

Choosing the right number of clusters

K-Means Intuition: Choosing the right number of clusters

Machine Learning A-Z

© SuperDataScience

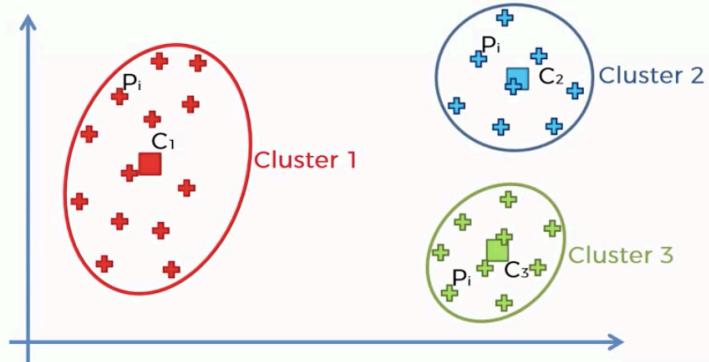
Choosing the right number of clusters



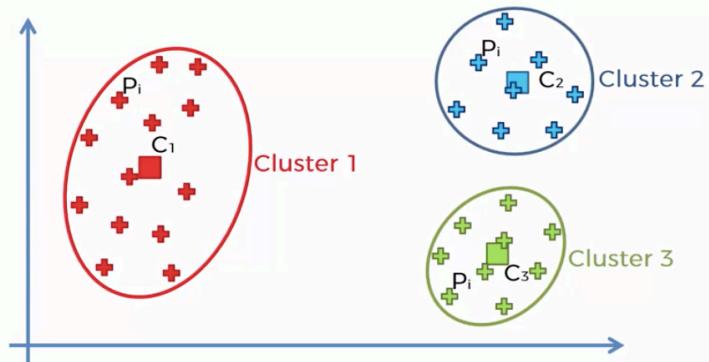
Machine Learning A-Z

© SuperDataScience

Choosing the right number of clusters



Choosing the right number of clusters

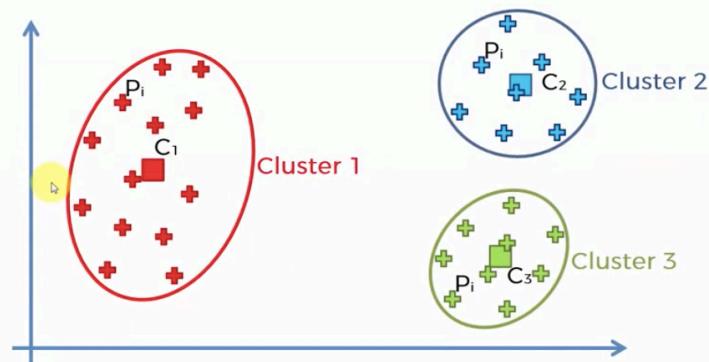


Choosing the right number of clusters

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

In here we want our WCSS be small as possible (not the smallest because the smallest is zero or in other word we should have clusters as much as our data point)

Choosing the right number of clusters



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

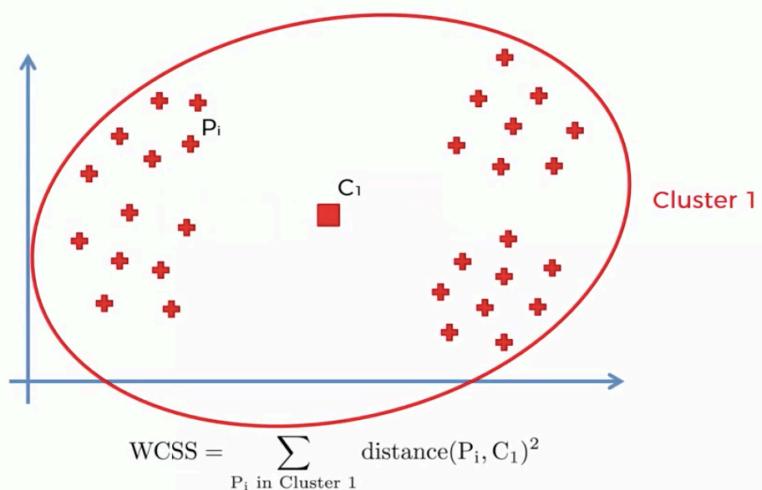
Choosing the right number of clusters

Rewind...

