

The background features a light gray grid pattern. Scattered around the grid are various hand-drawn doodles in a brownish-gold color. These include several loops and swirls at the top, a four-pointed star in the upper left, a wavy line at the bottom center, and some abstract scribbles and curved lines at the bottom and right edges.

K-Means Clustering

Unsupervised Learning
algorithm

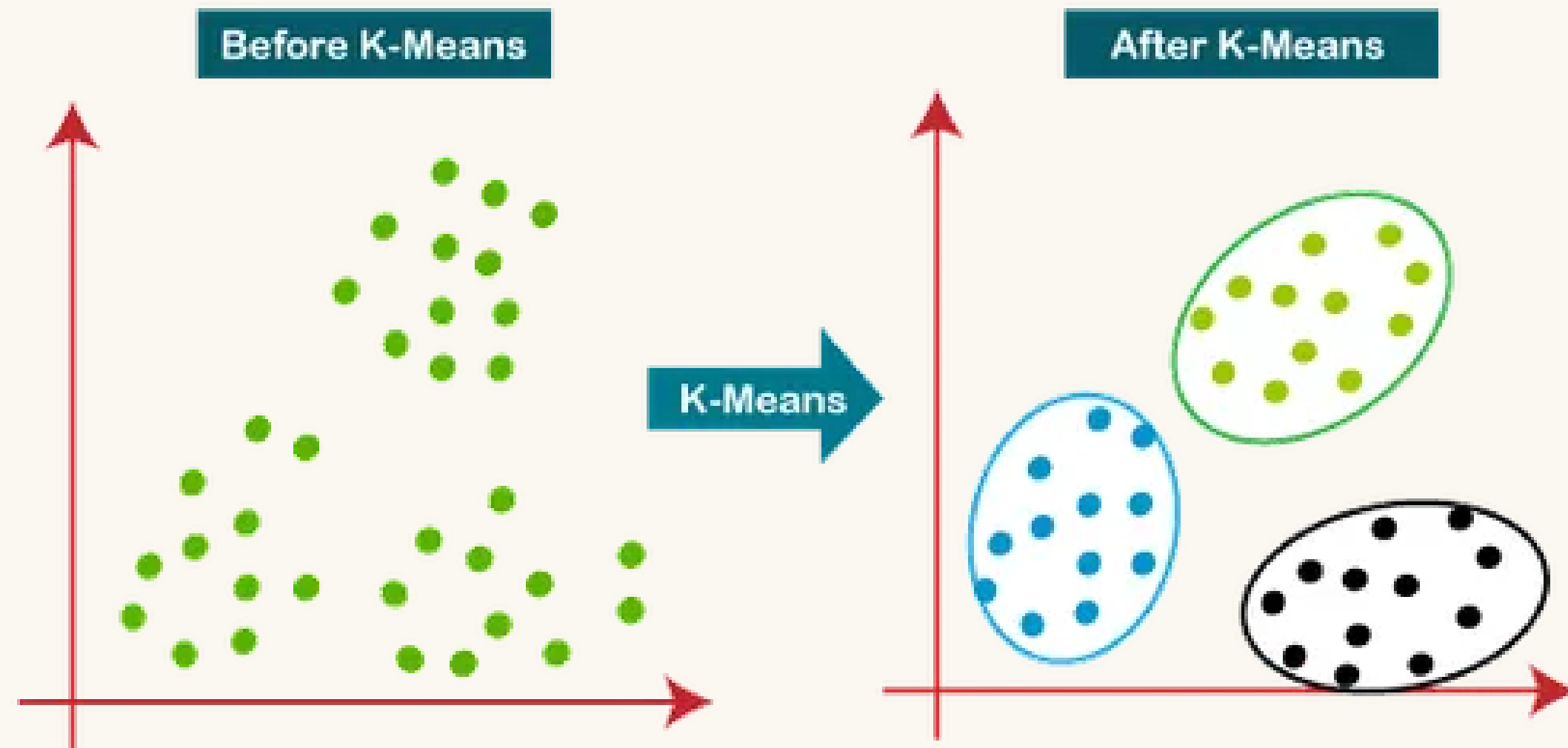
Introduction

It is used to solve the clustering problems in machine learning or data science.

It which groups the unlabeled dataset into different clusters.

Here, K defines the number of pre-defined clusters that need to be created in the process.

