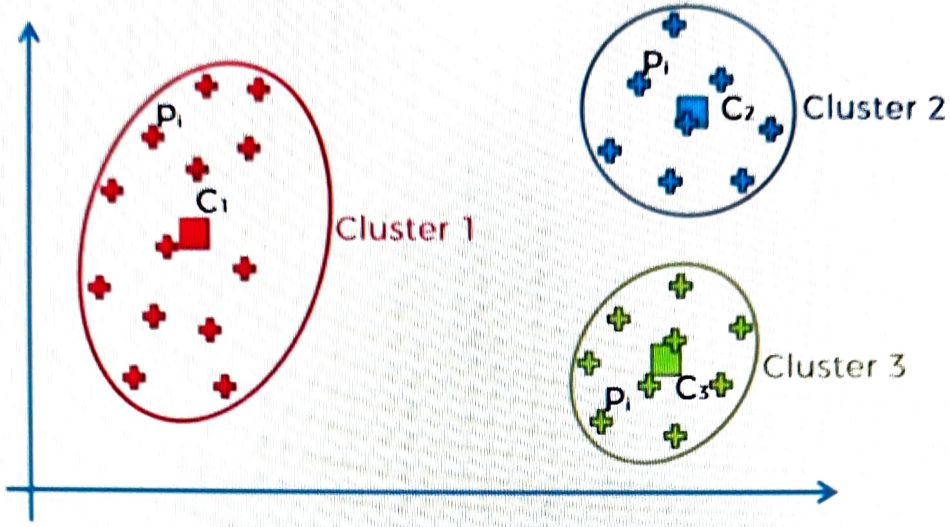
K-Means Intuition: Choosing the right number of clusters

Emmanuel Kamins algorithm which was lots of fun.

Choosing the right number of clusters



to be three then the result will look something like this.

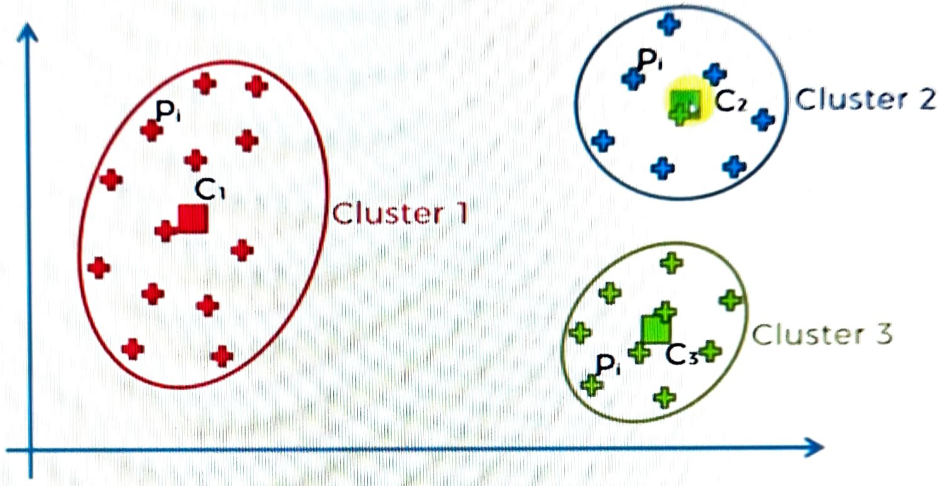
Choosing the right number of clusters

$$WCSS = \sum_{P_i \text{ in Cluster 1}} \text{distance}(P_i, C_1)^2 + \sum_{P_i \text{ in Cluster 2}} \text{distance}(P_i, C_2)^2 + \sum_{P_i \text{ in Cluster 3}} \text{distance}(P_i, C_3)^2$$

637 people have written a note here.

At first glance this formula might look a bit overwhelming or complex but in reality it's super simple

Choosing the right number of clusters



$$\text{WCSS} = \sum_{P_i \text{ in Cluster 1}} \text{distance}(P_i, C_1)^2 + \sum_{P_i \text{ in Cluster 2}} \text{distance}(P_i, C_2)^2 + \sum_{P_i \text{ in Cluster 3}} \text{distance}(P_i, C_3)^2$$

637 people have written a note here.

to every single point so to that point then we're square with squaring it to that point then risk wearing

Section 1: Welcome to the course! Here we will help you get started in the best conditions.

