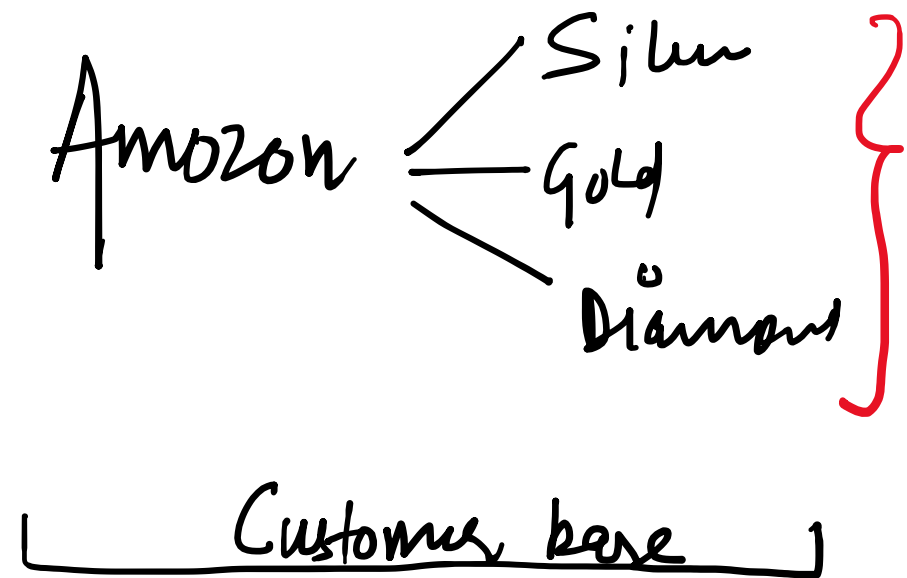


→ Clustering :-

* Segmentation of Customers



↳

① Distribute randomly

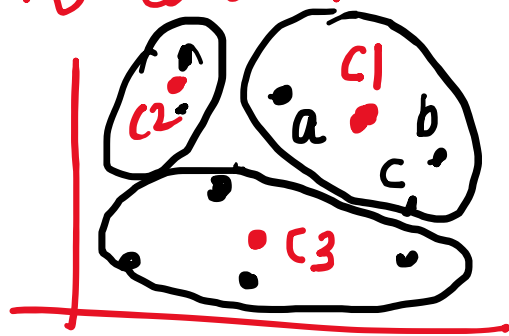
↳

② Segment the customers basis their engagement & then distribute the coupons

→ K-means Clustering :-

* Pre-requisite :- No. of clusters to be known before hand. ($K=3$)

① → Random Initialization : Randomly K centroids are initialised



$$c1 = \frac{a+b+c}{3}$$

→ Assignment :- Euclidean distance

→ Update :- Centroids will get updated with mean of datapoints assigned to it

→ Repetition :-

(x, y)

