

ChatGPT said:

## Divisive Clustering: Concept and Math

Divisive Clustering is a **top-down hierarchical clustering** algorithm. Unlike agglomerative clustering, which starts with each data point as its own cluster and progressively merges them, divisive clustering starts with all data points in a single cluster and **divides** them into smaller clusters recursively.

### Mathematical Overview of Divisive Clustering

The divisive clustering algorithm follows these basic steps:

1. **Start with one cluster:** The entire dataset is treated as a single cluster.
2. **Split the cluster:** The cluster is divided into two subclusters based on some **distance metric** (often using the **centroid** of the cluster and a distance measure, such as **Euclidean distance**).
3. **Iterate:** Each resulting subcluster is recursively divided until each data point is its own cluster (or until a stopping condition is met, like a desired number of clusters or a specific distance threshold).

### Math for Divisive Clustering

#### 1. Cluster Splitting:

- In each iteration, we pick the **most dissimilar cluster** and divide it into two subclusters. The dissimilarity can be calculated using various distance metrics, such as the **Euclidean distance** between the centroid of the cluster and the data points.
- We can use clustering algorithms (like **K-Means**) to split the most dissimilar cluster into two.

#### 2. Dissimilarity Measurement:

- The **dissimilarity** between clusters is often computed by looking at the **distance** between the centroid of the clusters or the variance within a cluster.

/////step 3 in pic

#### 4. Splitting the Cluster:

- The most dissimilar cluster is divided into two by using a method like **K-Means**. The idea is to minimize the **variance within each of the two subclusters** (similar to how K-Means optimizes centroids).
- 

## Steps for Divisive Clustering

1. **Initialize** with a single cluster containing all data points.
2. **Calculate the centroid** of the cluster.
3. **Split the cluster** into two subclusters (using a method like K-Means).
4. **Repeat** the process on the resulting subclusters, recursively splitting the most dissimilar clusters.
5. **Stop** when the desired number of clusters is reached or another stopping condition is met.