

★ Get unlimited access to the best of Medium for less than \$1/week. [Become a member](#)



The Elbow Method: Finding the Optimal Number of Clusters



ZalaRushirajsinh

Follow

4 min read · Nov 4, 2023



31



| *A step-by-step guide to Elbow Method*

A Clustering is a fundamental technique in data analysis and machine learning that involves **grouping similar data points** based on their characteristics. Whether you're working with customer segmentation, image compression, or any other application that involves grouping data, determining the optimal number of clusters is crucial for obtaining results. In this blog, we'll explore what the Elbow Method is, how it works, and its applications in various fields of Machine learning.

1. *What is the Elbow Method?*

The Elbow Method is a visual approach used to determine the ideal 'K' (number of clusters) in K-means clustering. It operates by calculating the Within-Cluster Sum of Squares (WCSS), which is the total of the squared distances between data points and their cluster center. However, there is a point where increasing K no longer leads to a significant decrease in WCSS,

and the rate of decrease slows down. This point is often referred to as the **elbow**.

1.1 How does the Elbow Method work?

The Elbow Method is a simple but effective technique used to determine the optimal number of clusters (K) in a K-Means clustering algorithm. It examines the relationship between the number of clusters and the within-cluster sum of squares (WCSS), a measure of the variance within each cluster.

1.2 Let's go through the steps involved in K-means clustering for better understanding:

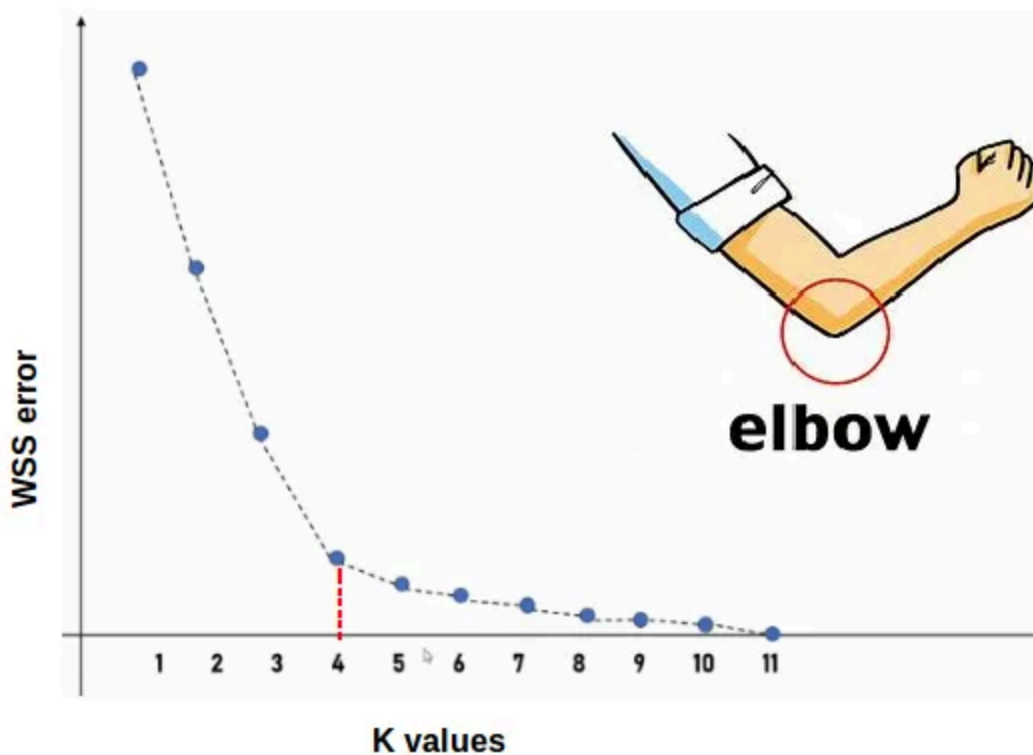
- Begin by selecting the desired number of clusters, denoted as K.
- Randomly select K data points from the dataset to serve as initial cluster centroids.
- Will use either Euclidean distance or Manhattan distance to measure or compute the distance between data points and the nearest centroid. Assign each data point to its nearest cluster centroid, thus forming K clusters.
- Calculate the new centroids for these clusters.
- Once again, reassign all data points to clusters based on the newly computed centroids and then repeat step 4. This process is iteratively carried out for a set number of iterations or until the centroid positions remain unchanged.

Identifying the ideal number of clusters is a crucial step in this algorithm. The **Elbow Method** is a widely used technique to determine the optimal value of K.

K Means Clustering Using the Elbow Method

In the Elbow Method, we systematically experiment with different numbers of clusters (K) ranging from 1 to 10. With each K value, we compute the Within-Cluster Sum of Squares (WCSS). When we plot WCSS against K, the resulting graph resembles an elbow. As we increase the number of clusters, the WCSS value begins to decrease. Notably, the WCSS is at its highest when $K=1$.

