

Density-Based Clustering:

- **Goal:** cluster together points that are densely packed together.
 - How should we define density?
 - Given a fixed radius ϵ around a point, if there are at least **min_pts** number of points in that area, then this section is dense.
- We need to distinguish between points at the core of a dense region and points at the border of a dense region.
 - Let's define:
 - **Core** point: if its ϵ -neighborhood contains at least **min_pts**
 - **Border** point: if it is in the ϵ -neighborhood of a core point
 - **Noise** point: if it is neither a core nor border point
- DBScan Algorithm:
 - ϵ and **min_pts** given:
 - 1. Find the ϵ -neighborhood of each point
 - 2. Label the point as **core** if it contains at least **min_pts**
 - 3. Label points in its neighborhood that are not **core** as **border**
 - 4. Label points as **noise** if they are neither **core** nor **border**
 - 5. For each **core** point, assign to the same cluster all **core** points in its neighborhood
 - 6. Assign border points to nearby clusters
 - Benefits:
 - 1. Can identify clusters of different shapes and sizes
 - 2. Resistant to noise
 - Limitations:
 - 1. Can fail to identify clusters of varying densities.
 - 2. Tends to create clusters of the same density.
 - 3. Notion of density is problematic in high-dimensional spaces