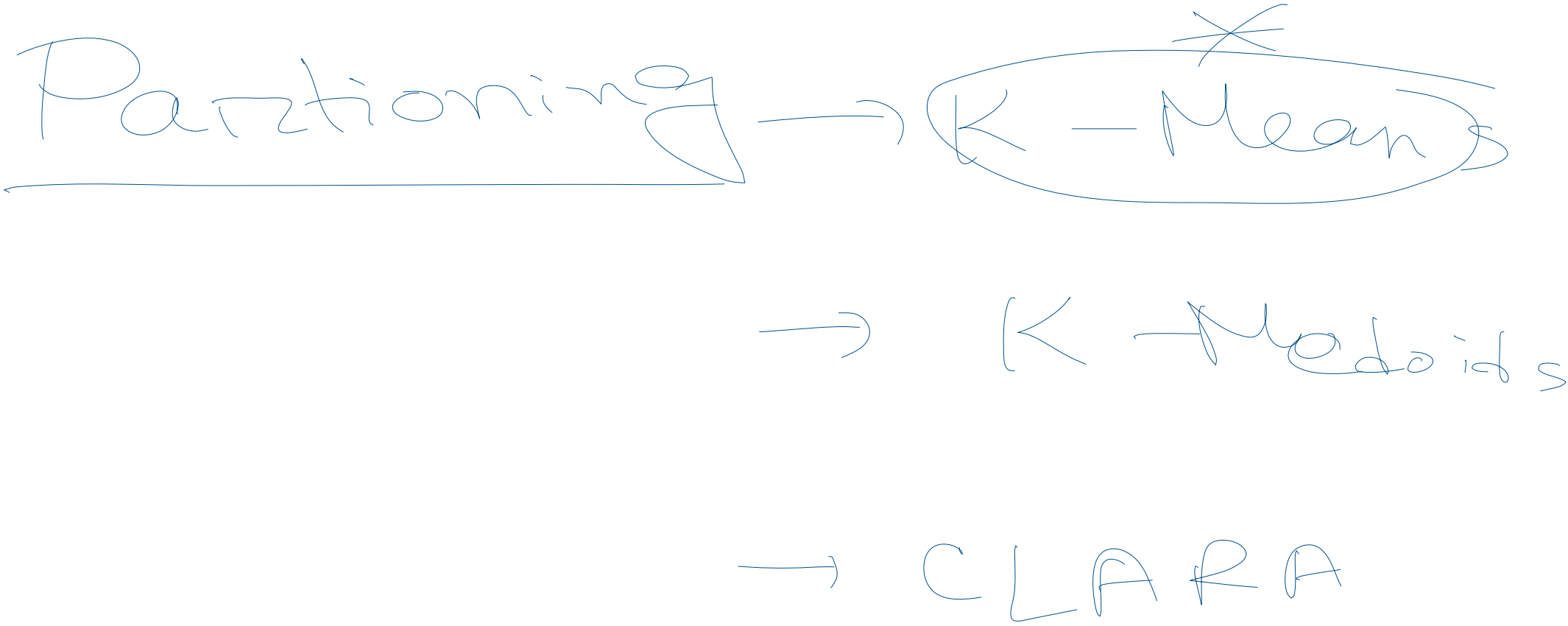
Clustering: Partitioning Methods, Hierarchical Methods, Density-Based Methods, Grid-Based Methods.



