

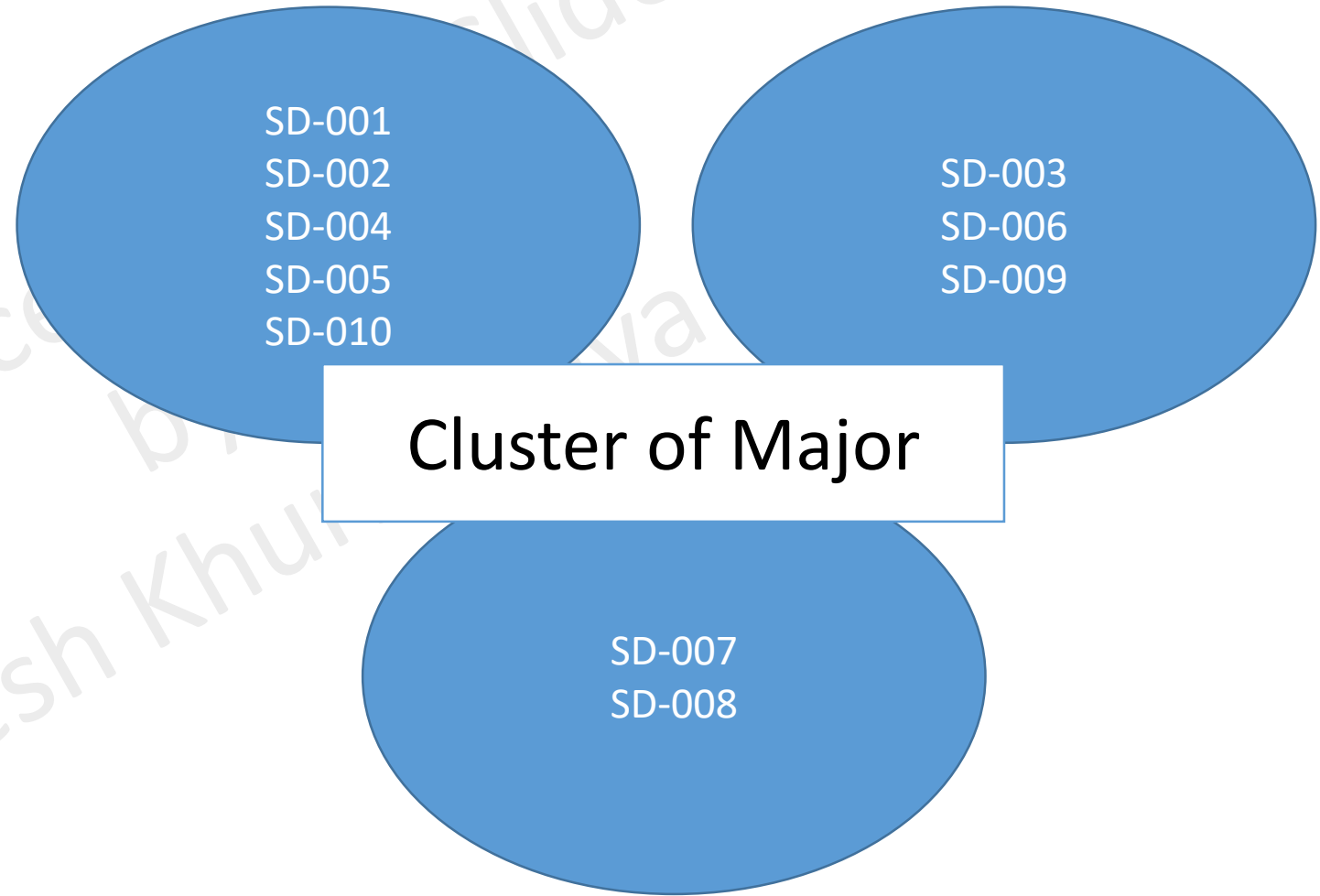
Complete Data Science and Machine Learning Using Python

By
Jitesh Khurkhuriya

Clustering

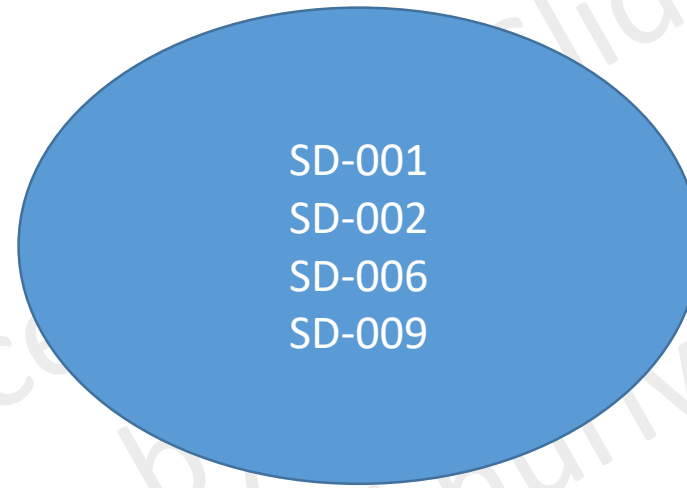
What is Clustering?

Student ID	Gender	Major	Grade
SD-001	M	Math	A+
SD-002	M	Math	A
SD-003	F	Statistics	A+
SD-004	F	Math	A
SD-005	F	Math	B
SD-006	M	Statistics	B
SD-007	F	Physics	A+
SD-008	F	Physics	A
SD-009	M	Statistics	B+
SD-010	F	Math	B+



What is Clustering?

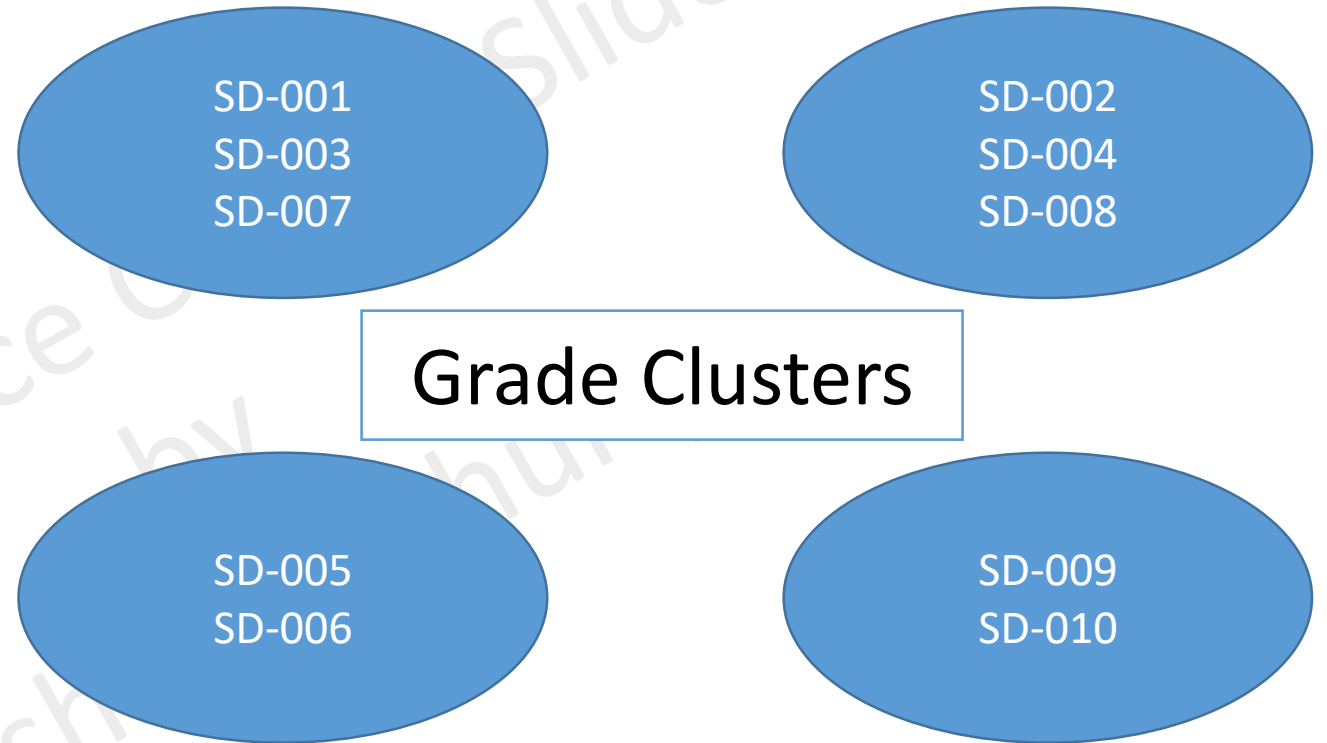
Student ID	Gender	Major	Grade
SD-001	M	Math	A+
SD-002	M	Math	A
SD-003	F	Statistics	A+
SD-004	F	Math	A
SD-005	F	Math	B
SD-006	M	Statistics	B
SD-007	F	Physics	A+
SD-008	F	Physics	A
SD-009	M	Statistics	B+
SD-010	F	Math	B+



Gender Specific Clusters

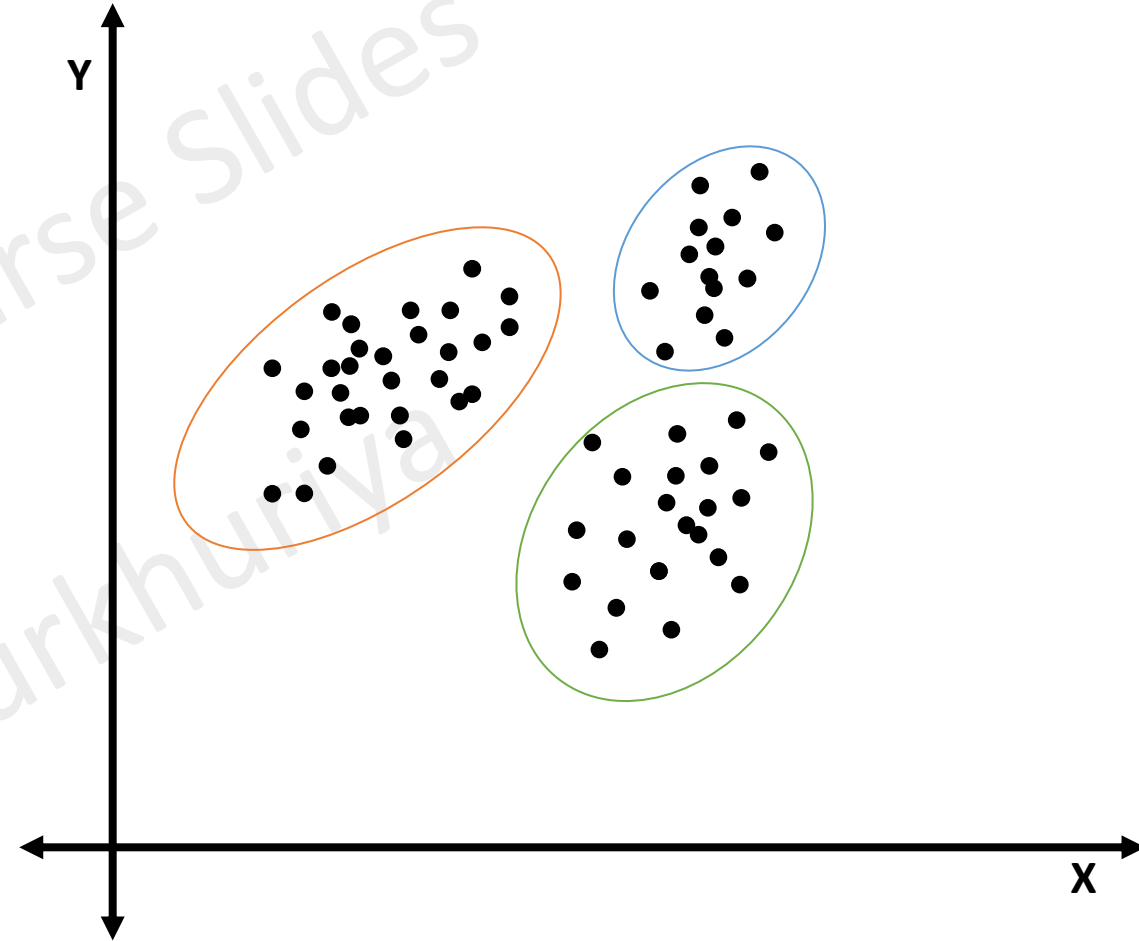
What is Clustering?