13 / 14 | 42min

Section 2: ---------- Part 1: Data Preprocessing ----------

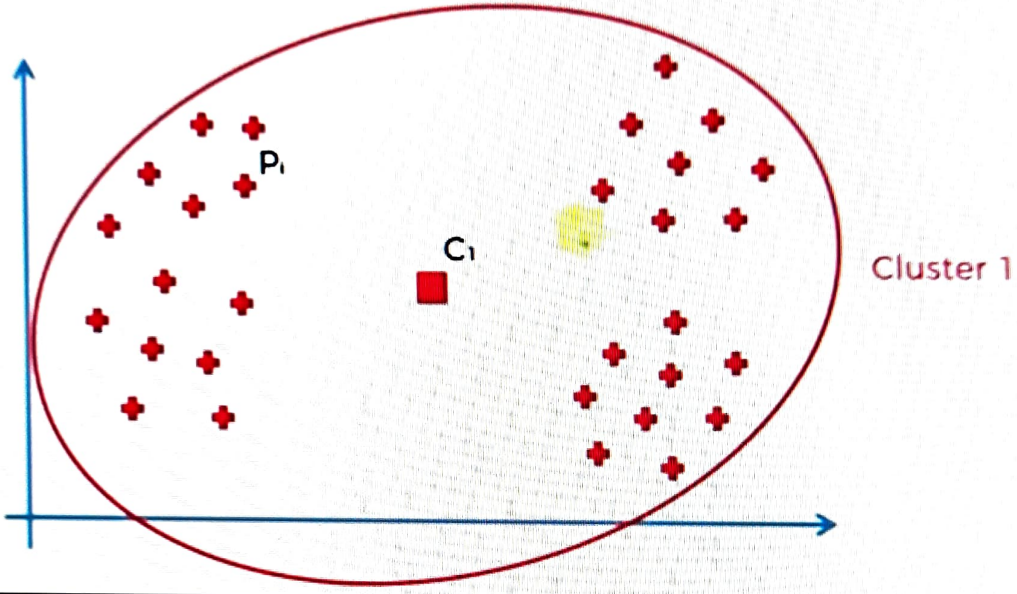
Choosing the right number of clusters

Rewind...

637 people have written a note here.

it's going to change as we increase the number of clusters.

Choosing the right number of clusters

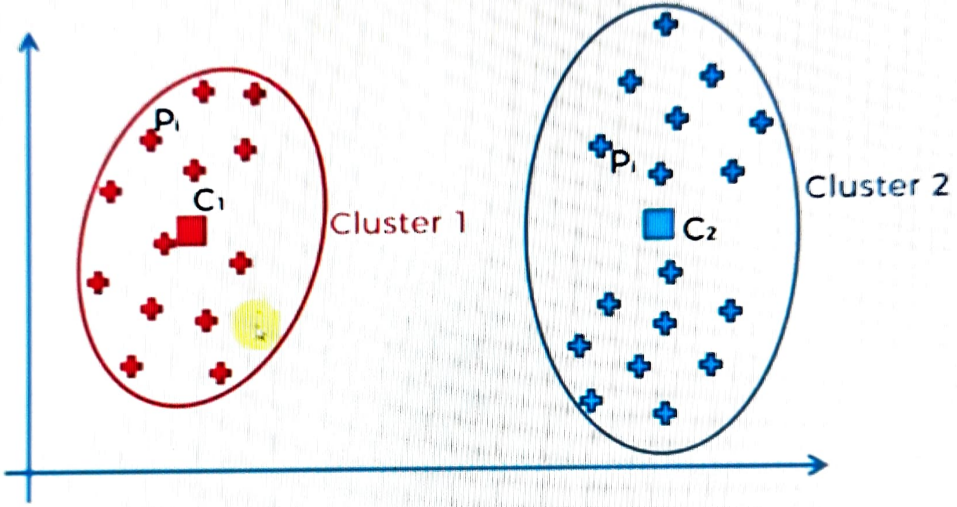


637 people have written a note here.

$\text{ance}(P_i, C_1)^2$

to reach towards it.