Properties

1. Scalable
2. High dimensionality
3. Diff data
4. Unstructured data

Partioning → K-Means

16, 16, 17, 20, 21, 21, 22, 23

K = 2

Centroid (C_1) = 16

Centroid (C_2) = 21

$S_1 = [15, 16, 17] \rightarrow 16.33$

$S_2 = [20, 21, 21, 22, 23] \rightarrow 21.$

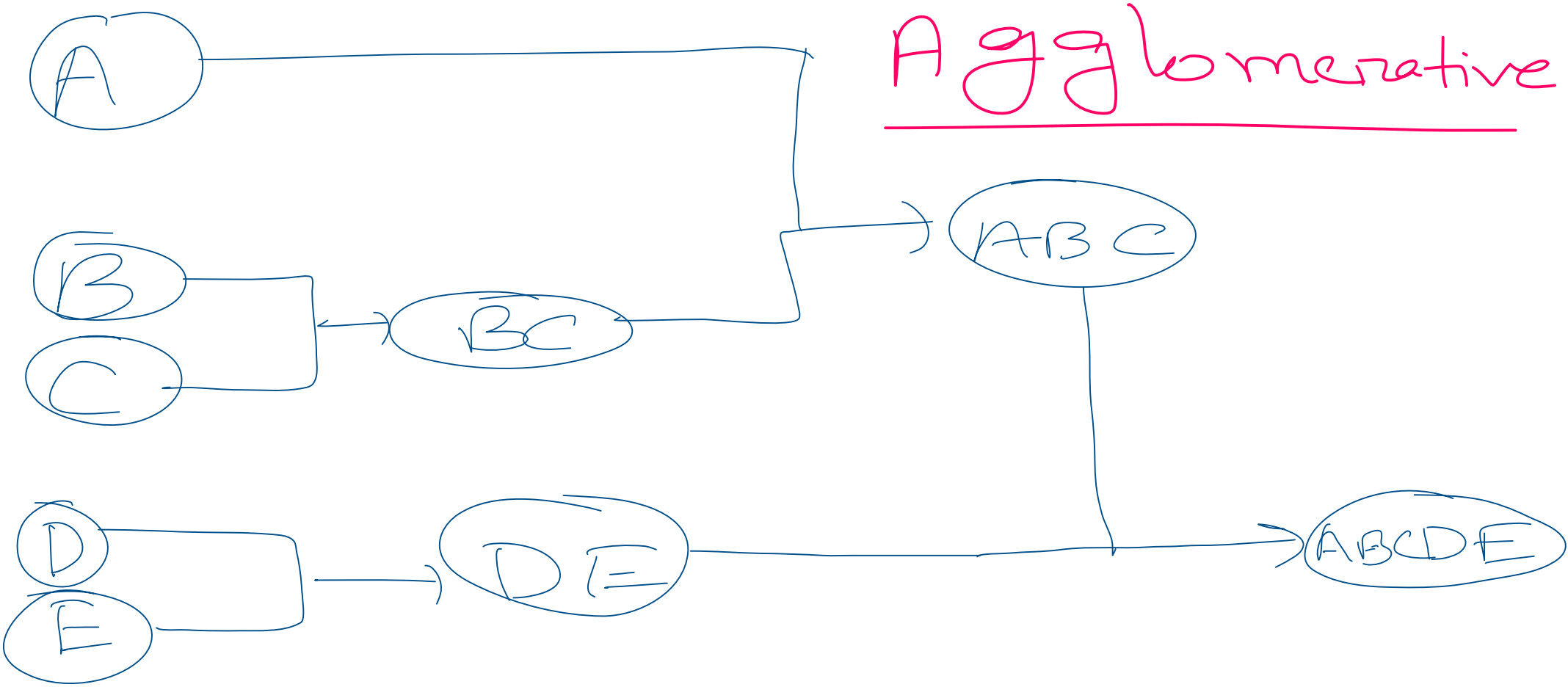
STOP

7

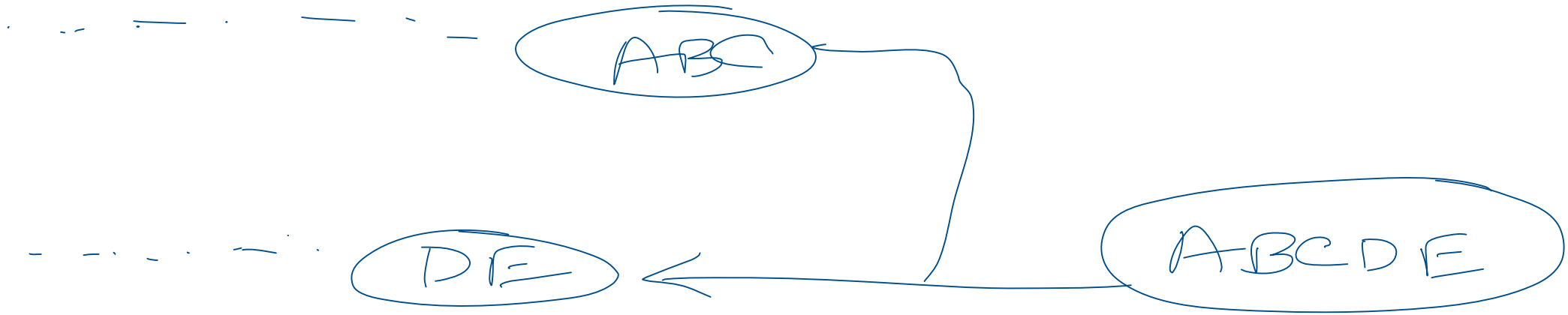
Hierarchical

- Tree of clusters
- Bottom-up

Agglomerative