Machine Learning A-Z

© SuperDataScience

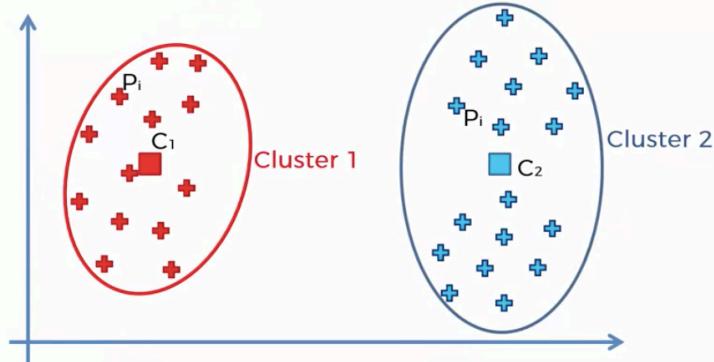
Choosing the right number of clusters



Machine Learning A-Z

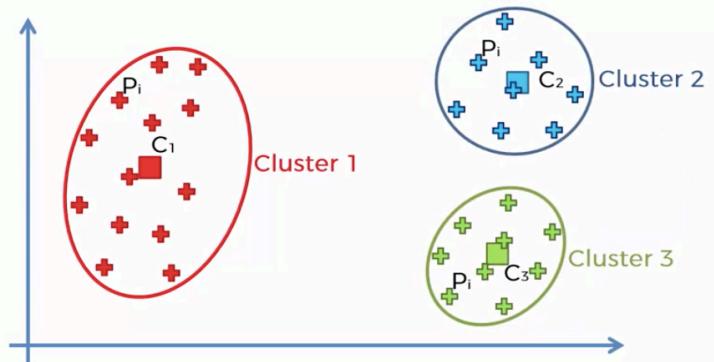
© SuperDataScience

Choosing the right number of clusters



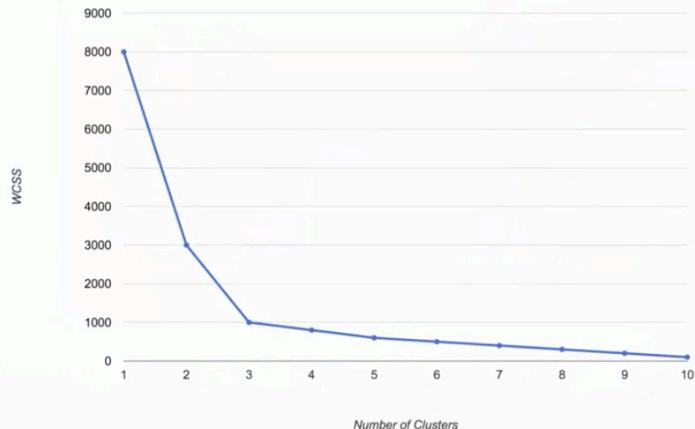
$$\text{WCSS} = \sum_{P_i \text{ in Cluster 1}} \text{distance}(P_i, C_1)^2 + \sum_{P_i \text{ in Cluster 2}} \text{distance}(P_i, C_2)^2$$

Choosing the right number of clusters



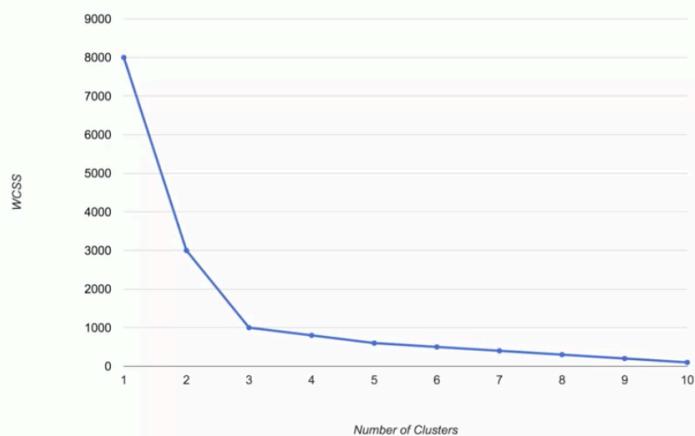
$$\text{WCSS} = \sum_{P_i \text{ in Cluster 1}} \text{distance}(P_i, C_1)^2 + \sum_{P_i \text{ in Cluster 2}} \text{distance}(P_i, C_2)^2 + \sum_{P_i \text{ in Cluster 3}} \text{distance}(P_i, C_3)^2$$

Choosing the right number of clusters



Choosing the right number of clusters

The Elbow Method



The Elbow Method means that we should have substantial decrease between different clusters in order to accept it. Here the shape is look like a arm which we have to choose it's elbow. In here we have a substantial decrease until 3 and non-substantial decrease afterward so we choose 3 as our optimal number of clusters.

Choosing the right number of clusters

The Elbow Method