Student ID	Gender	Major	Grade
SD-001	M	Math	A+
SD-002	M	Math	A
SD-003	F	Statistics	A+
SD-004	F	Math	A
SD-005	F	Math	B
SD-006	M	Statistics	B
SD-007	F	Physics	A+
SD-008	F	Physics	A
SD-009	M	Statistics	B+
SD-010	F	Math	B+



What is Clustering or Cluster Analysis?

- Clustering is the task of grouping a set of objects
- Unsupervised Learning model
- Discovering distinct groups in customer databases
- Used for creating strategies to adopt for certain segments

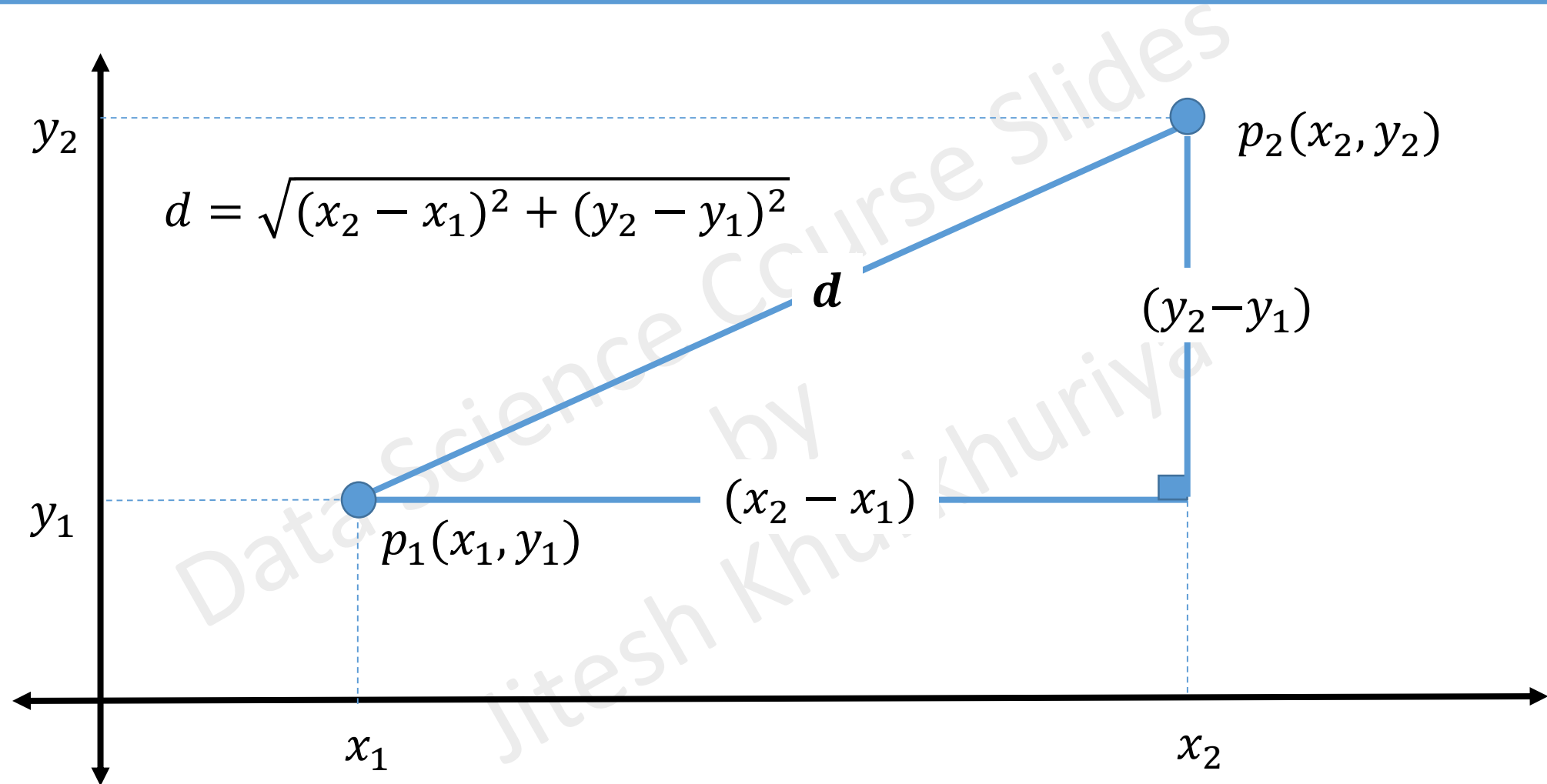


Examples of Clustering

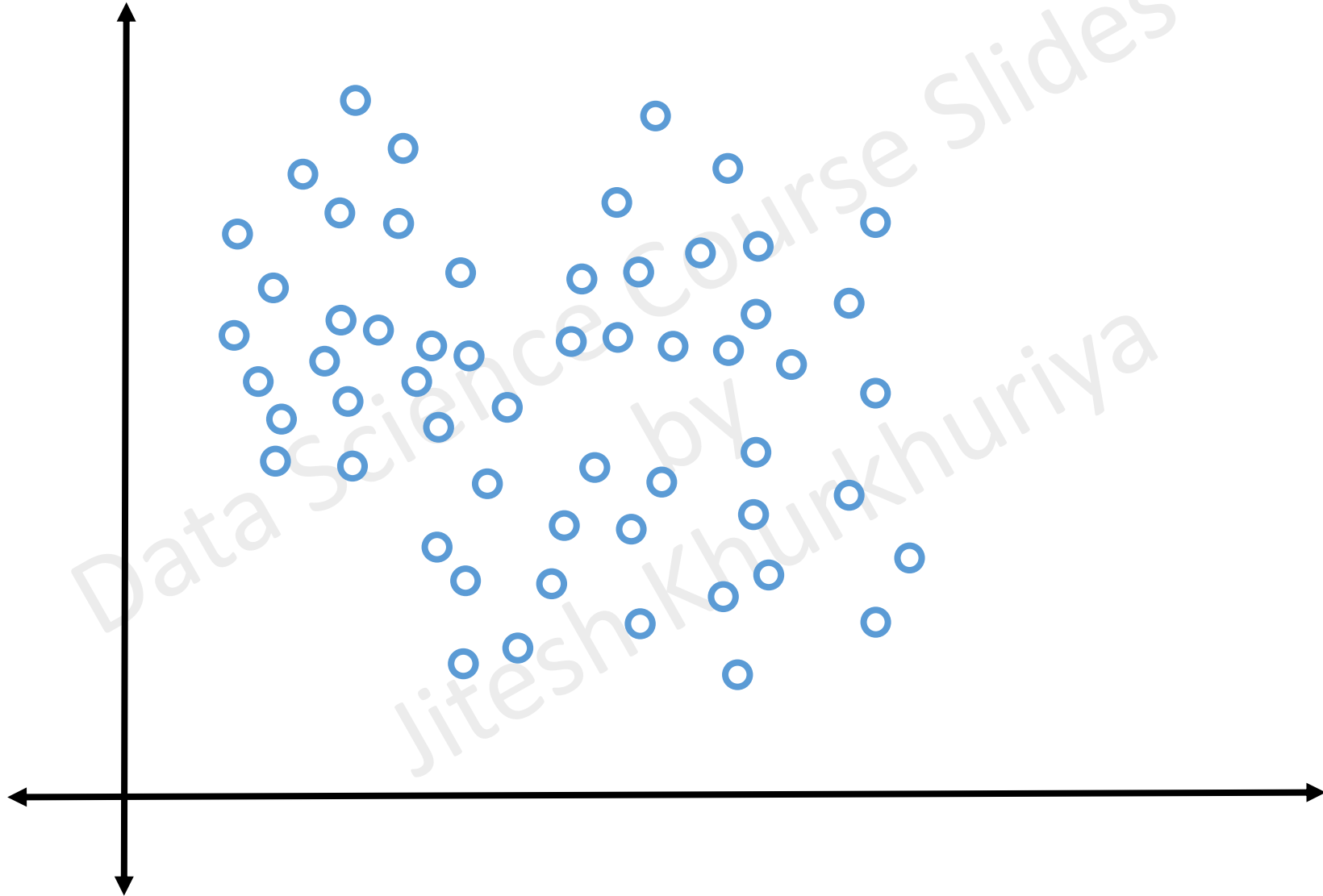
- Recommendation engines
- Market segmentation
- Social network analysis
- Medical/Health
- Image segmentation
- Anomaly detection

How Clustering works?

