

Number of Clusters and the Elbow Rule: Takeaways



by Dataquest Labs, Inc. - All rights reserved © 2024

Concepts

- The inertia is the measure of how far the data points assigned to a cluster are from that cluster's centroid.
- Mathematically, inertia is the sum of the squared distances from each data point to the centroid as given by the following formula: $\sum_{i=1}^n (x_i - c_k)^2$.
- The ideal clustering is made of low inertia and low number of **k**, but this is a trade-off. The objective is to find the lower number of clusters that produces the lowest inertia.
- The Elbow Curve is a line chart of the inertias over the number of clusters. As the values on the y-axis decrease, the values on the x-axis increase.
- As the number of clusters increases, the decrease in inertia becomes lower and lower, which tends to create a sharp "elbow" on the curve.
- The ideal number of clusters is where inertia begins to decrease at lower rates.

Resources

- [Elbow Curve](#)
- [Diminishing Returns](#)

Takeaways by Dataquest Labs, Inc. - All rights reserved © 2024