•

A

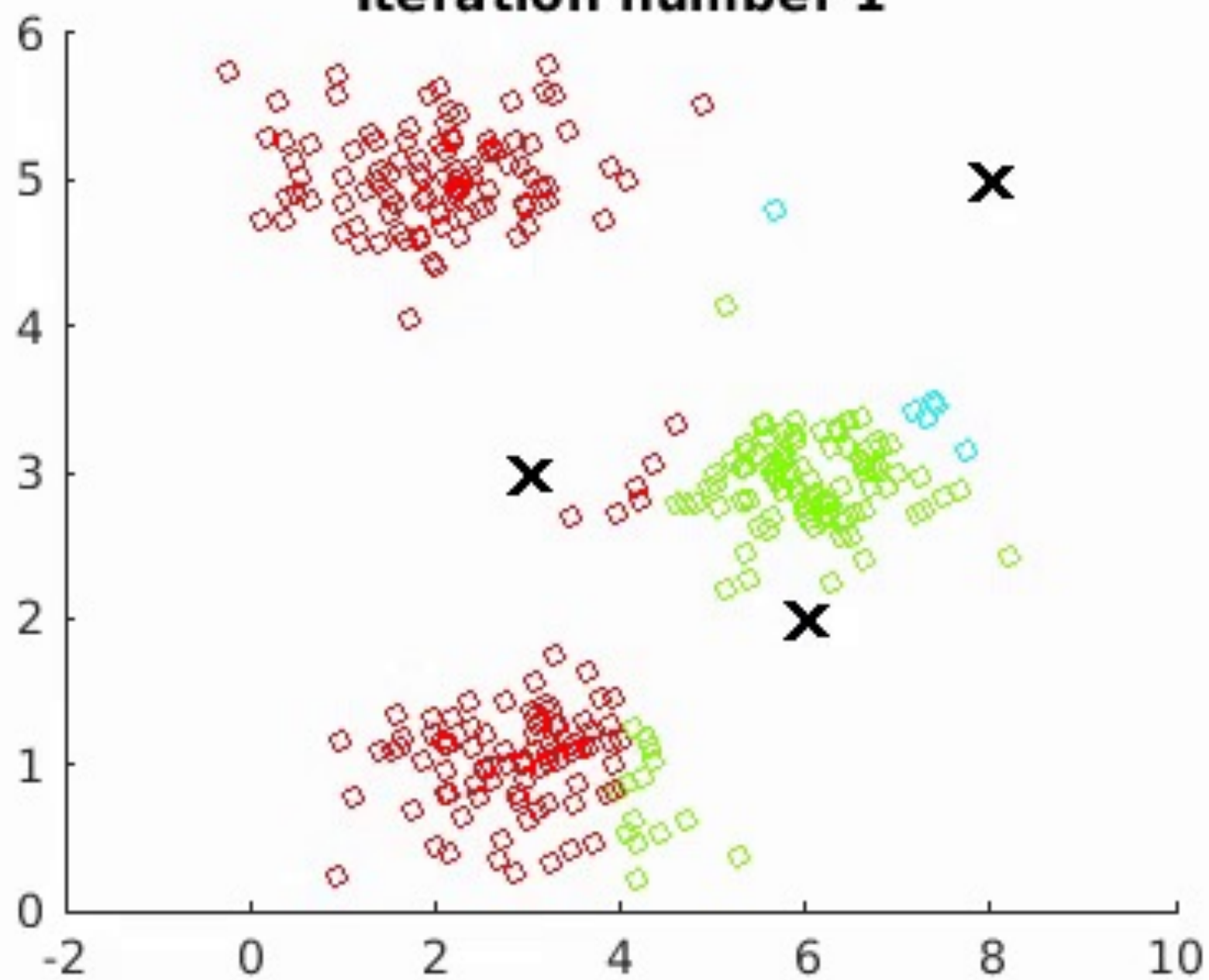
(c, d)