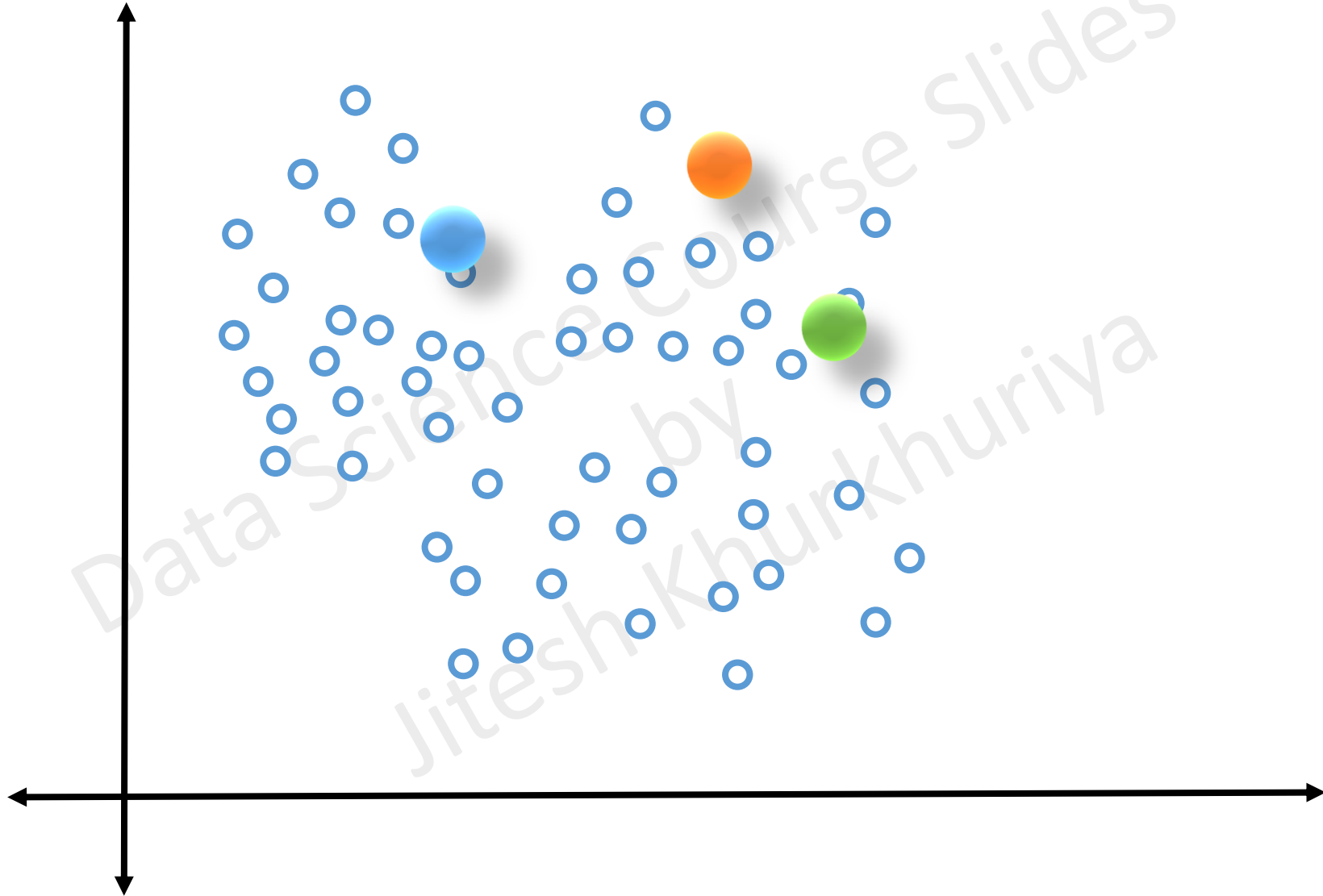
Euclidean Distance



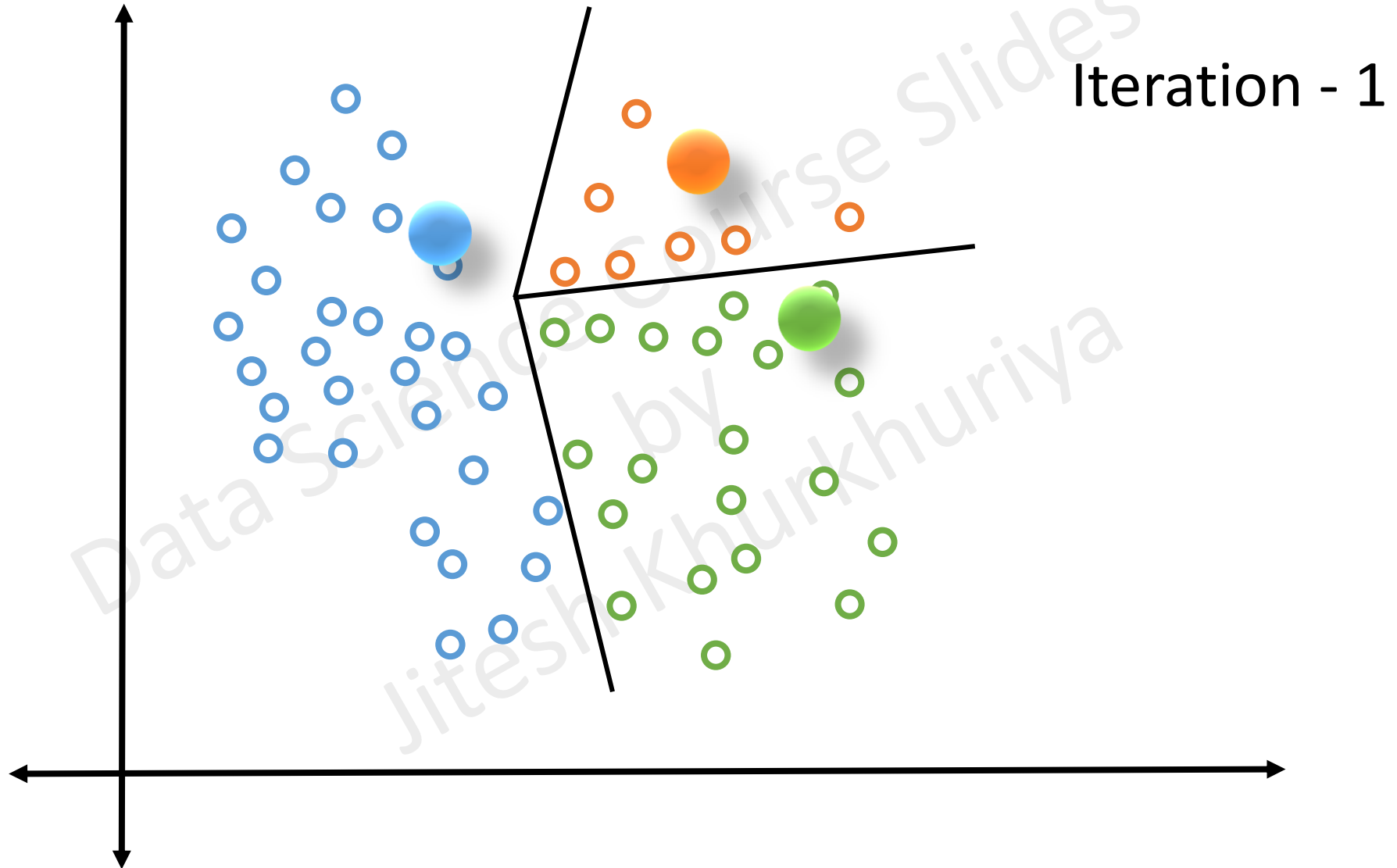
How Clusters are formed?



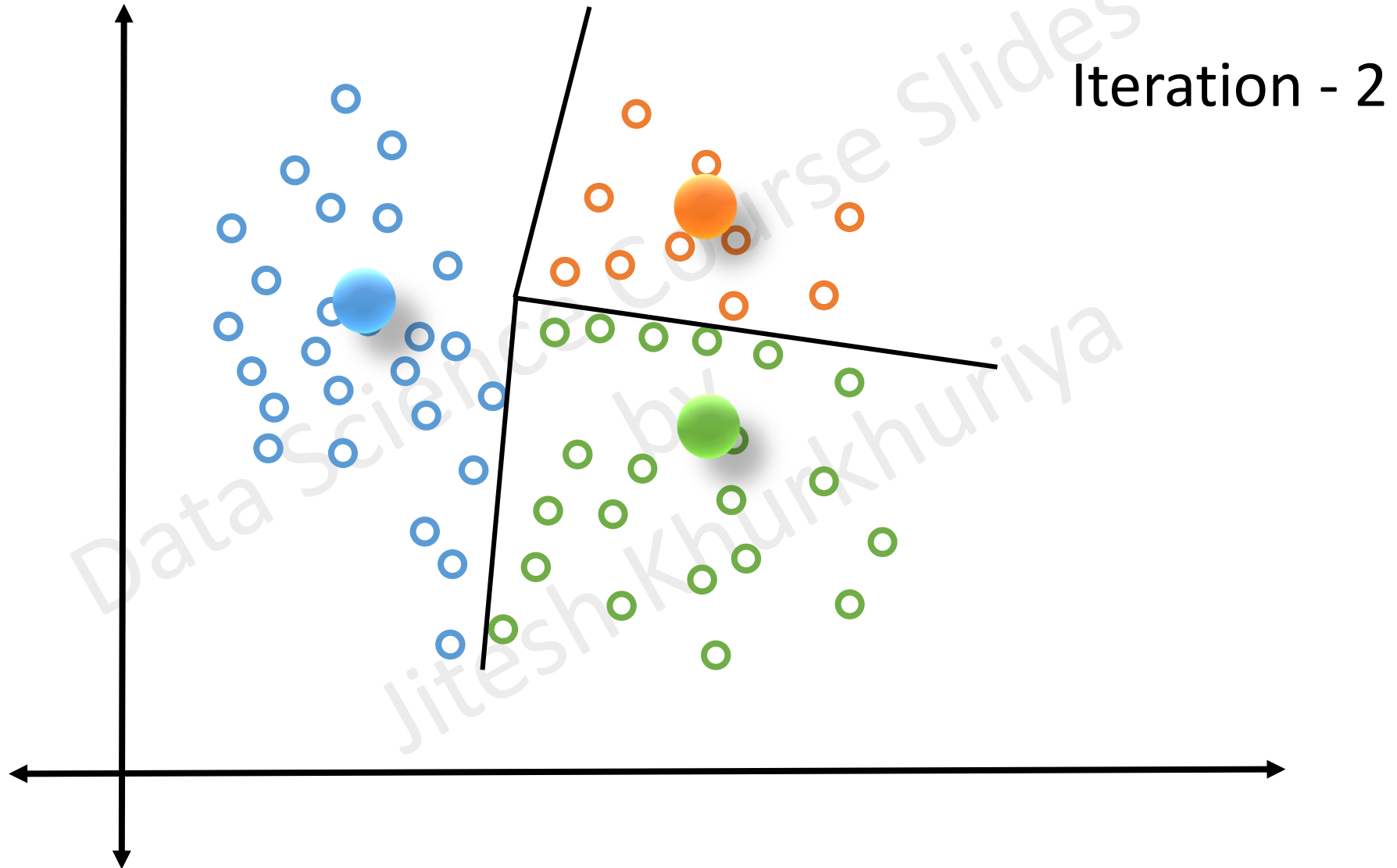
How Clusters are formed?



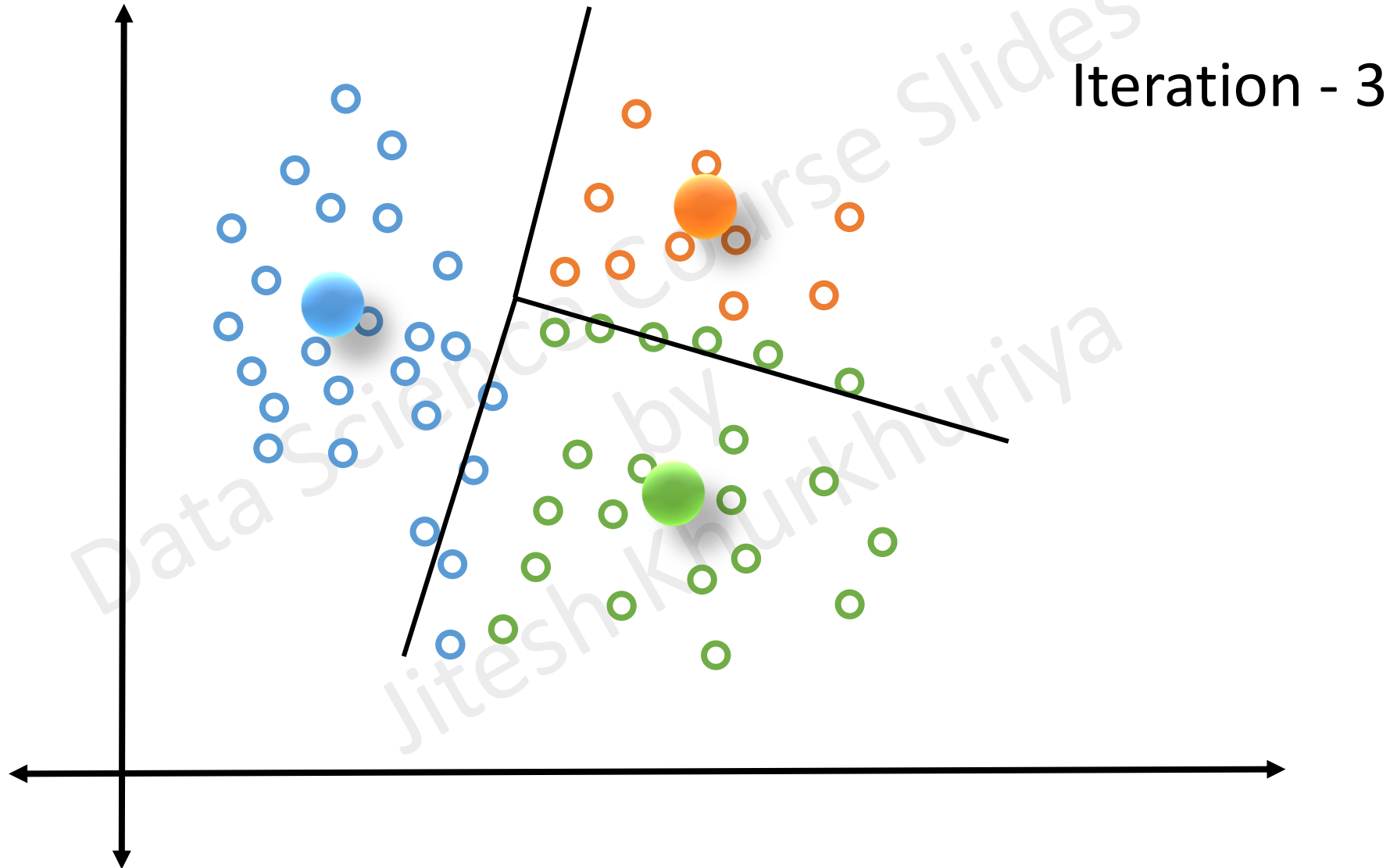
How Clusters are formed?



