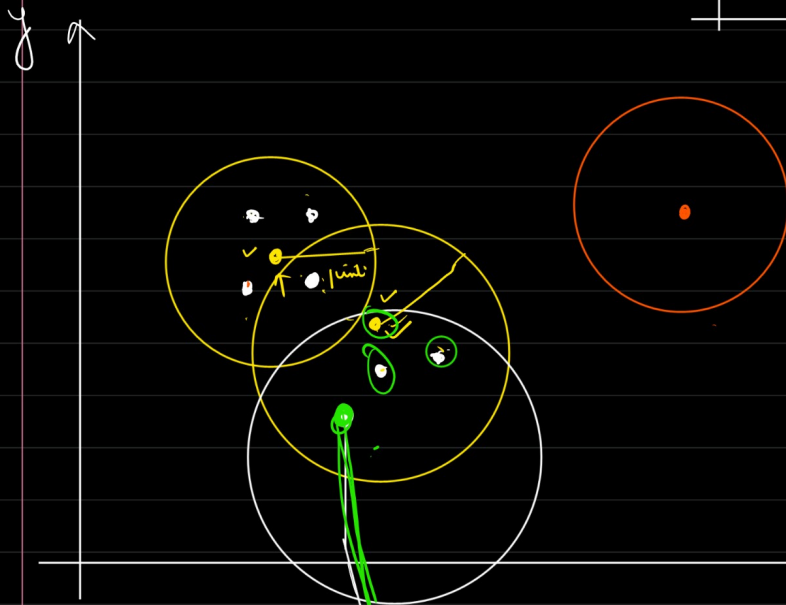


DBSCAN (density based spatial clustering of application with noise)

Distance based Algorithm
such as K means & hierarchical can not be used here (Non Linear cluster)



- → Core datapoint
- → border datapoint
- Outlier

* min no. of dp = 4 ✓
* radius (ϵ) = 1 ✓

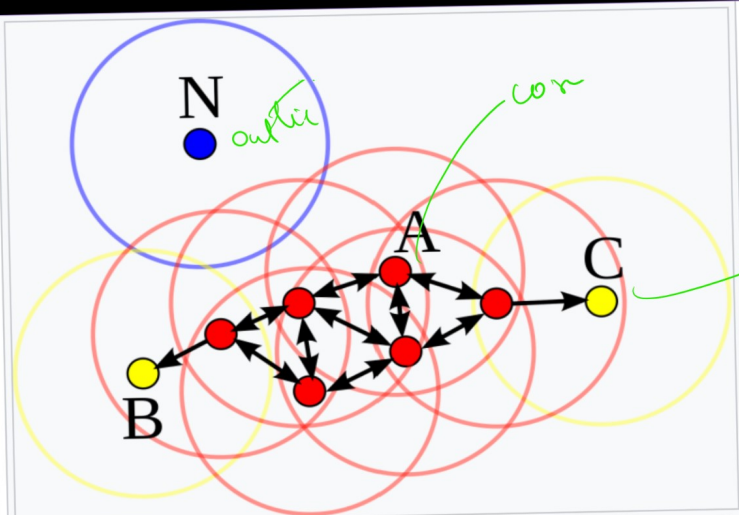
Core → the dp have atleast
✓ minimum no of dp (4)
✓ in its radius ($\epsilon = 1$ unit)

border → No. of dp within
its boundary (ϵ) will
be less than minimum
dp.

Why DBSCAN ?

- Find Patterns in non-linear data
- Outlier

* Outlier - No dp in
radius (ϵ)



In this diagram, $\text{minPts} = 4$. Point A and the other red points are core points, because the area surrounding these points in an ϵ radius contain at least 4 points (including the point itself). Because they are all reachable from one another, they form a single cluster. Points B and C are not core points, but are reachable from A (via other core points) and thus belong to the cluster as well. Point N is a noise point that is neither a core point nor directly-reachable.

Border

ϵ
minNo of dpts

