

## #DBSCAN - Density-Based Clustering

dbscan groups data based on point density instead of distance to a center. for a point  $x$ , its  $\epsilon$ -neighborhood is:

$$N_{\epsilon}(x) = \{y \mid \|y - x\| \leq \epsilon\}$$

a point is a core point if:

$$|N_{\epsilon}(x)| \geq \text{minPts}$$

a point is density-reachable from a core point if it lies inside its  $\epsilon$ -neighborhood. two points are density-connected if there exists a chain of density-reachable points between them.

a cluster is a maximal set of density-connected points. points that are not density-connected to any cluster are labeled as noise.

this formulation allows dbscan to discover arbitrarily shaped clusters and naturally ignore outliers.