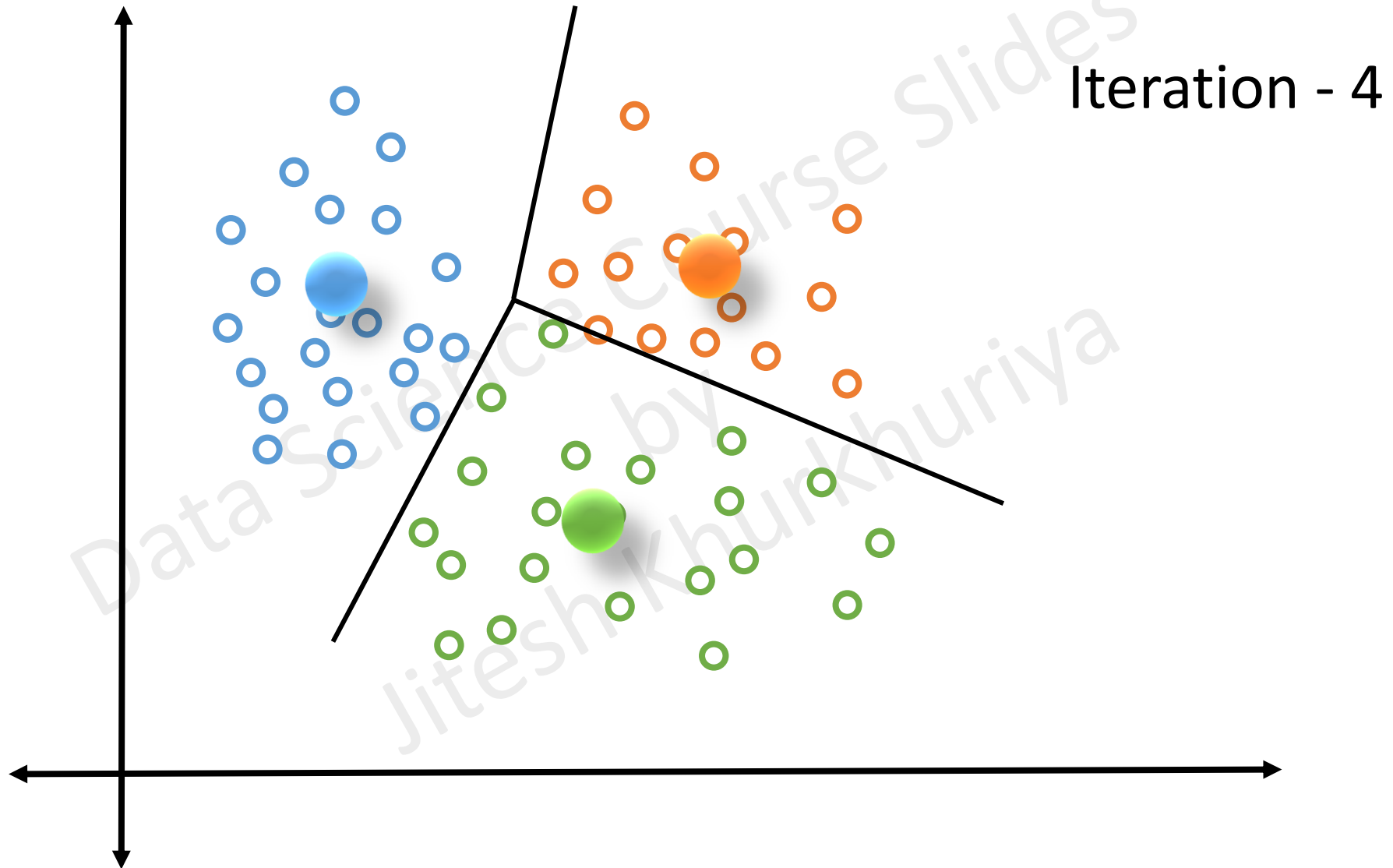
How Clusters are formed?



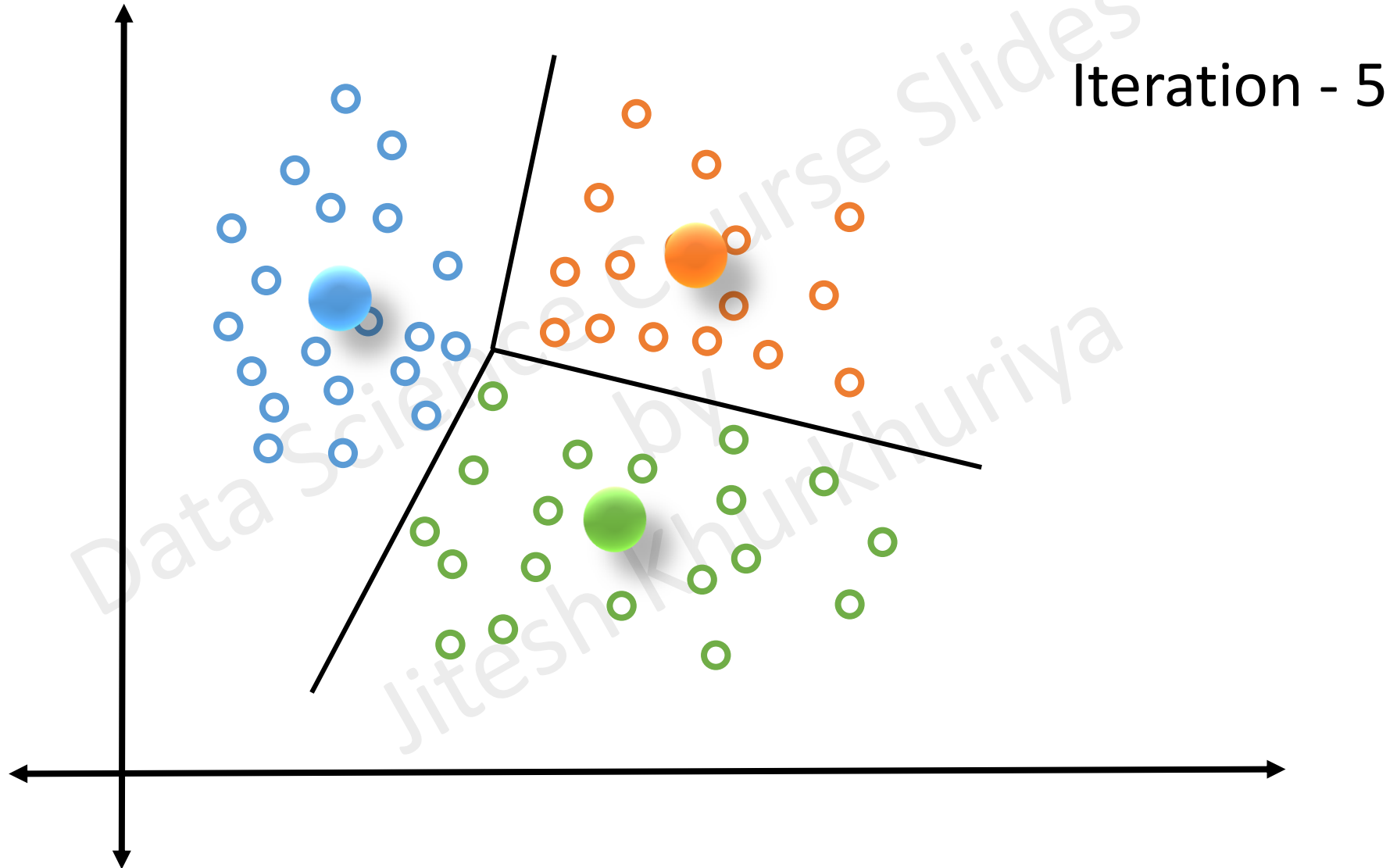
How Clusters are formed?



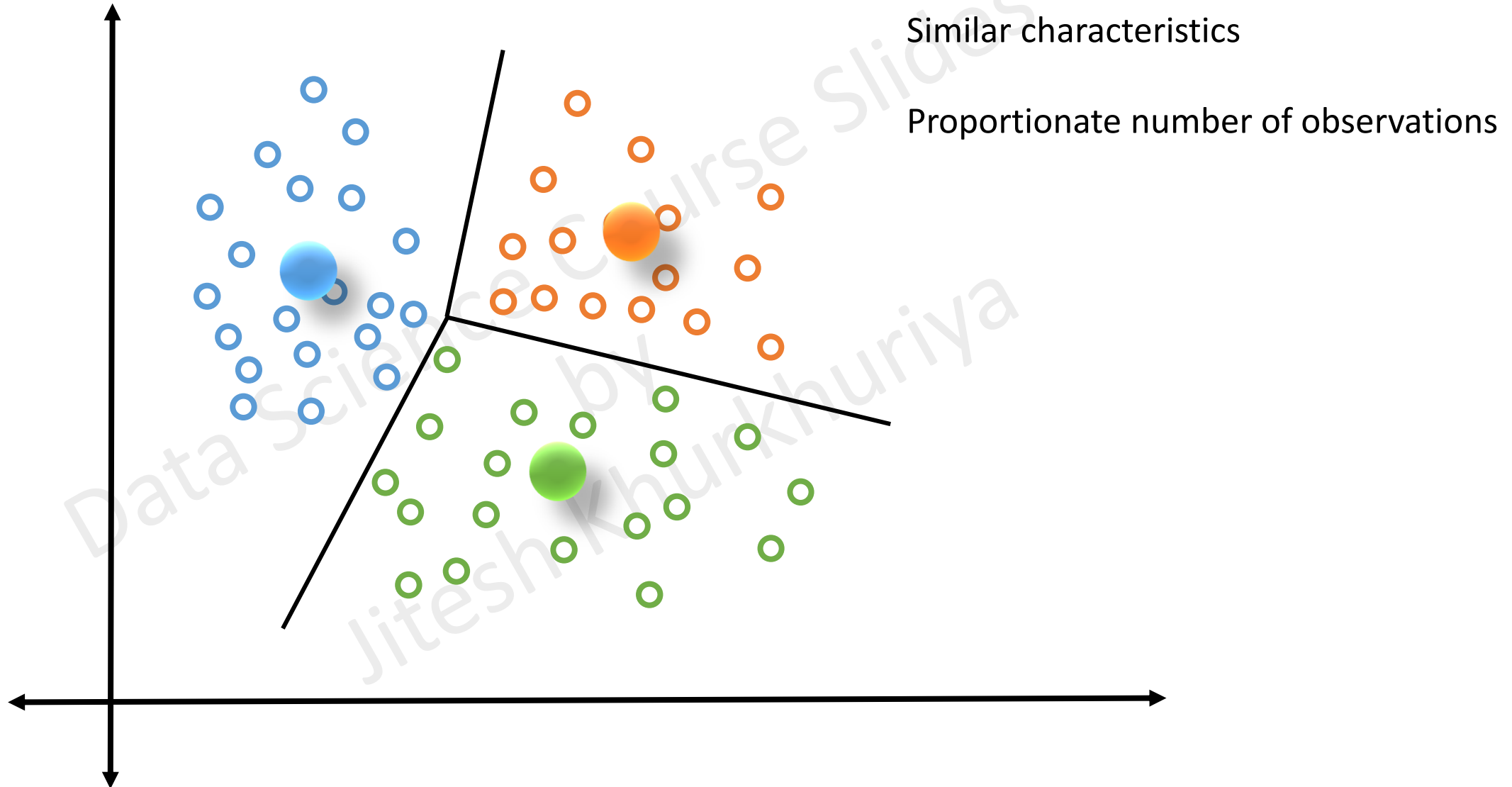
How Clusters are formed?



