1.12 = 0.75$

BAYES BUSINESS SCHOOL
CITY, UNIVERSITY OF LONDON

Different linkages give different results

Same dataset with different types of linkages can give completely different results:



Source: James et al., 2013

In general:

- **Complete, average and Ward** linkage tend to yield **evenly sized** clusters
- **Single linkage** tends to yield extended clusters to which **single leaves** are fused one by one
- **Rule of thumb: complete, average, or Ward**. Go with the dendrogram.

Clustering in practice

Segmenting customers of a mall

- You have data about customers to a mall: basic demographics and shopping behaviors
- You want to identify relevant customer segments using the clustering methods just learned
- These segments should then help you developing targeted marketing strategies

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

Source: Kaggle

Let’s try it out

- Go through the code in the notebook and try to understand the key Python concepts
  → Let me know if you get stuck anywhere!
- Try to interpret the clustering process and the clusters based on the questions in the notebook
- Think of a strategy for targeted marketing based on the segments and be ready to discuss this in class

BAYES BUSINESS SCHOOL
CITY, UNIVERSITY OF LONDON

Activity recap

- How many clusters to use based on the dendogram? Why?
- How many clusters to use based on the elbow plot? Why?
- How would you interpret the clusters? Do the clusters match between the two methods?
- What marketing strategy would you suggest to the mall provider?

Clustering in the group assignment

Boats – the case

- Mary is Senior Manager in the Customer Insights department of CreeqBoat.
- She has conducted market research around needs in the boating industry.

Overarching goal: using the study to
- **understand the market segments** that make up her potential client basis
- **understand the key purchase drivers** in each segment
- **develop a targeted strategy** based on the understanding of customer segments

Usual caveats:
- **Limited by time:** this generally would take weeks and require many different people. The assignment is consequently simplified and can be improved.
- **No one best answer:** many different answers informed by our backgrounds and levels of knowledge of the area. It is important to justify your reasoning!

Boats, Part 1: The key components that inform segmentation

Goal: Boil down the 29 (!!) questions about customer attitudes to a few key factors.

How to do this?

Use PCA!

- **How many components to keep?**
- **Association between original 29 variables and the components?**
- **Interpretation of components?**

Sparse PCA tries to compromise between two things:

- doing the “real” PCA
- pushing for a result where a lot of the loadings are zero

This is controlled by parameter alpha. If alpha=0, real PCA, if alpha large, a lot of zeros ⇒ take alpha=5.

BAYES
BUSINESS SCHOOL
CITY, UNIVERSITY OF LONDON

Boats, Part 2: The actual customer segmentation

Goal: We now have the key factors that "make up" each client. We need to understand now if there are certain typical client profiles in terms of needs.

How to do this?

Use clustering

- **How many clusters to use?**
- **Check robustness of clusters**
- **Use Q2-Q16 to profile these clusters**
- **Develop a targeting strategy**

See you next week!