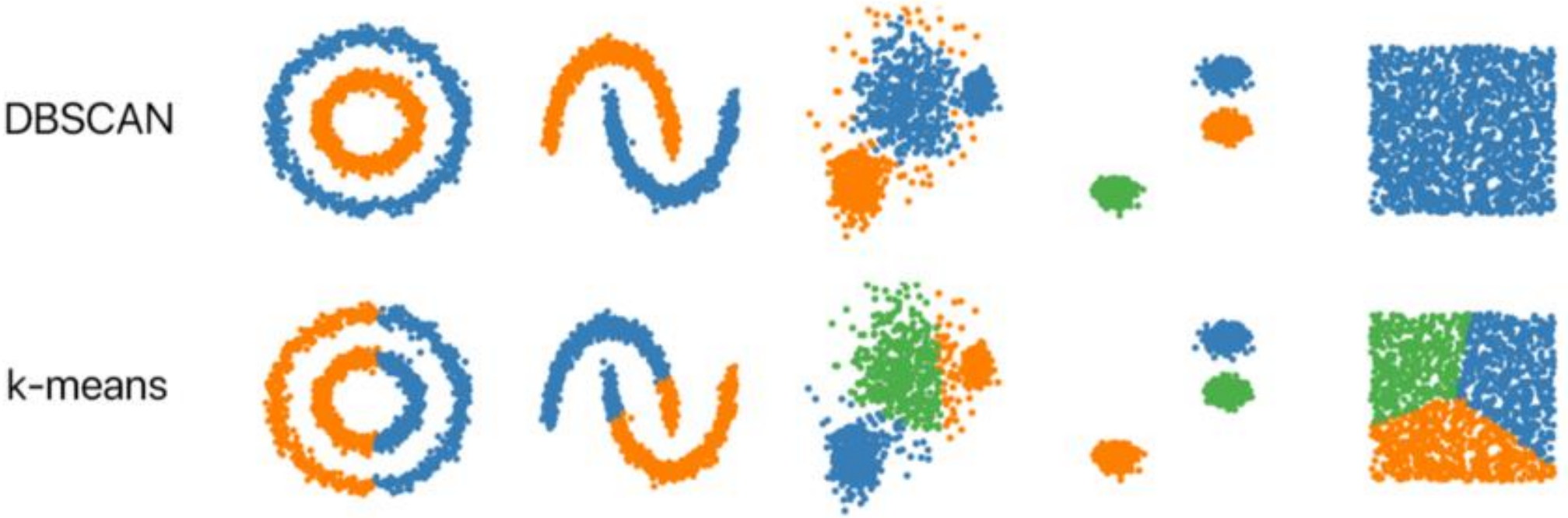
DBSCAN(Density-Based Spatial Clustering of Applications with Noise)
DBSCAN does not need to specify the number of clusters ;
can automatically detect the number of clusters based
DBSCAN can find arbitrary shape clusters that k-means are not able to find
DBSCAN can handle noise and outliers better



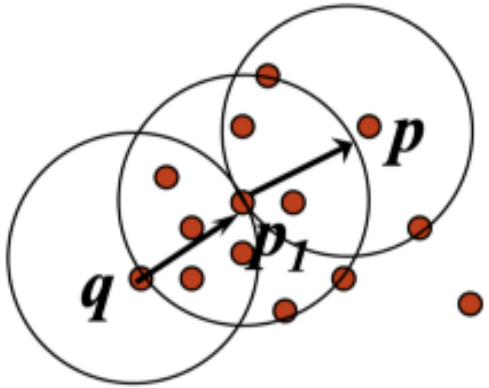
Some parameters used by DBSCAN

Eps: Maximum radius of the neighborhood

MinPts: Minimum number of points in an Eps-neighbourhood of that point

Directly density-reachable: A point p is directly density reachable from a point q w.r.t. Eps, MinPts, if NEps (q): {p belongs to D | dist(p,q) ≤ Eps} and |N Eps (q)| ≥ MinPts ; Minpts = 5, Eps = 1