How Clusters are formed?

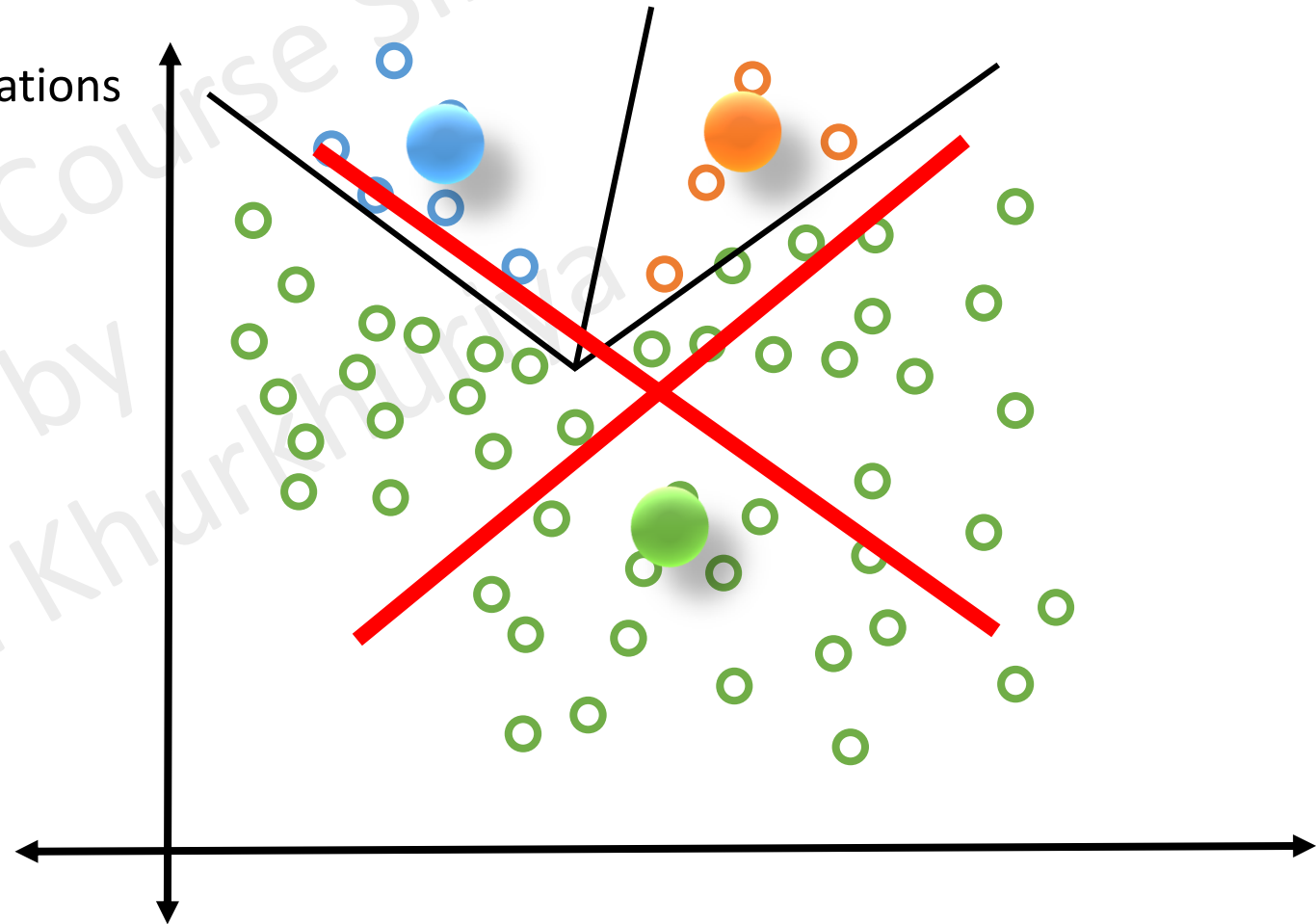


How Clusters are formed?



Good Cluster Analysis

- Observations in the same group share similar characteristics
- Clusters have proportionate number observations

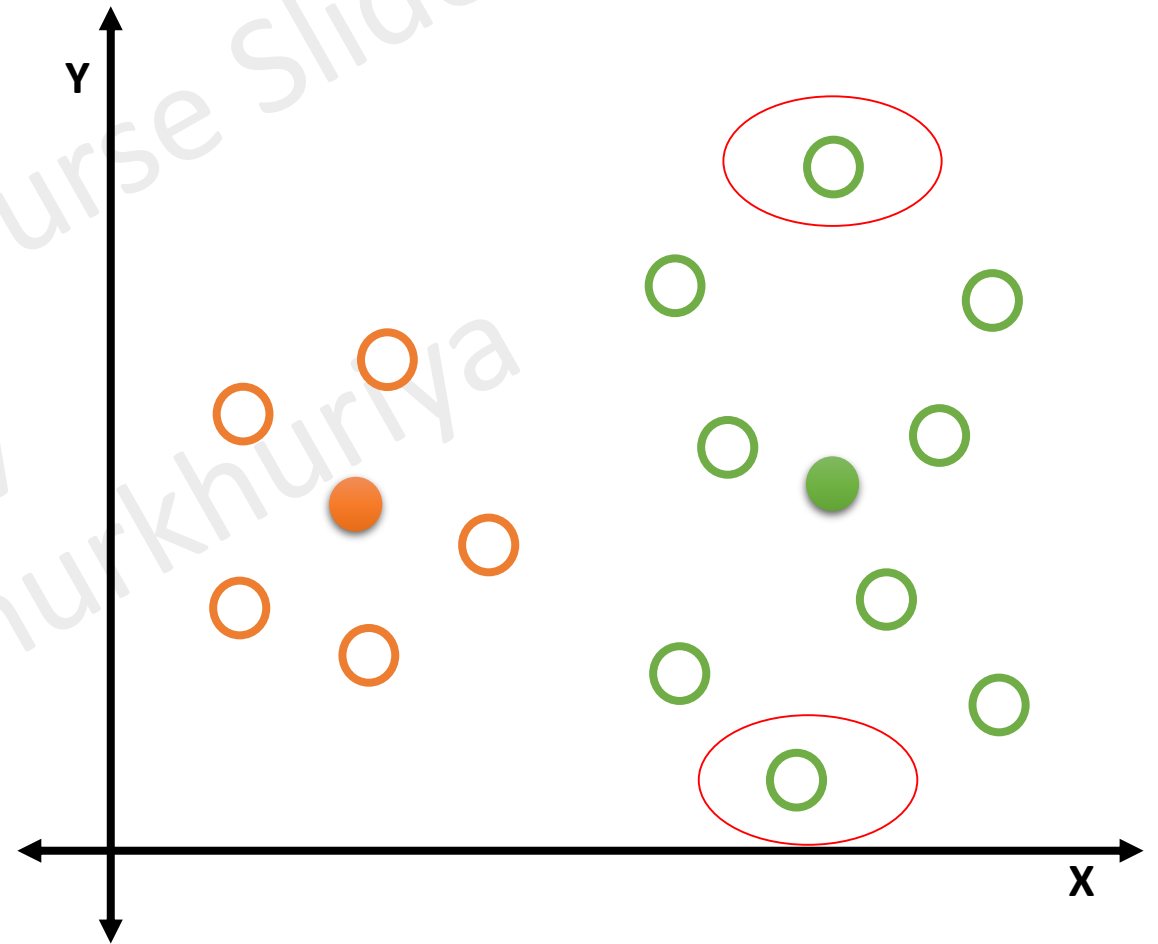
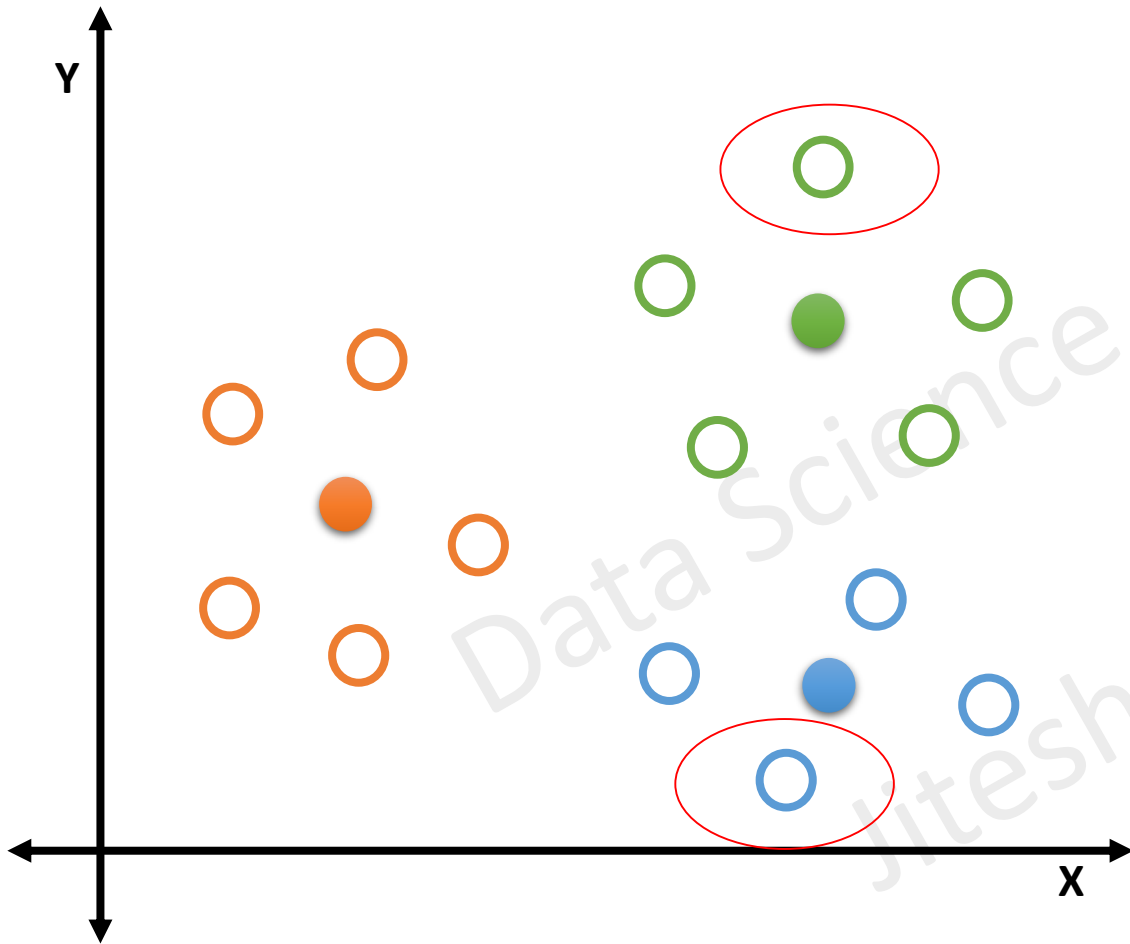


How to Decide Number of Clusters?