For example: if $K=3$, there will be three clusters



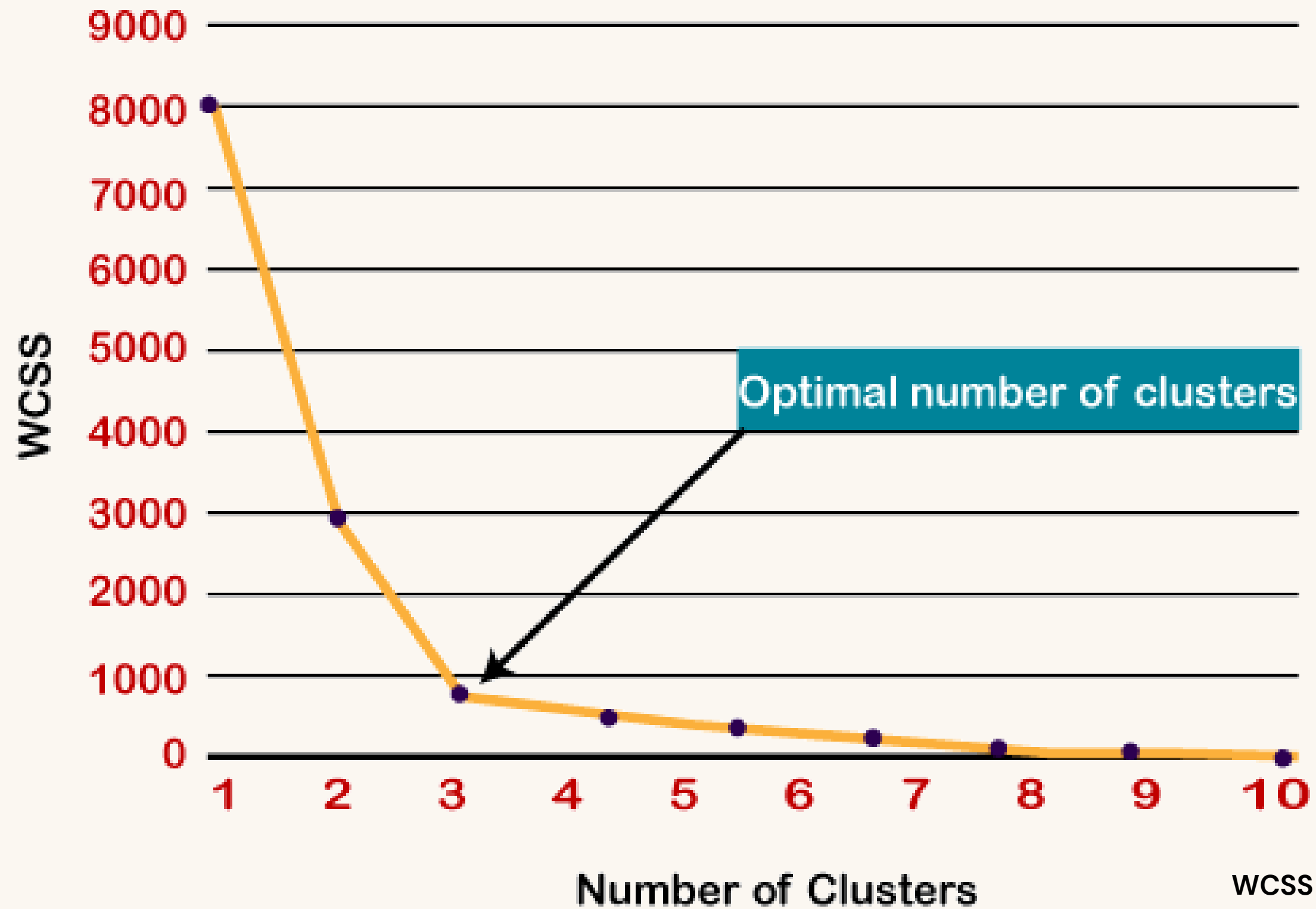
HOW IT WORKS?

1. Select the number K to decide the number of clusters.
2. Select random K points or centroids.
3. Assign each data point to their closest centroid, which will form the predefined K clusters.
4. Calculate the variance and place a new centroid of each cluster.
5. Repeat the third steps, which means reassign each datapoint to the new closest centroid of each cluster.
6. If any reassignment occurs, then go to step-4 else go to FINISH



How to choose the k value?

ELBOW METHOD



WCSS = Within Cluster Sum of Squares, which defines the total variations within a cluster.