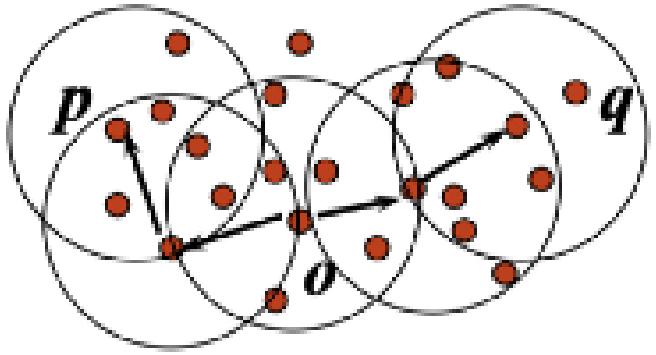
- Density-reachable:
  - A point *p* is density-reachable from a point *q* w.r.t. *Eps*, *MinPts* if there is a chain of points $p_1, \ldots, p_n, p_1 = q, p_n = p$ such that p_{i+1} is directly density-reachable from p_i



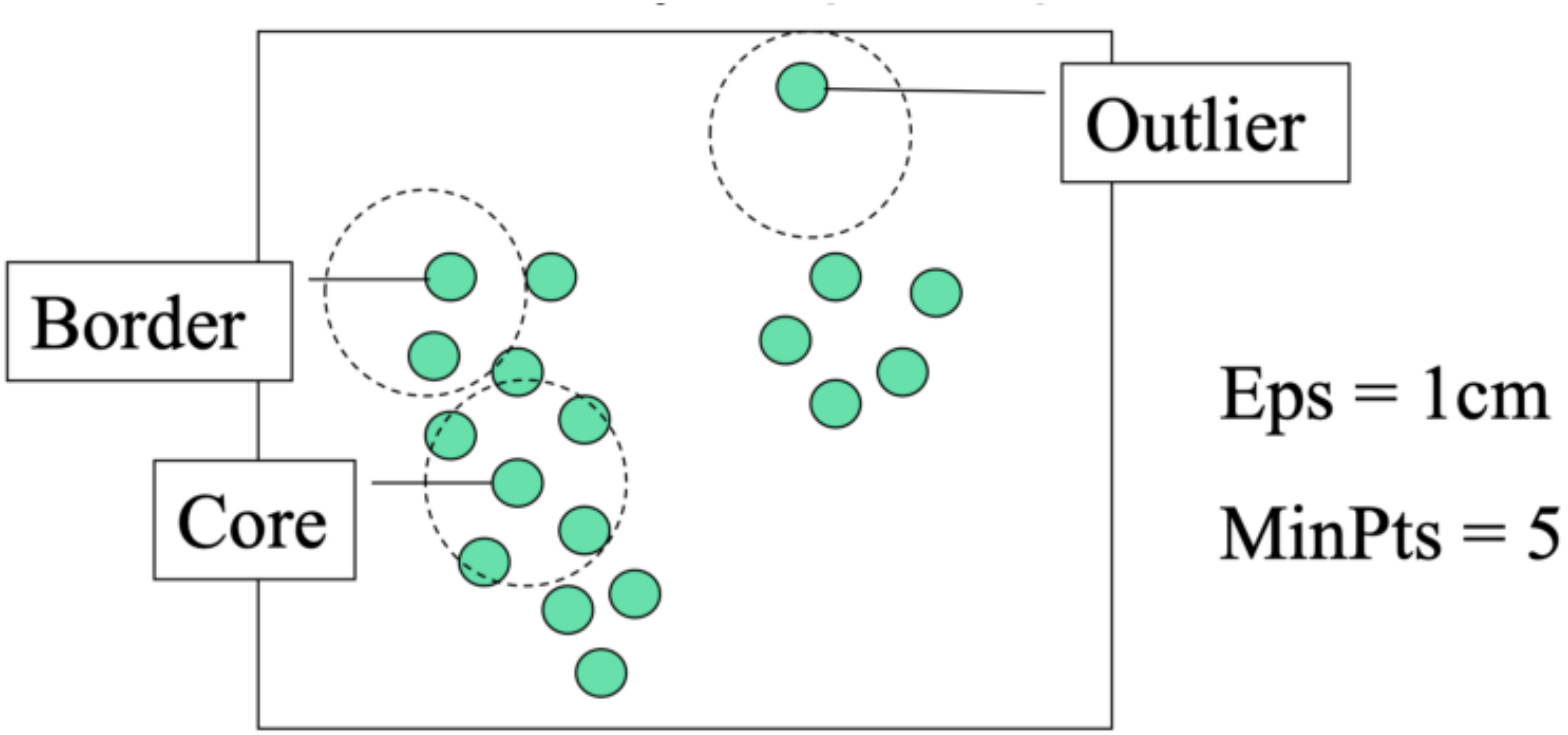
p belongs to D | dist(p,q) ≤ Eps} and |N Eps (q)| ≥ MinPts ;