Observations in the same clusters are related to each other.

Distance of the observations from centroid decides their cluster assignment.

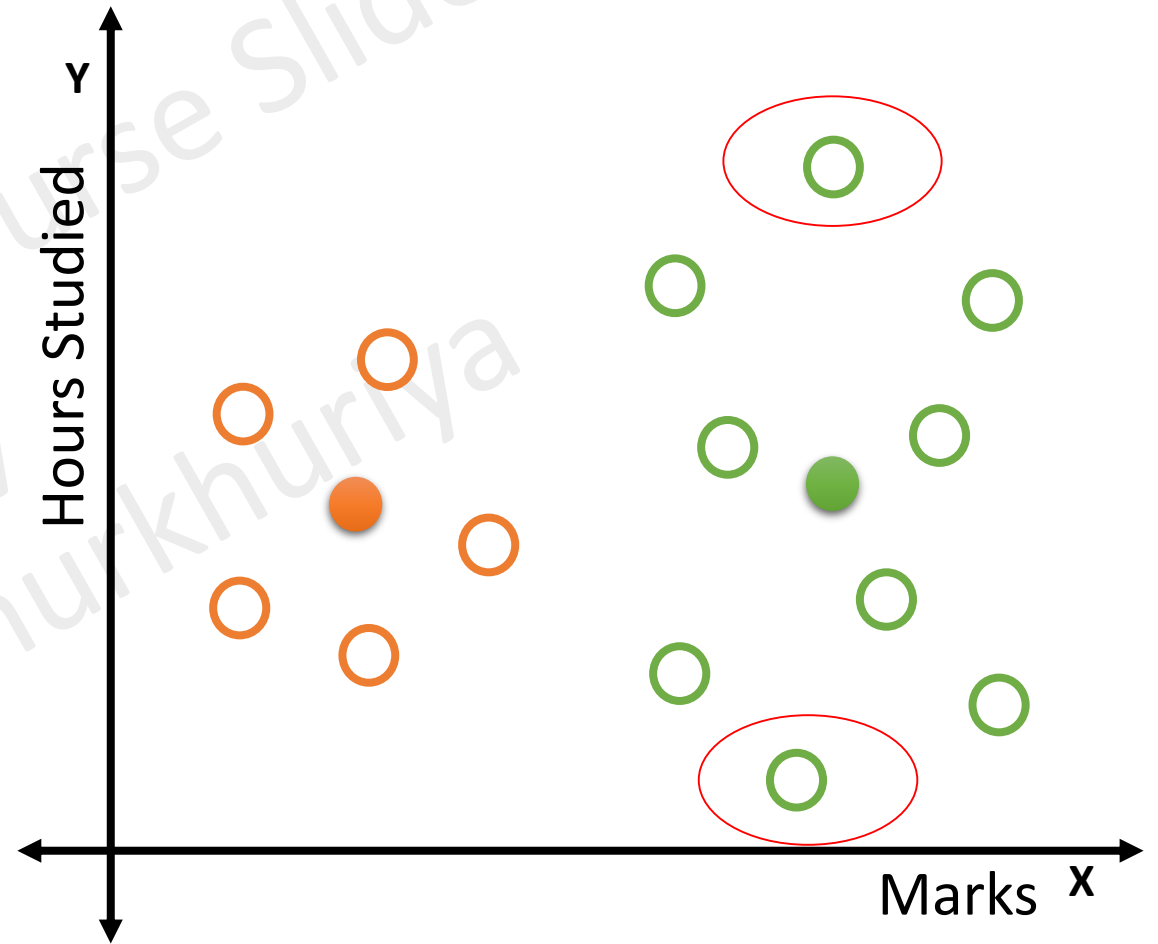
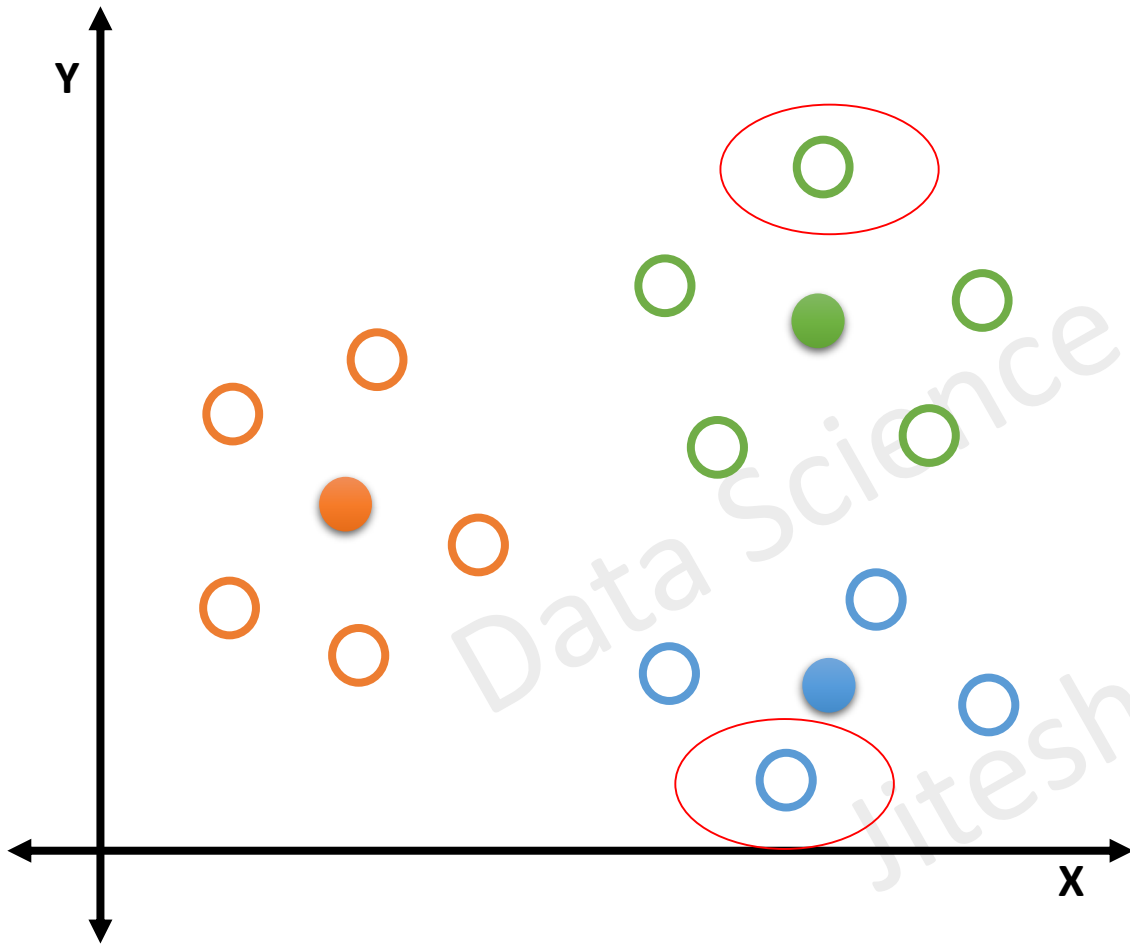
Lesser the distance,
Better the relationship



Observations in the same clusters are related to each other.

Distance of the observations from centroid decides their cluster assignment.

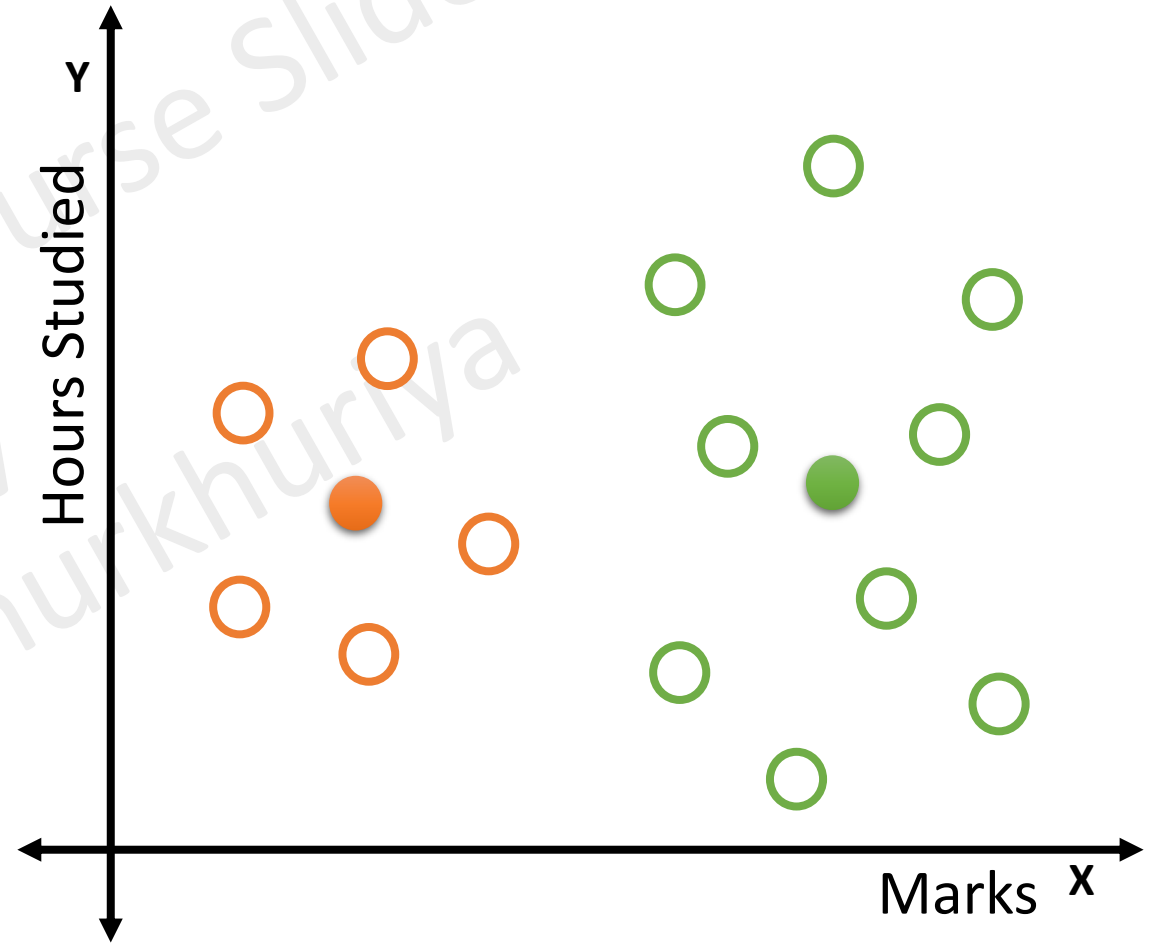
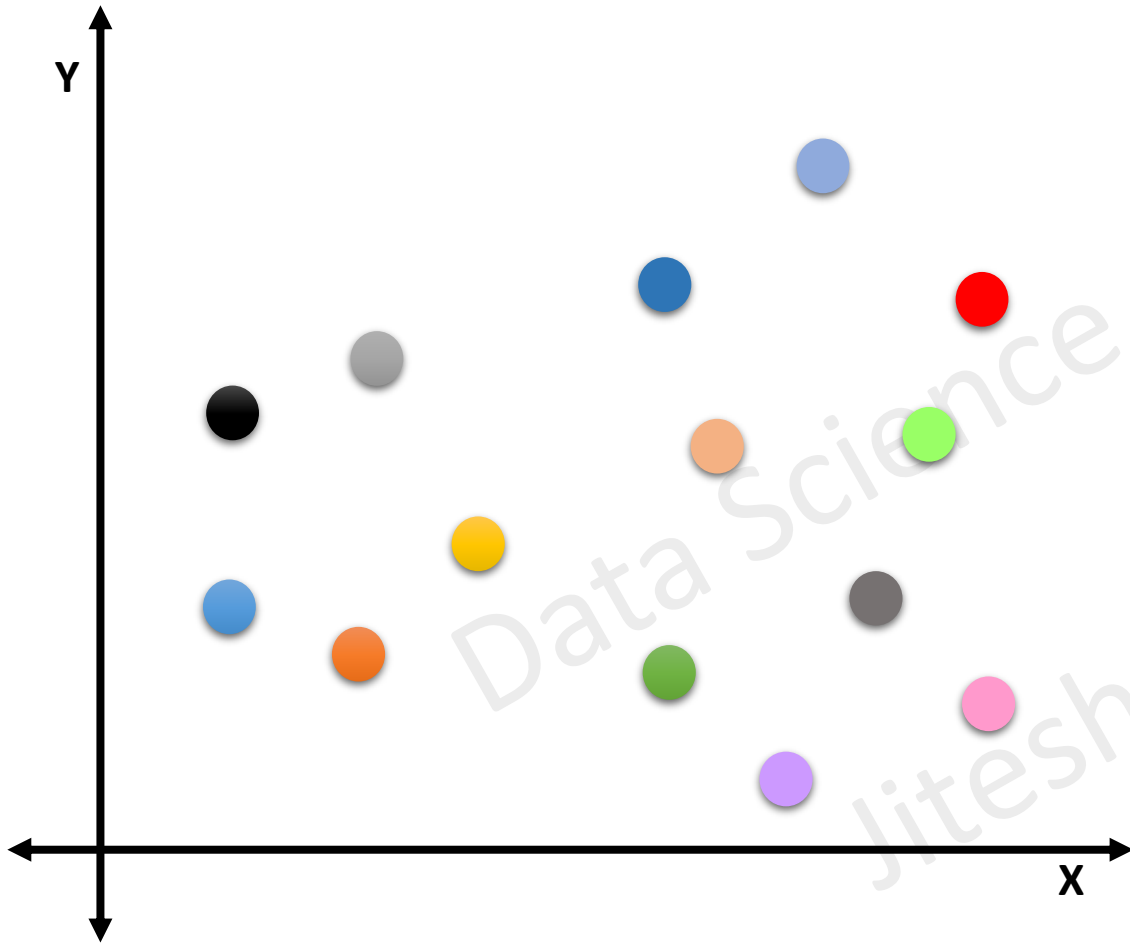
Lesser the distance,
Better the relationship