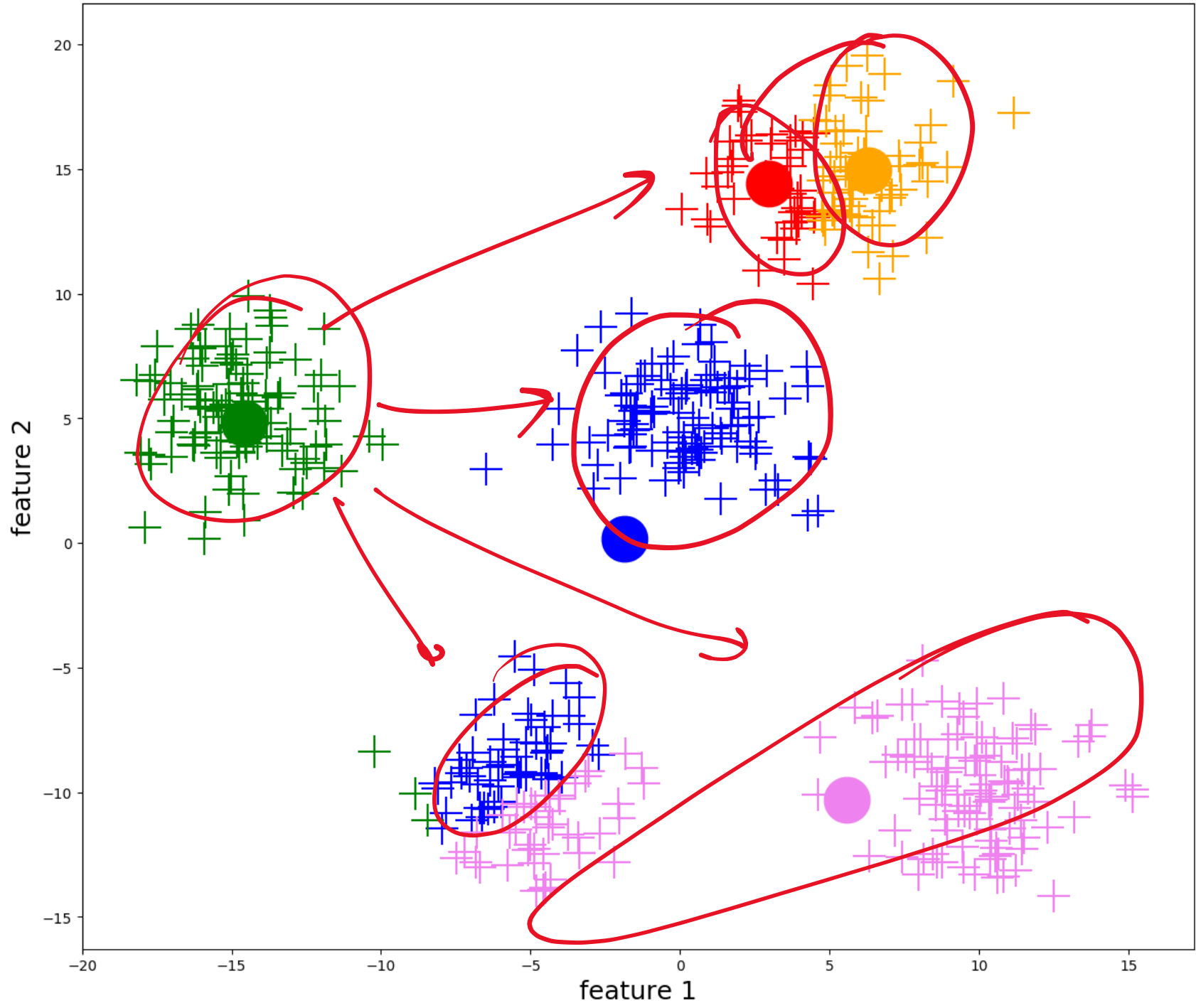
•

C

$$d = \sqrt{(x - c)^2 + (y - d)^2}$$

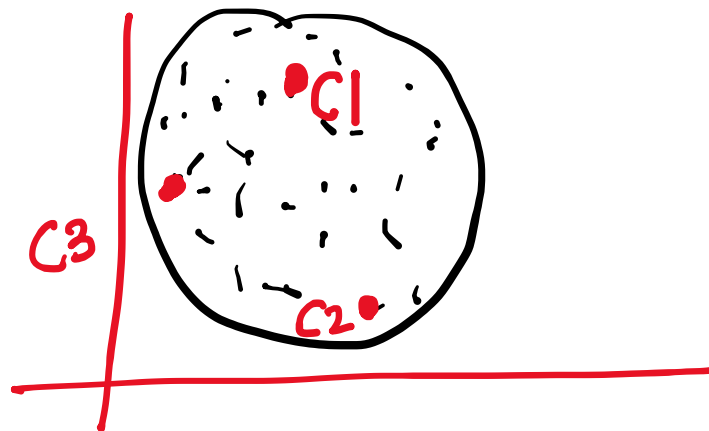
Iteration number 1