Minpts = 5, Eps = 1

- Density-connected
  - A point *p* is density-connected to a point *q* w.r.t. *Eps*, *MinPts* if there is a point *o* such that both, *p* and *q* are density-reachable from *o* w.r.t. *Eps* and *MinPts*



- core point if it has more than a specified number of points (MinPts) within Eps. These are points that are at the interior of a cluster A.
  border point has fewer than MinPts within Eps, but is in the neighborhood of a core point.
  outlier(noise) point, which is the points that are neither core nor border points.

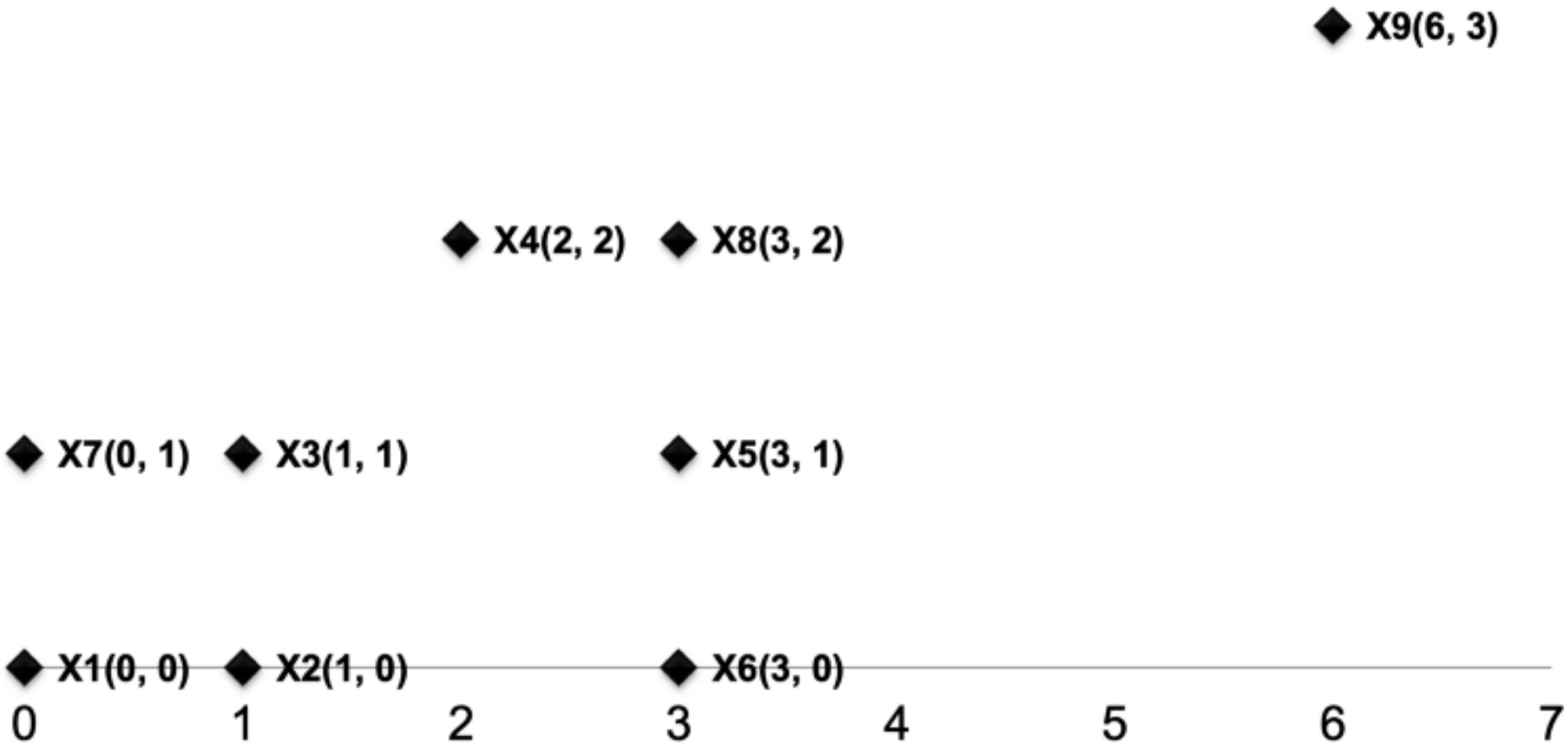


1. 1. Arbitrary select a point p
   2. Retrieve all points density-reachable from p based on Eps and MinPts
   3. If p is a core point, a cluster is formed
   4. If p is a border point, no points are density-reachable from p and DBSCAN visits the next point of the database
   5. Continue the process until all of the points have been processed

   Consider the following data to be clustered;

   x1(0,0), x2(1,0), x3(1,1), x4(2,2), x5(3,1), x6(3,0), x7(0,1), x8(3,2), x9(6,3)
   Eps =1 and MinPts = 3. Find all core points, border points and noise points, and show the final clusters using DBCSAN algorithm