Observations in the same clusters are related to each other.

Distance of the observations from centroid decides their cluster assignment.

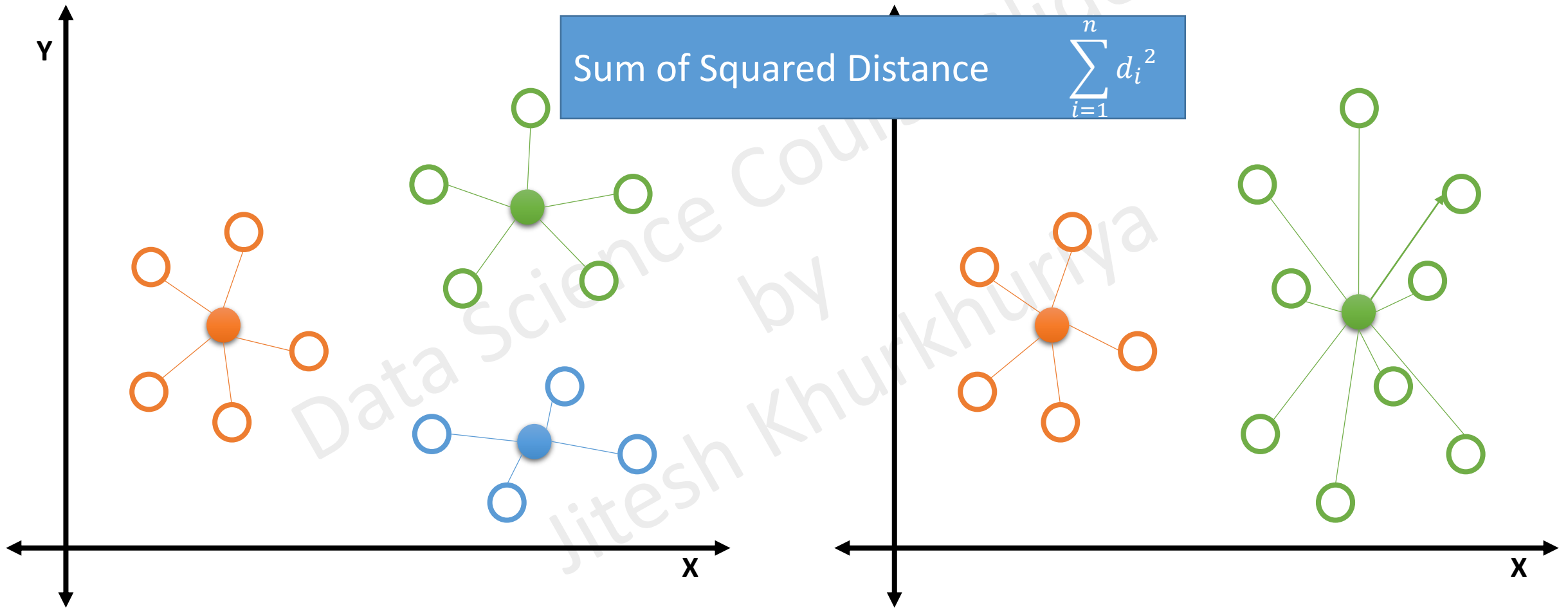
Lesser the distance,
Better the relationship



Observations in the same clusters are related to each other.

Distance of the observations from centroid decides their cluster assignment.

Lesser the distance,
Better the relationship

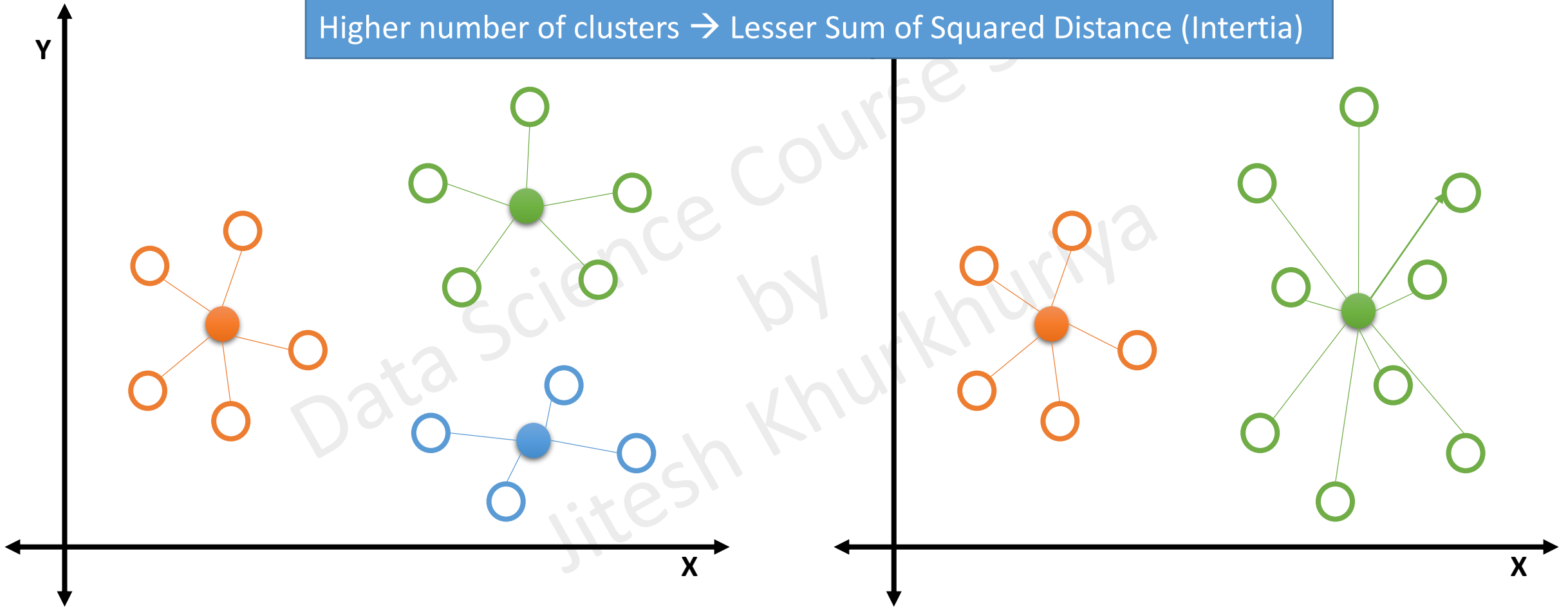


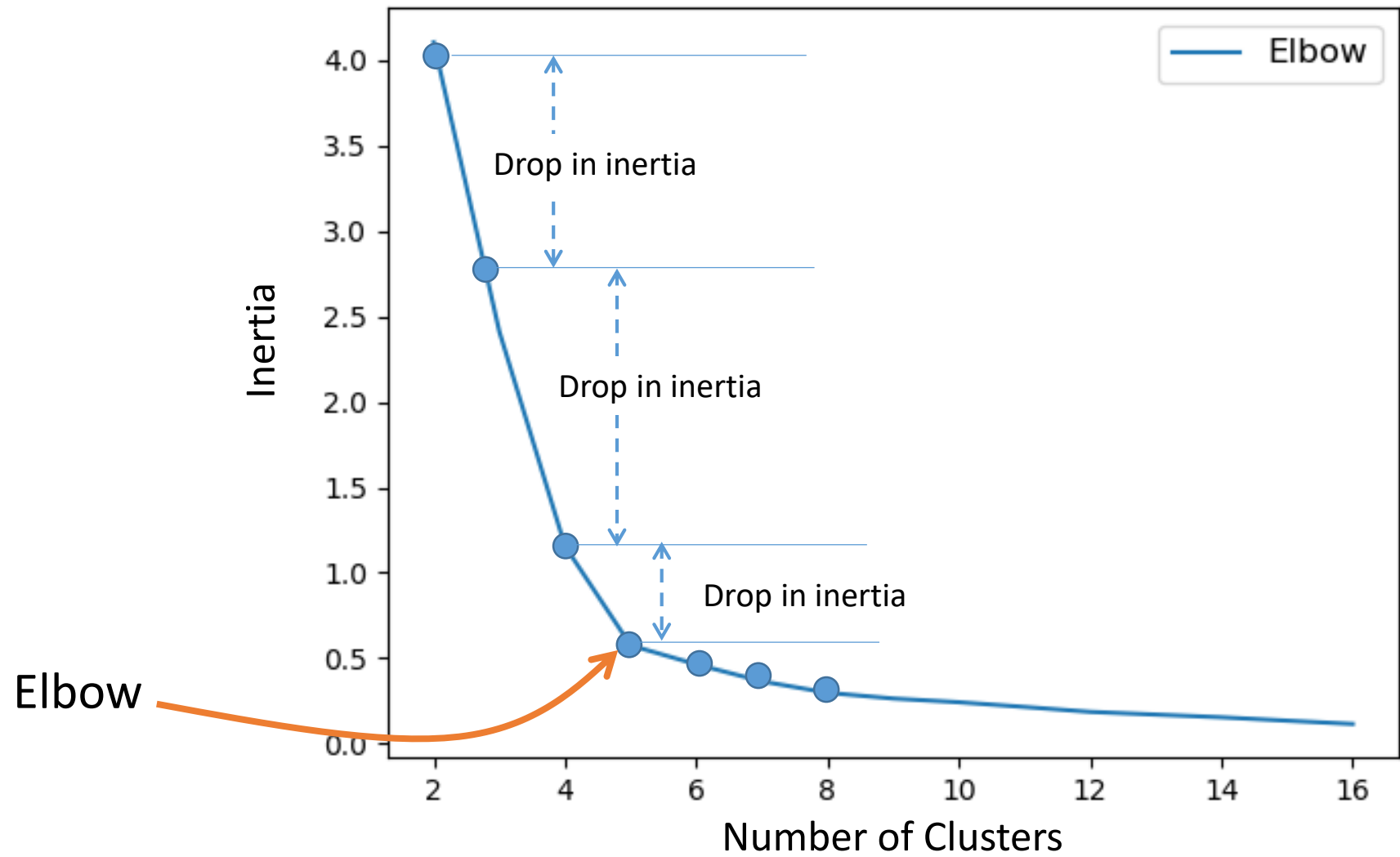
Observations in the same clusters are related to each other.