Choosing the right number of clusters

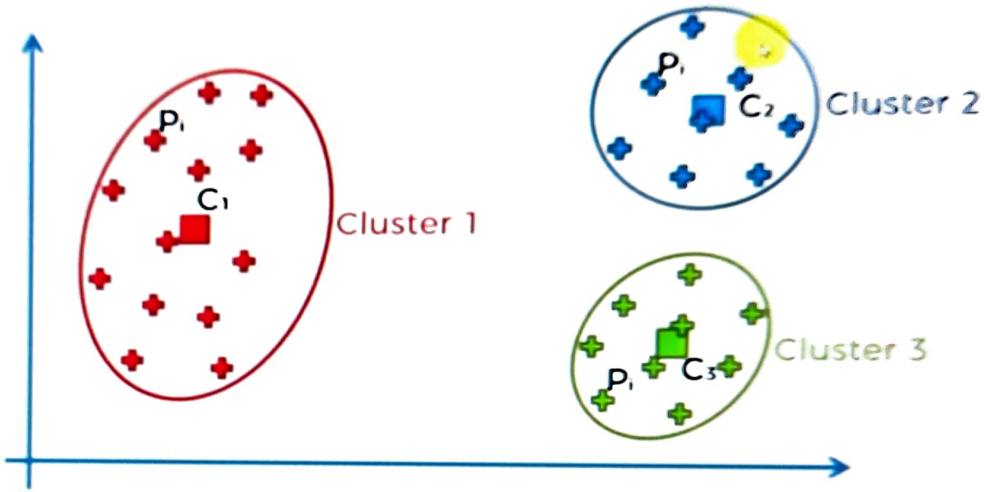


$$\sum \text{distance}(P_i, C_2)^2$$

637 people have written a note here.

All you have to do is just calculate distance to this centroid as it gets less and then for the Central

Choosing the right number of clusters

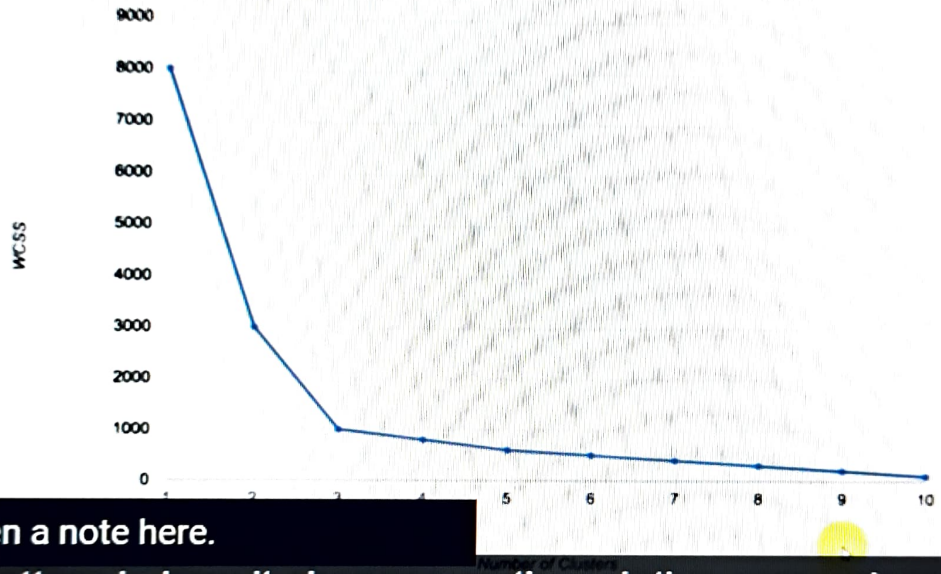


$$\text{WCSS} = \sum_{P_i \text{ in Cluster 1}} \text{distance}(P_i, C_1)^2 + \sum_{P_i \text{ in Cluster 2}} \text{distance}(P_i, C_2)^2 + \sum_{P_i \text{ in Cluster 3}} \text{distance}(P_i, C_3)^2$$

Machine Lea... DataScience

These two sums have changed so now it's sort of a one cluster.