Data Points



N(p), Eps-neighborhood of point p

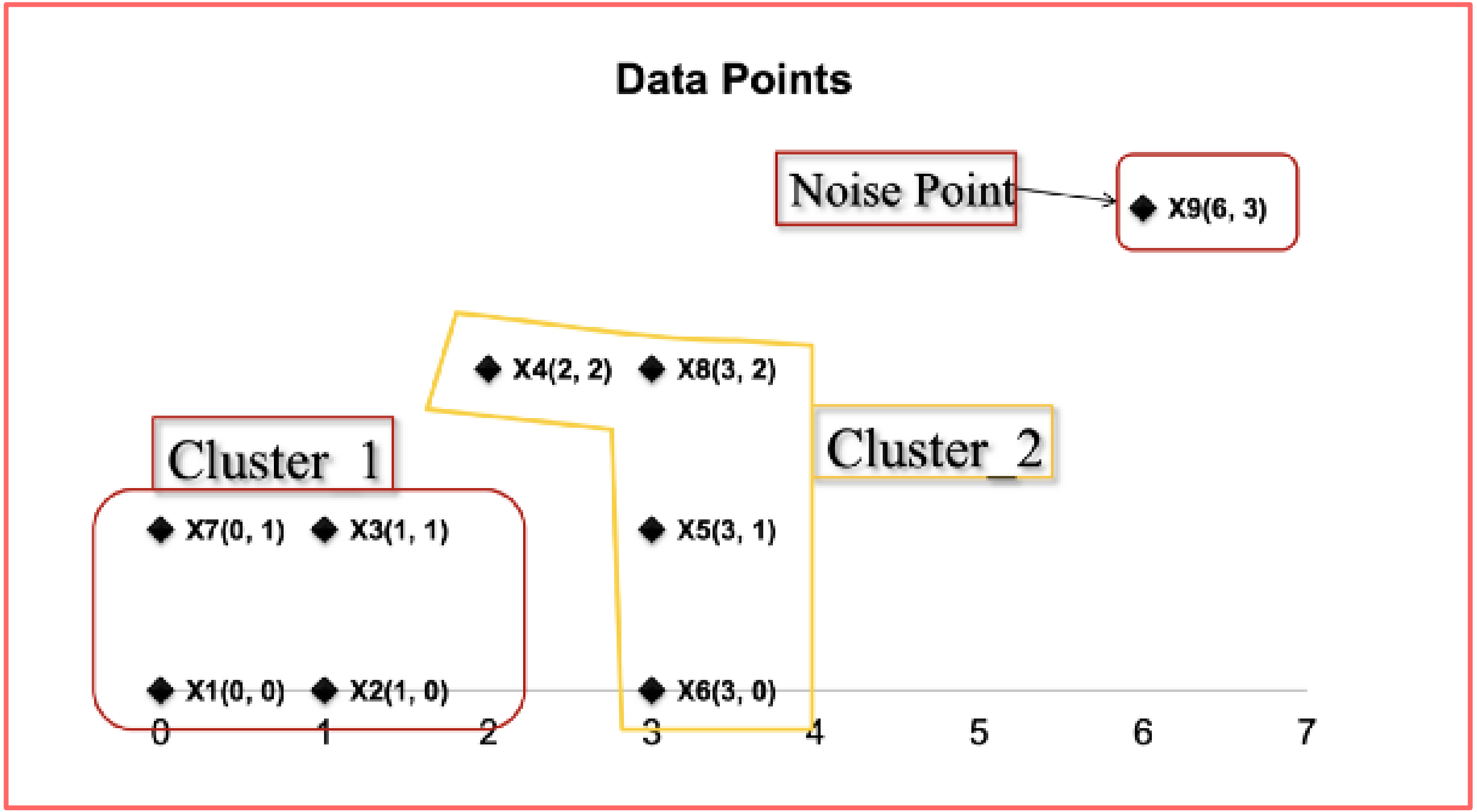
N(x1) = {x1, x2, x7}
N(x2) = {x2, x1, x3}
N(x3) = {x3, x2, x7}
N(x4) = {x4, x8}
N(x5) = {x5, x6, x8}
N(x6) = {x6, x5}
N(x7) = {x7, x1, x3}
N(x8) = {x8, x4, x5}
N(x9) = {x9}

If the size of N(p) is at least MinPts, then p is said to be a core point.

MinPts is 3, thus the size of N(p) is at least 3. **Thus core points are:{x1, x2, x3, x5, x7, x8}**

given a point p, p is said to be a border point if it is not a core point but N(p) contains at least one core point. N(x4) = {x4, x8}, N(x6) = {x6, x5}. here x8 and x5 are core points, So both **x4 and x6 are border points**.

Arbitrary select a point p, now we choose x1
Retrieve all points density-reachable from x1: {x2, x3, x7}
Here x1 is a core point, a cluster is formed. So we have Cluster_1: {x1, x2, x3, x7}
Next, we choose x5, Retrieve all points density-reachable from x5: {x8, x4, x6}
Here x5 is a core point, a cluster is formed.So we have Cluster_2: {x5, x4, x8, x6}
Next, we choose x9, x9 is a noise point, noise points do NOT belong to any clusters.


Data Points
Noise Point
X9(6, 3)
X4(2, 2)
X8(3, 2)
Cluster_1
Cluster_2
X7(0, 1)
X3(1, 1)
X5(3, 1)
X1(0, 0)
X2(1, 0)
X6(3, 0)
0
1
2
3
4
5
6
7