Elbow method



1. Graphical Representation of Elbow Method

Practical Implementation of the Elbow Method

- The dataset we are using here is the People working in the Company (<https://github.com/nishithkotak/machine-learning/blob/master/KMeans.csv>). It's data that contains the details of a

people working in a company(features like age, Annual Income(k\$), and spending score[1–100])

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

X=pd.read_csv('/content/KMeans.csv')
X

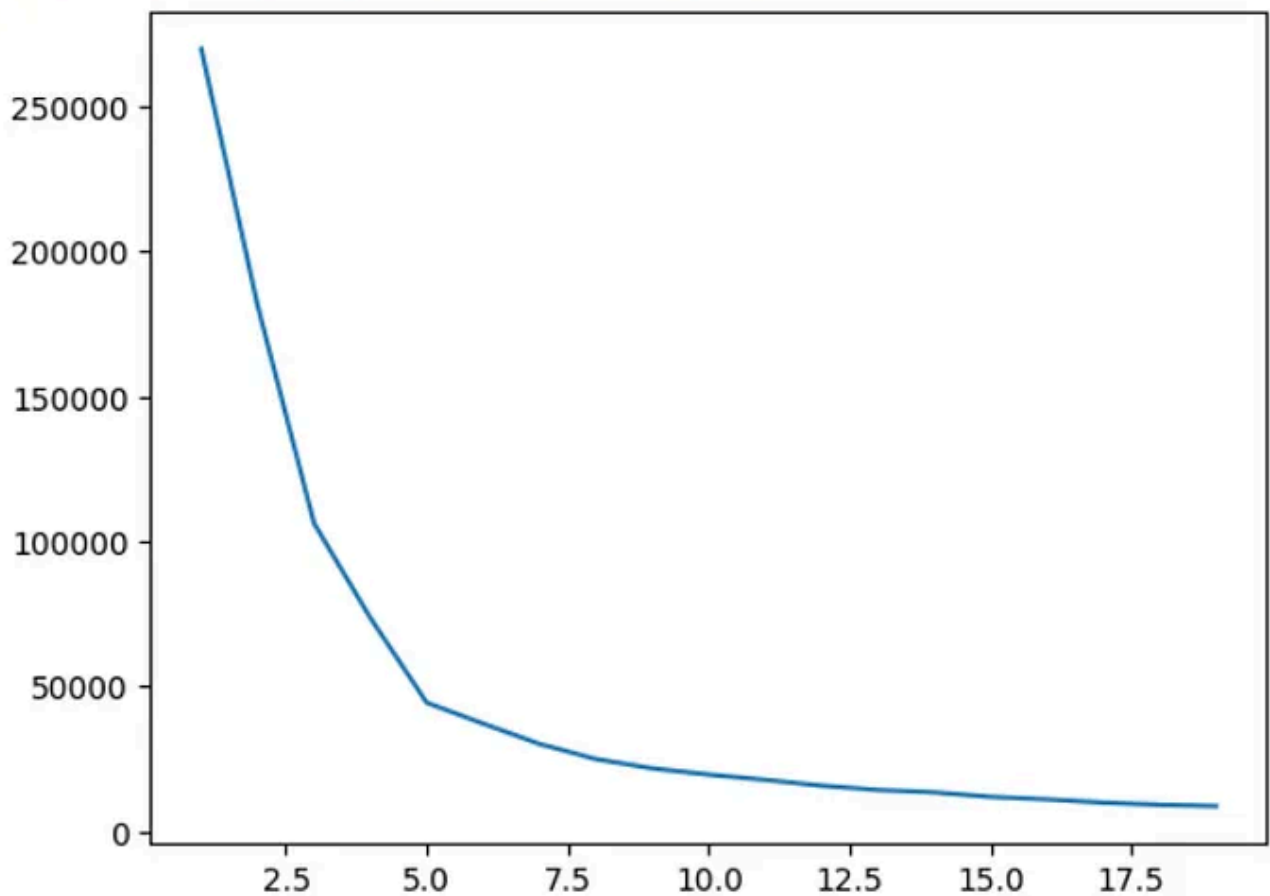
dataset=X.iloc[:,[1,2]].values
dataset

from sklearn.cluster import KMeans
WCSS=[]
for i in range(1,20):
    kmeans=KMeans(n_clusters=i,init='k-means++')
    kmeans.fit(dataset)
    WCSS.append(kmeans.inertia_)
WCSS

plt.plot(range(1,20),WCSS)
```

OUTPUT

[<matplotlib.lines.Line2D at 0x7908c14cf2e0>]



1.1 Output

2. Applications of the Elbow Method:

- 1. Customer Segmentation:** Businesses often use the Elbow Method to determine the optimal number of customer segments for personalized marketing and product recommendations. It helps identify distinct customer groups based on their behaviors, preferences, and demographics.
- 2. Image Compression:** In image processing, the Elbow Method can be applied to find the optimal number of colors or features to represent an image efficiently. This is crucial for reducing the storage space or bandwidth required for image data.
- 3. Outliers Detection:** Anomaly detection is the identification of rare or unusual data points. By clustering data into different groups and

observing clusters with significantly fewer data points, the Elbow Method can assist in identifying anomalies or outliers more effectively.

4. **Recommender Systems:** Recommendation engines, like those used by e-commerce platforms or content streaming services, can benefit from the Elbow Method to find the right number of user or item clusters. This, in turn, can improve the accuracy and relevance of recommendations.
5. **Genomic Data Analysis:** In genomics, researchers use the Elbow Method to identify clusters of genes or proteins for tasks like disease classification, gene expression analysis, or identifying functionally related genes.

3. Conclusion:

So, the Elbow Method is a helpful trick for data experts. It helps them figure out how many groups to put data into. It's not a rule, but it's a good way to make the groups work better and understand the data more. So, when you need to decide how many groups to use, try the Elbow Method for better results...

Machine Learning

Programming

Technology

Cluster



Written by ZalaRushirajsinh

4 followers · 1 following

Follow