

Hi Team,

Please go through this document on DBSCAN. As you read, take a moment to think through the concepts and check if they make sense to you. If anything feels unclear, we can discuss it together. The content is divided into three parts — A, B, and C. Try to ensure you understand each of them.

A.

What DBSCAN Does

Instead of guessing how many groups there are, DBSCAN says:

“I will walk around, and if I see enough dots near me, I’ll start a group right here!”

eps (ϵ): Maximum distance between two points to be considered as neighbors.

min_samples/min_points: Minimum number of points required to form a dense region (i.e., a core point).

Core Point: A point with at least `min_samples` within distance `eps`.

Border Point: Has fewer than `min_samples` within `eps`, but is in the neighborhood of a core point.

Noise Point (Outlier): Neither a core point nor a border point.

B.

For each point, count how many points are within distance **eps**.

If this count \geq **min_samples**, it's a **core point**.

Core points form the **center of a cluster**. Expand the cluster by including all directly density-reachable points.

Border points are included in a cluster but don't expand it.

Points not reachable from any core point are marked as **noise**.

C.

Core Points as Cluster Centers

In DBSCAN:

- A **core point** is a point that has **at least min_samples neighbors** (including itself) **within a distance eps**.
- These core points are considered the **"heart" or center** of a potential cluster.

Cluster Expansion: Directly Density-Reachable

Now, what does *"expand the cluster by including all directly density-reachable points"* mean?

♦ "Directly density-reachable" = close & dense

- A point **A** is directly density-reachable from point **B** if:
 - **B is a core point**, and
 - **A is within eps distance from B**.

So, from a core point, DBSCAN **looks around in its `eps` neighborhood** and pulls all nearby points into the same cluster — whether they are:

- other **core points** (which in turn expand the cluster even further), or
 - **border points** (which do not expand the cluster themselves but are included).
-

Think of it like this:

1. Imagine placing a **circle of radius `eps`** around a point.
2. If that circle contains enough other points (\geq `min_samples`), it's a **core point**.
3. Then DBSCAN **grows** a cluster from this point by:
 - Adding all other **points within this circle**.
 - If those new points are also **core points**, it repeats the process from them.
4. It **keeps expanding** until no new points are reachable.