Divisive



AdU

1. Diff size, density
2. Missing & noisy
3. Relationship

Disadv

① Computational cost ↑↑

② Criterion → #clusters
stopping

DBSCAN

1000 Queries

1

- large diff in density

Density-based spatial Clustering of Applications with Noise

DBSCAN

① eps ⟶ neighborhood

dist, d[A, B] ≤ eps

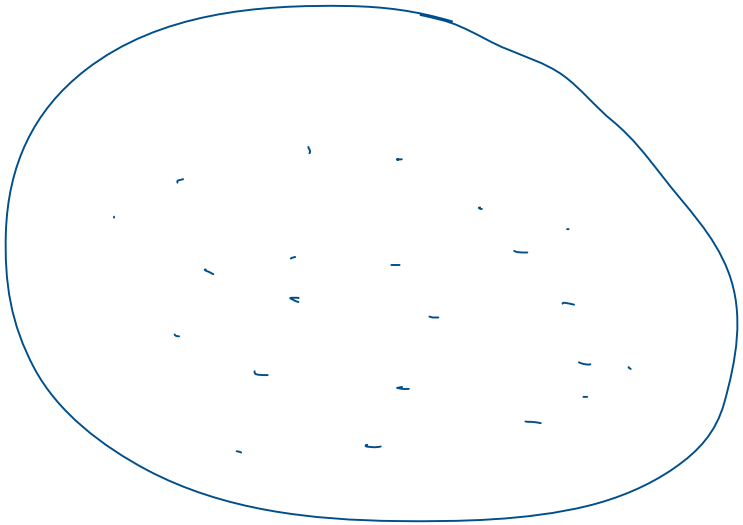
neighbor

— k - distance graph

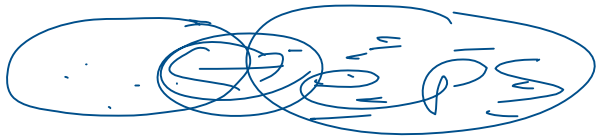