Inter cluster ↓
Intra cluster ↑

→ K-Means ++ :-



① R α

② A ✓

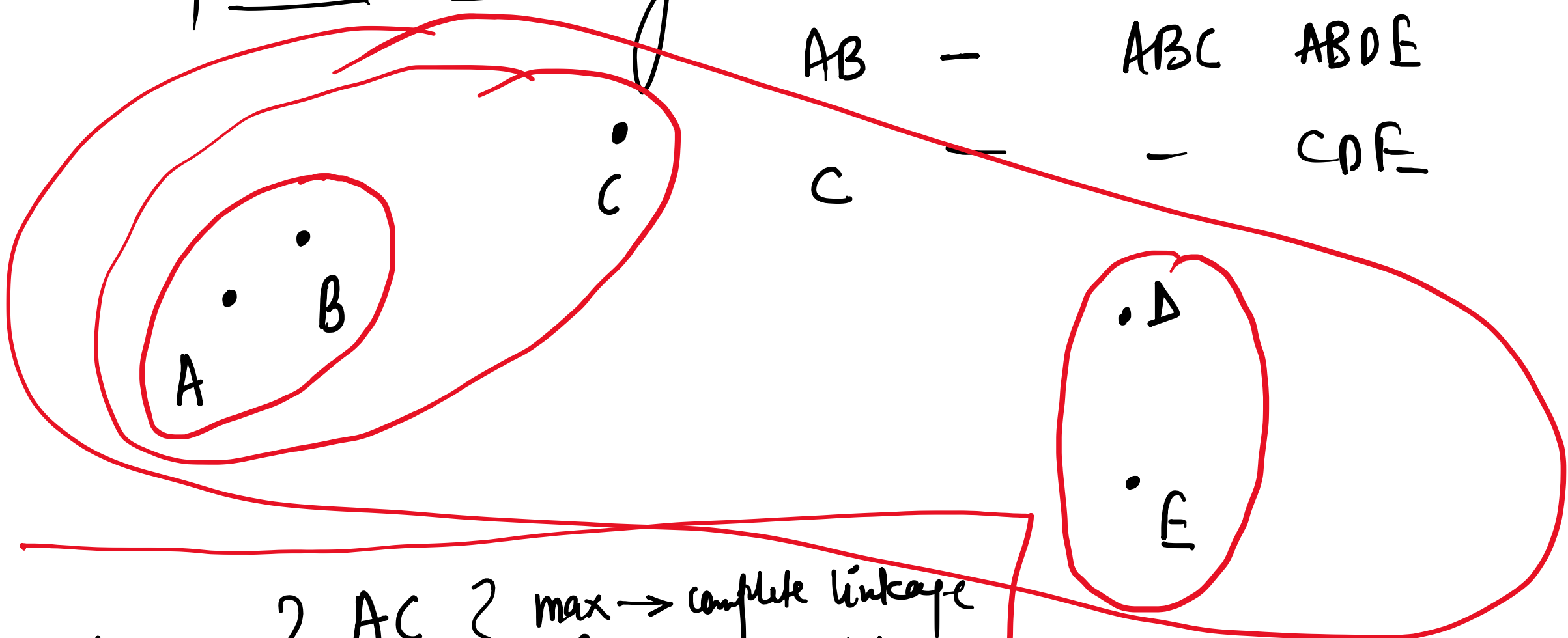
③ U ✓

④ R ✓

① Initialise randomly, but in data space.