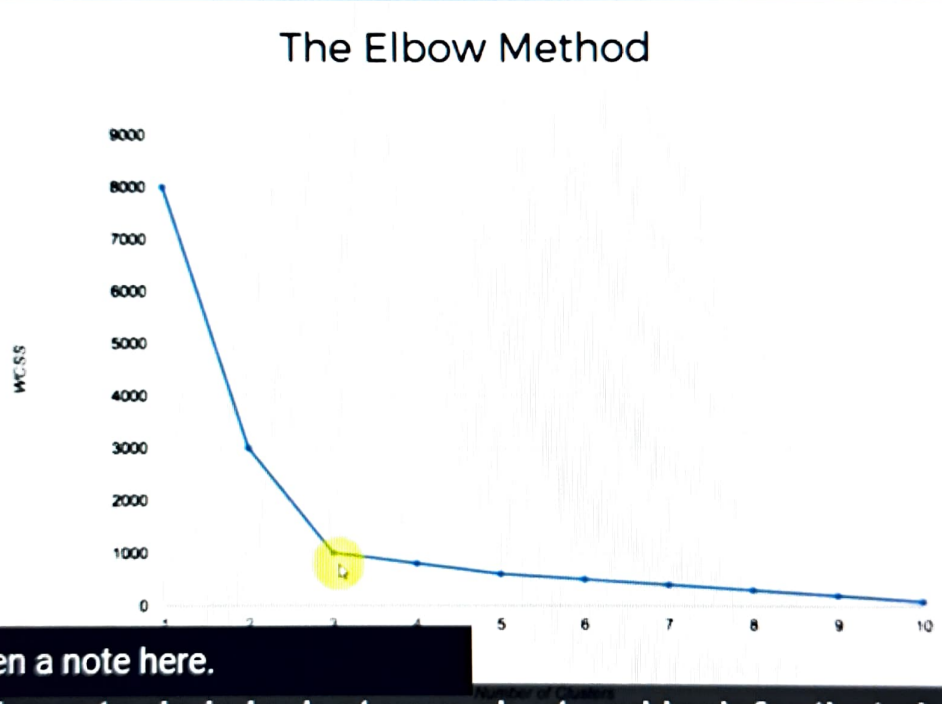
Choosing the right number of clusters



284 people have written a note here.

What matters is how it changes so the relative comparison between different k means clustering methods

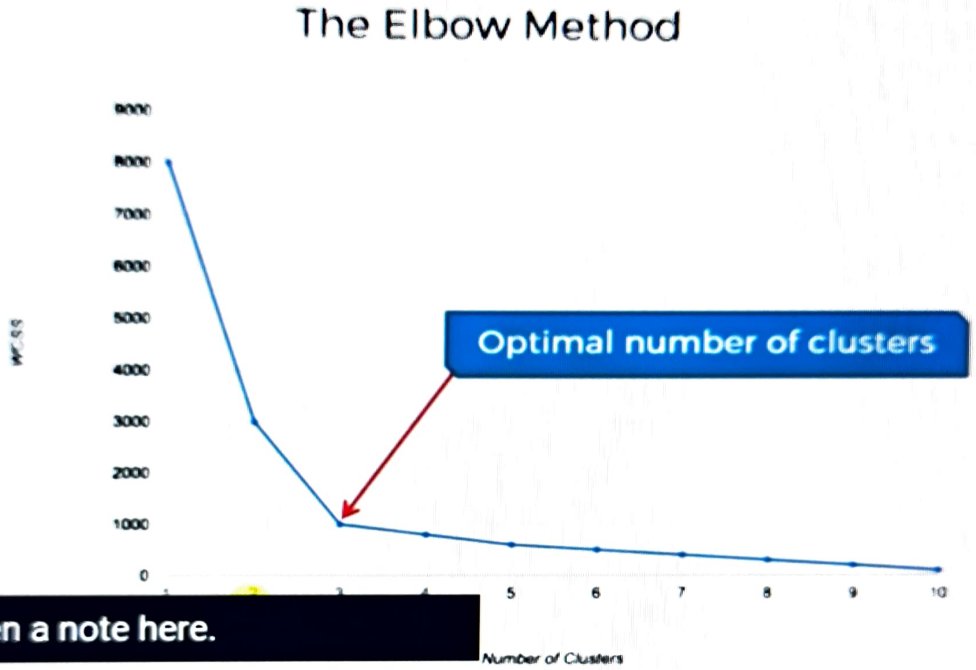
Choosing the right number of clusters



284 people have written a note here.

All you have to do is look at your chart and look for that change or that's kind of like it does look

Choosing the right number of clusters



453 people have written a note here.

And as you can imagine this method is quite arbitrary.