eps = 2

2. MinPts → min number of datapoints (neighbor)

≥ 3



- dataset ↑ MinPts ↑

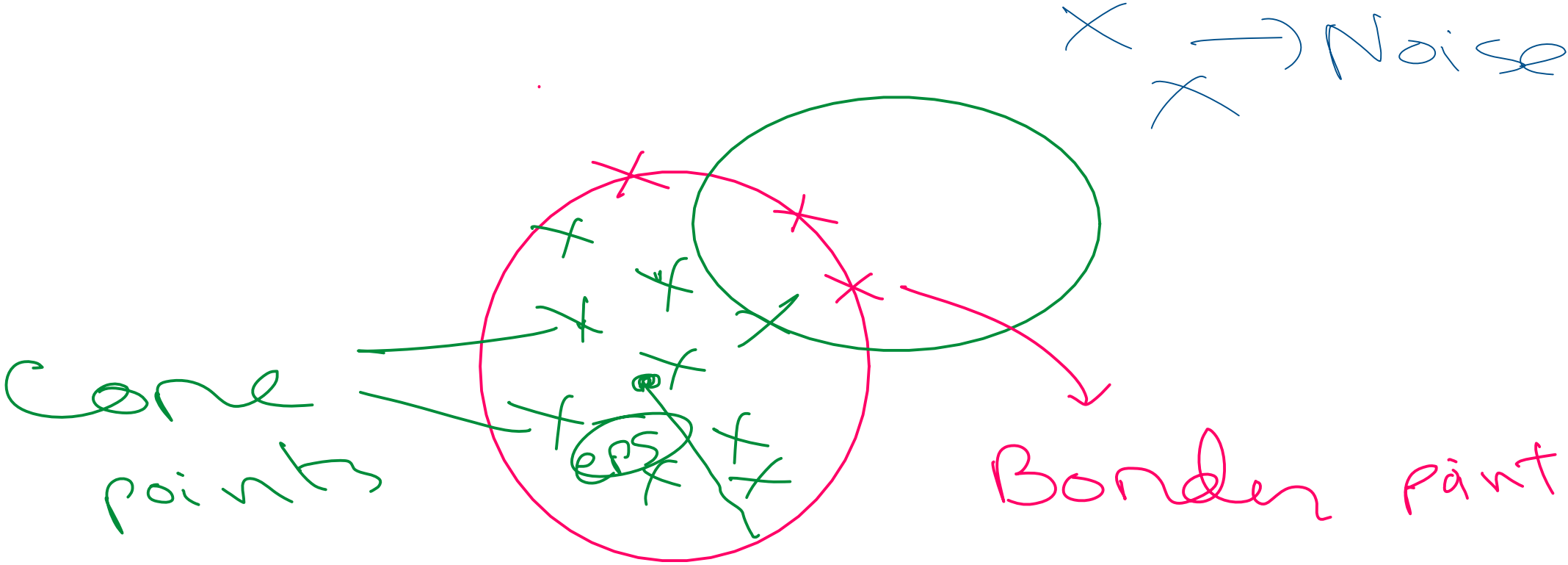


3 - types of data - points

① Core → > Minpoints

② Borden → < Minpts & in the neigh of core point

③ Noise/Outlier → Neither Core / Borden


Noise
Core points
eps
Borden point

Steps

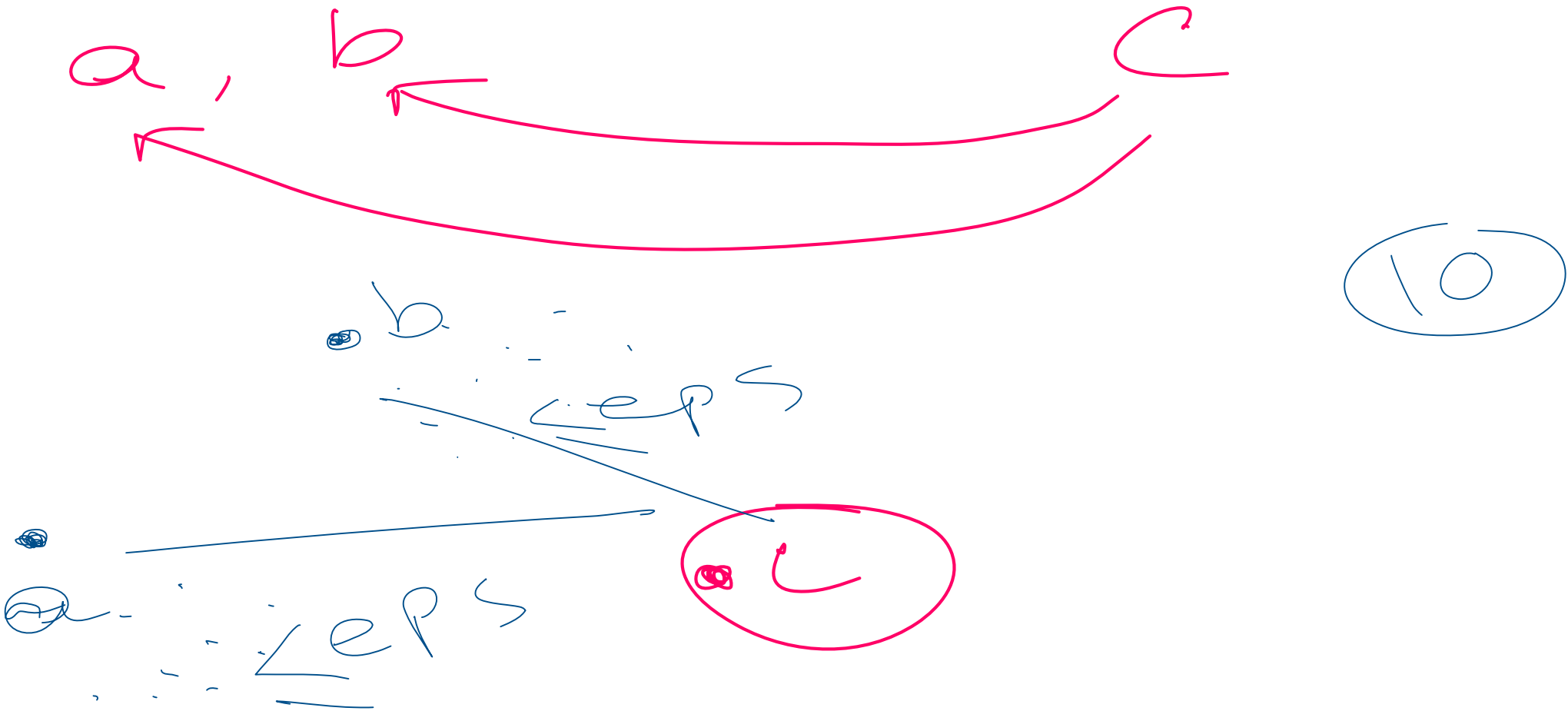
① Find all neighbor points within eps and identify core pts

