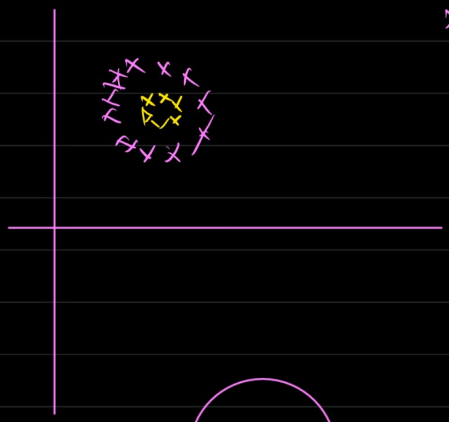
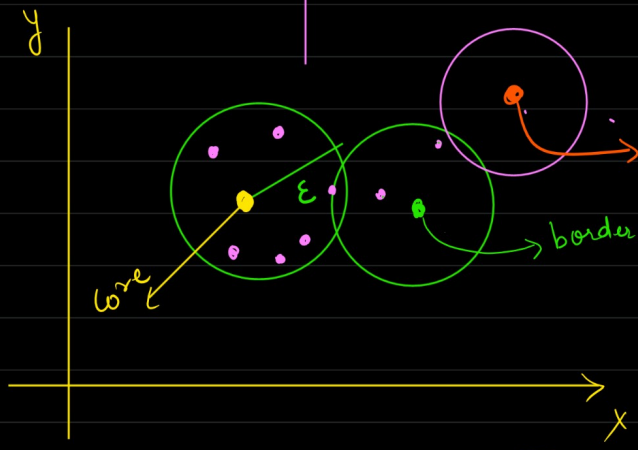
DBSCAN (density based Spatial Clustering of Application with noise)

* Characteristics of DBSCAN
  - → Finds groups / pattern
  - → Finds outliers / noise

if already we had Kmean / hierarchical, then why DBSCAN?



* Distance based Algorithm such as K means & hierarchical can not be used here (non linear data / non linear clustering)



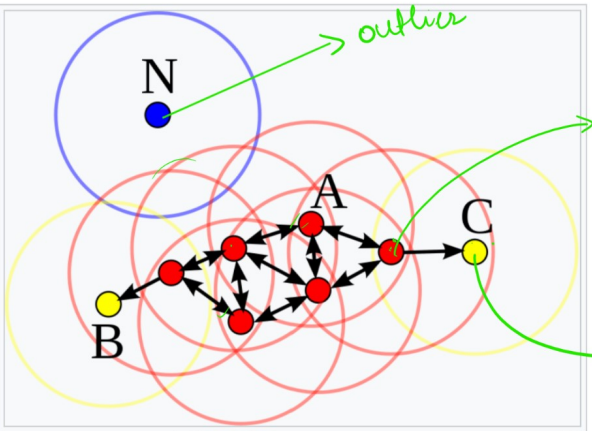
- ● → Core data point
- ● → border data point
- ● → outlier dp.

* minimum no of dp = 4
* radius (ε) = 1 m

* Core dp :-
  The dp's which hv atleast minimum no. of dp in its radius ε (4)

* Border dp → no. of dp within its boundary (ε) will be less than minimum dp

* Outlier (Noise) → No dp in radius ε



outlier

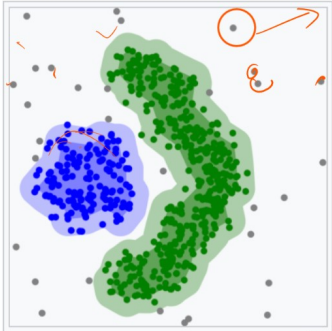
core dp

border dp

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.

* Some Examples of DBSCAN working very well with non linear data.

outlier



DBSCAN can find non-linearly separable clusters. This dataset cannot be adequately clustered with k-means or Gaussian Mixture EM clustering.

The left image depicts a more traditional clustering method that does not account for multi-dimensionality. Whereas the right image shows how DBSCAN can contort the data into different shapes and dimensions in order to find similar clusters.