WELCOME TO

# TECH I.S.

## DBSCAN

# DBSCAN Clustering

**DBSCAN is a Density-Based Clustering algorithm**

**In density based clustering we partition points into dense regions separated by not-so-dense regions.**

↓

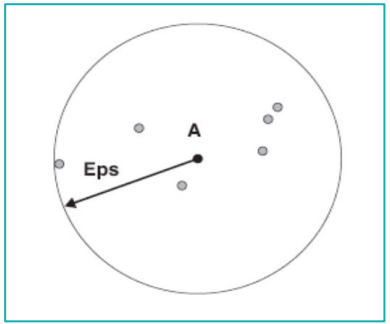
**Important Questions:**

**How do we measure density? AND What is a dense region?**

↓

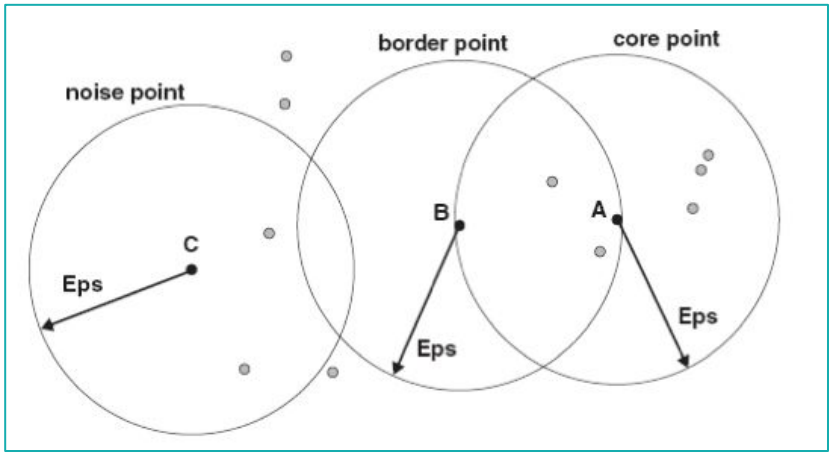
**Density at point p: number of points within a circle of radius Eps.**

**Dense Region: A circle of radius Eps that contains at least MinPts points**



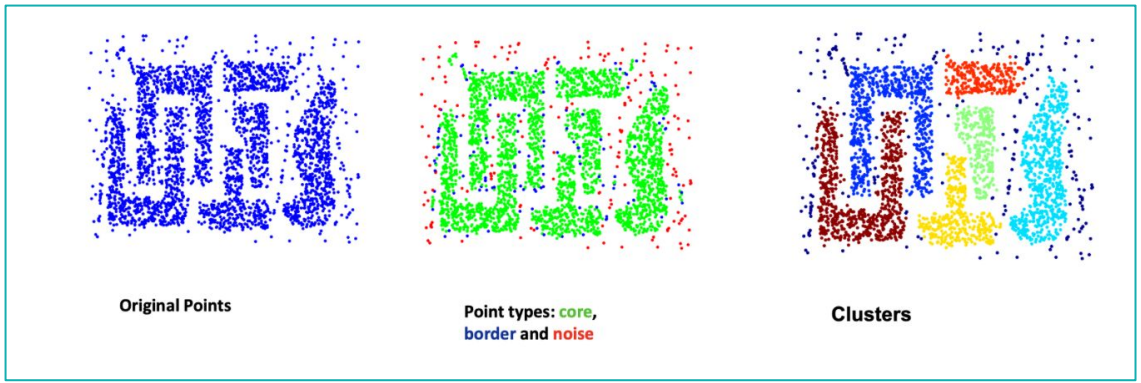
# Characterization of point

- A point is a core point if it has more than a specified number of points (MinPts) within Eps
  - These points belong in a dense region and are at the interior of a cluster
- A border point has fewer than MinPts within Eps, but is in the neighborhood of a core point
- A noise point is any point that is not a core point or a border point.