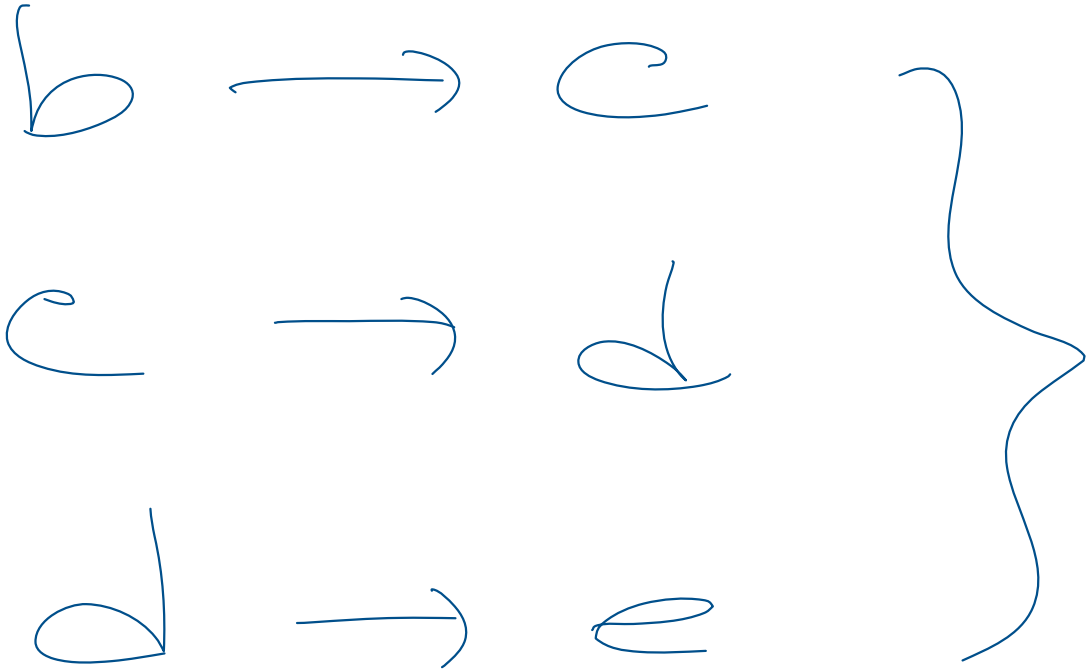
② For each core pts, if not assigned to a cluster → new cluster