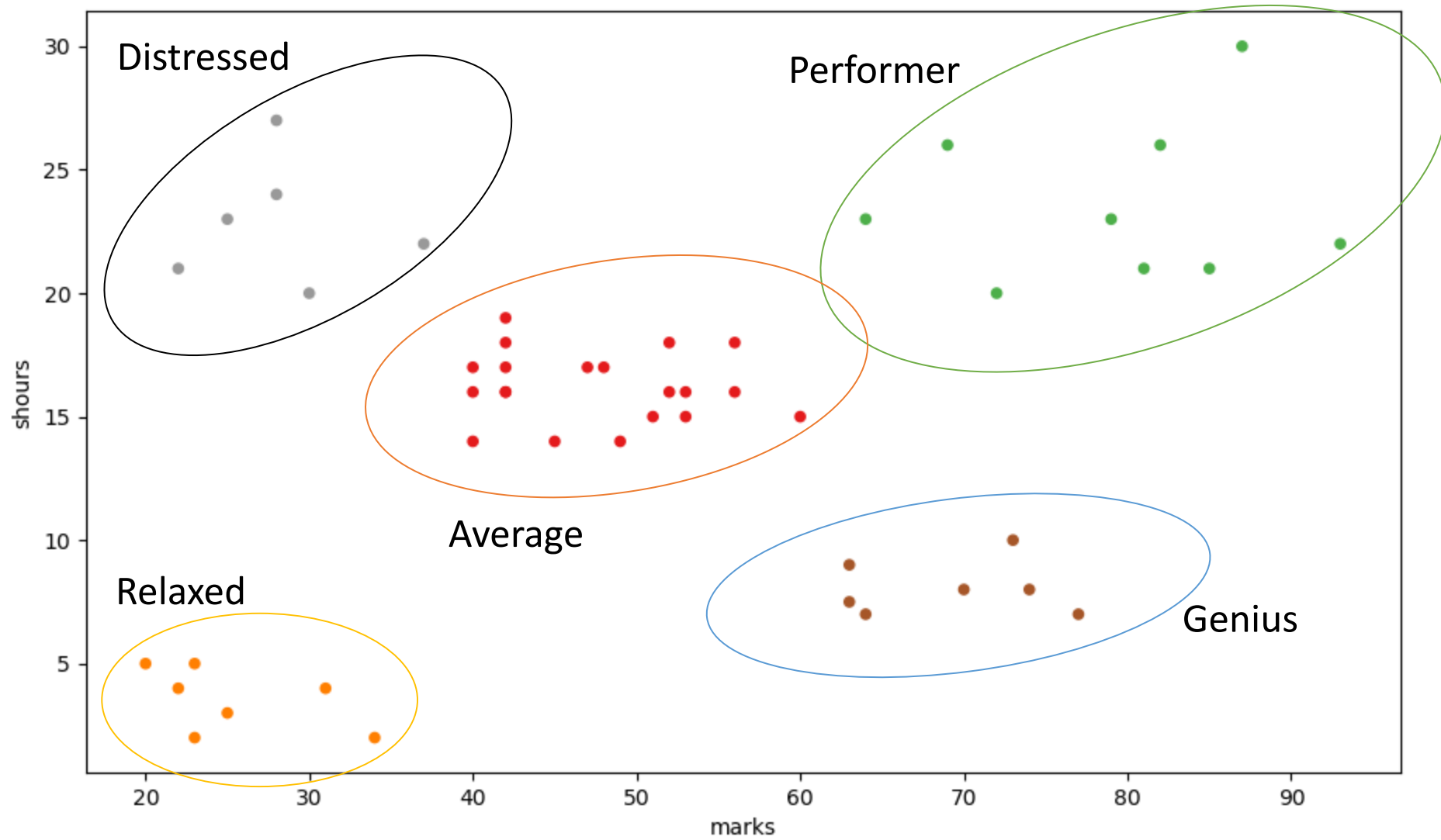
Distance of the observations from centroid decides their cluster assignment.

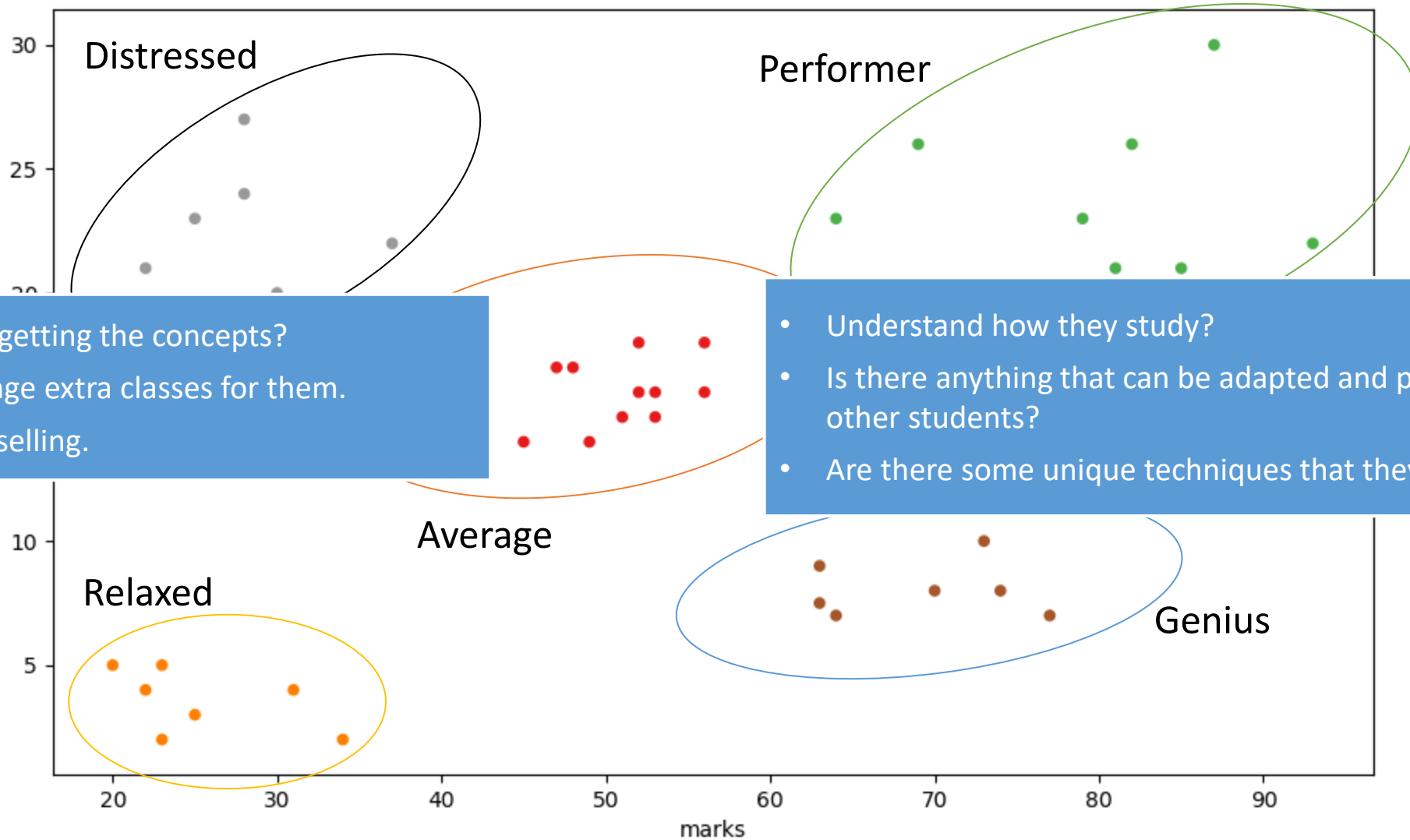
Lesser the distance,
Better the relationship

Higher number of clusters \rightarrow Lesser Sum of Squared Distance (Intertia)





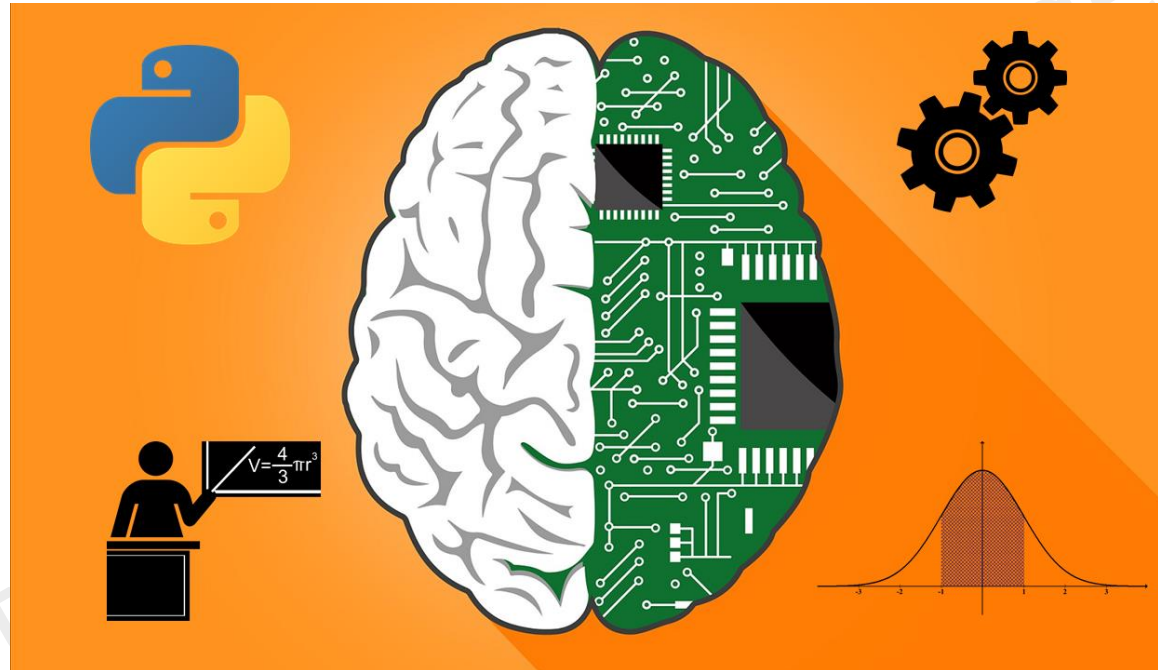




- Are they not getting the concepts?
- May be Arrange extra classes for them.
- May be counselling.

- Understand how they study?
- Is there anything that can be adapted and provided to other students?
- Are there some unique techniques that they are following?

Complete Data Science and Machine Learning Using Python



Thank You!