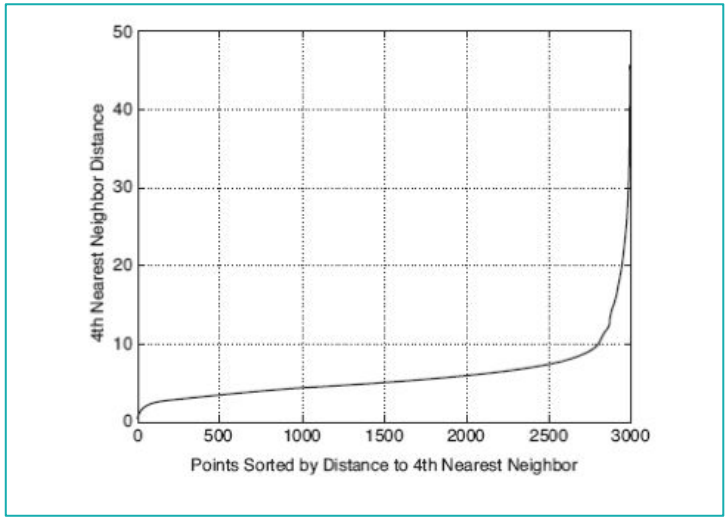
# DBSCAN Algorithm

1. Label points as core, border and noise
2. Eliminate noise points
3. For every core point p that has not been assigned to a cluster
    - Create a new cluster with the point p and all the points that are density-connected to p.
4. Assign border points to the cluster of the closest core point.



# Determining EPS and MinPts

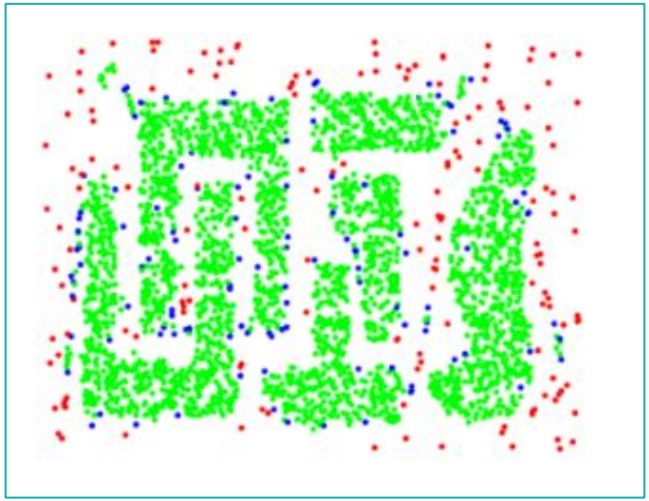
- Idea is that for points in a cluster, their kth nearest neighbors are at roughly the same distance
  - Noise points have the kth nearest neighbor at farther distance
  - So, plot sorted distance of every point to its kth nearest neighbor



# Strengths of DBSCAN Clustering

**DBSCAN works well with:**

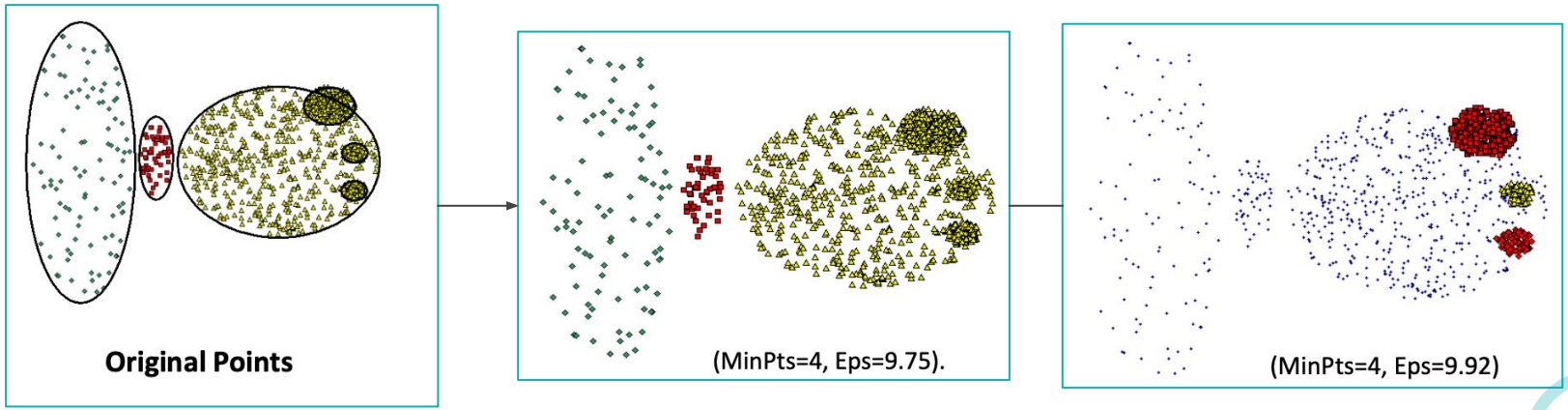
- **Resistant to Noise**
- **Can handle clusters of different shapes and sizes**



# Limitations of DBSCAN Clustering

**DBSCAN Struggles with:**

- Data with Varying densities.
- High-dimensional data



# Much obliged.

TECH I.S.