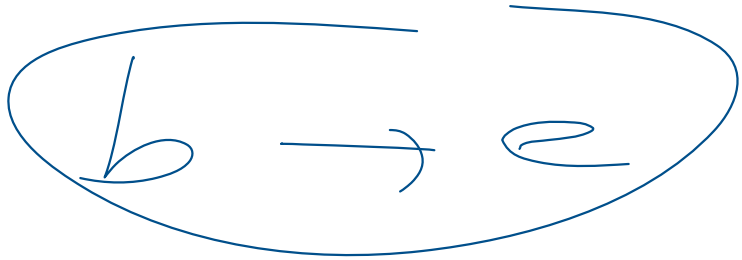
③ Recursively find all density-connected points. Assign them to same cluster as the

Assign them to same cluster as the core point.

Density - connected points



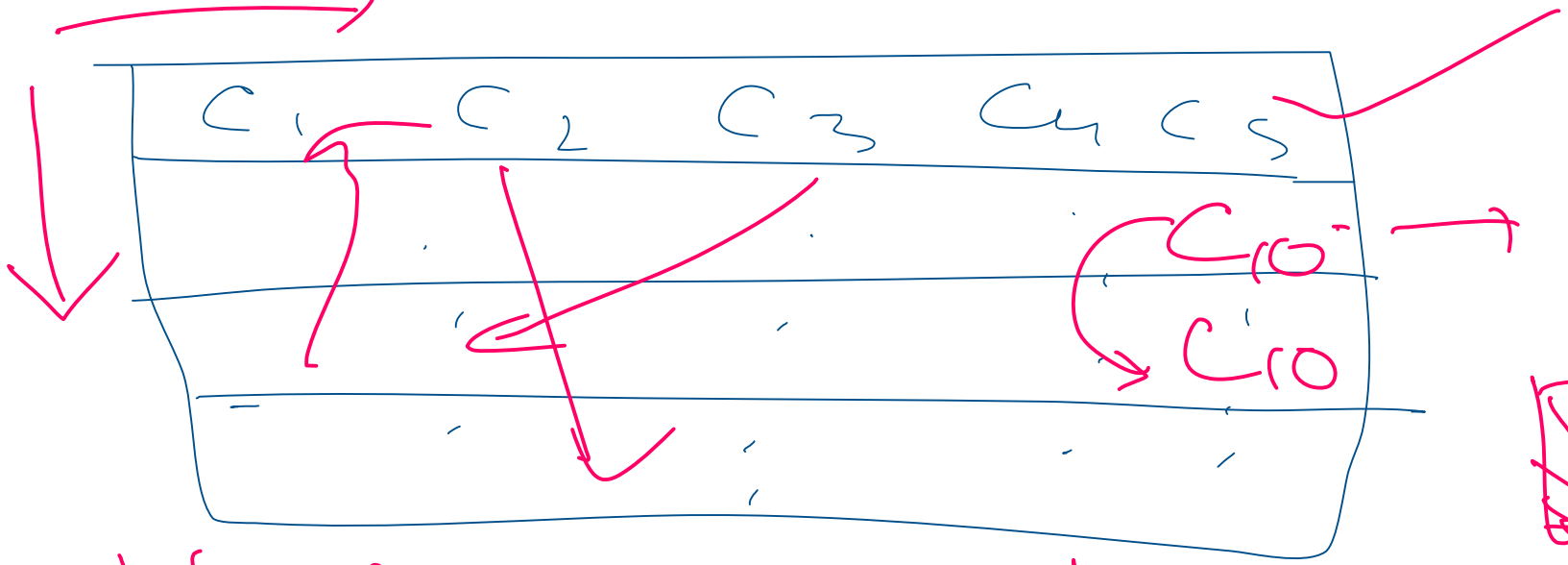




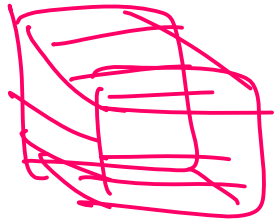
Grid - based → STING

Statistical Info.

- hierarchical



No diag - bound



OPTICS → Density based

Ordering Point to

Identify clustering Struct

Saturday, July 8, 2023 10:24 PM