② Every other centroid is initialised in a way that is farthest from existing centroids

→ Hierarchical Clustering :-



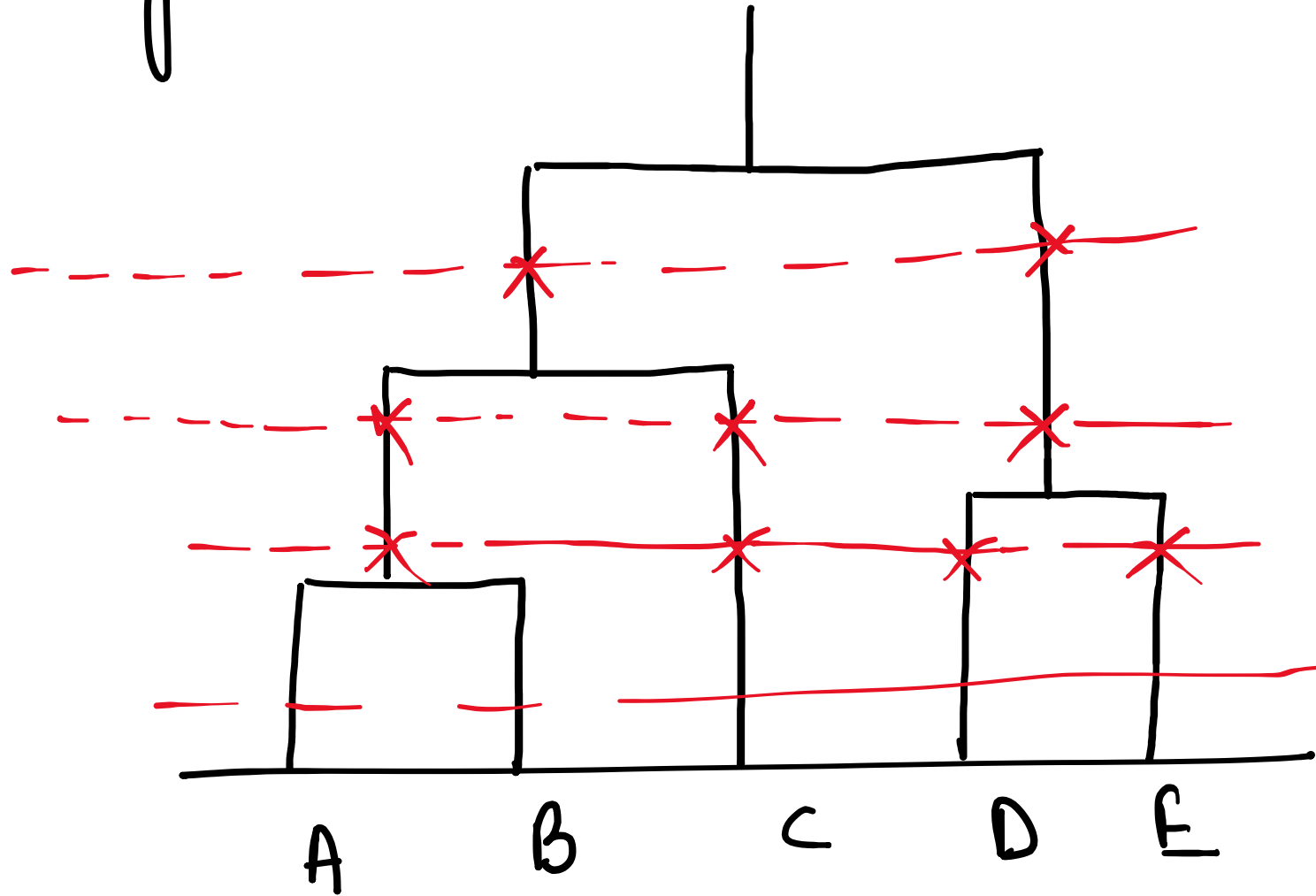
	AB	C	DE
AB	-	ABC	ABDE
C	-	-	CDE

ABC }
 AC }
 BC }
 max → complete linkage
 min — single linkage
 average → average linkage

→ Dendrogram :-

No. of Clusters = 2

ABC DE

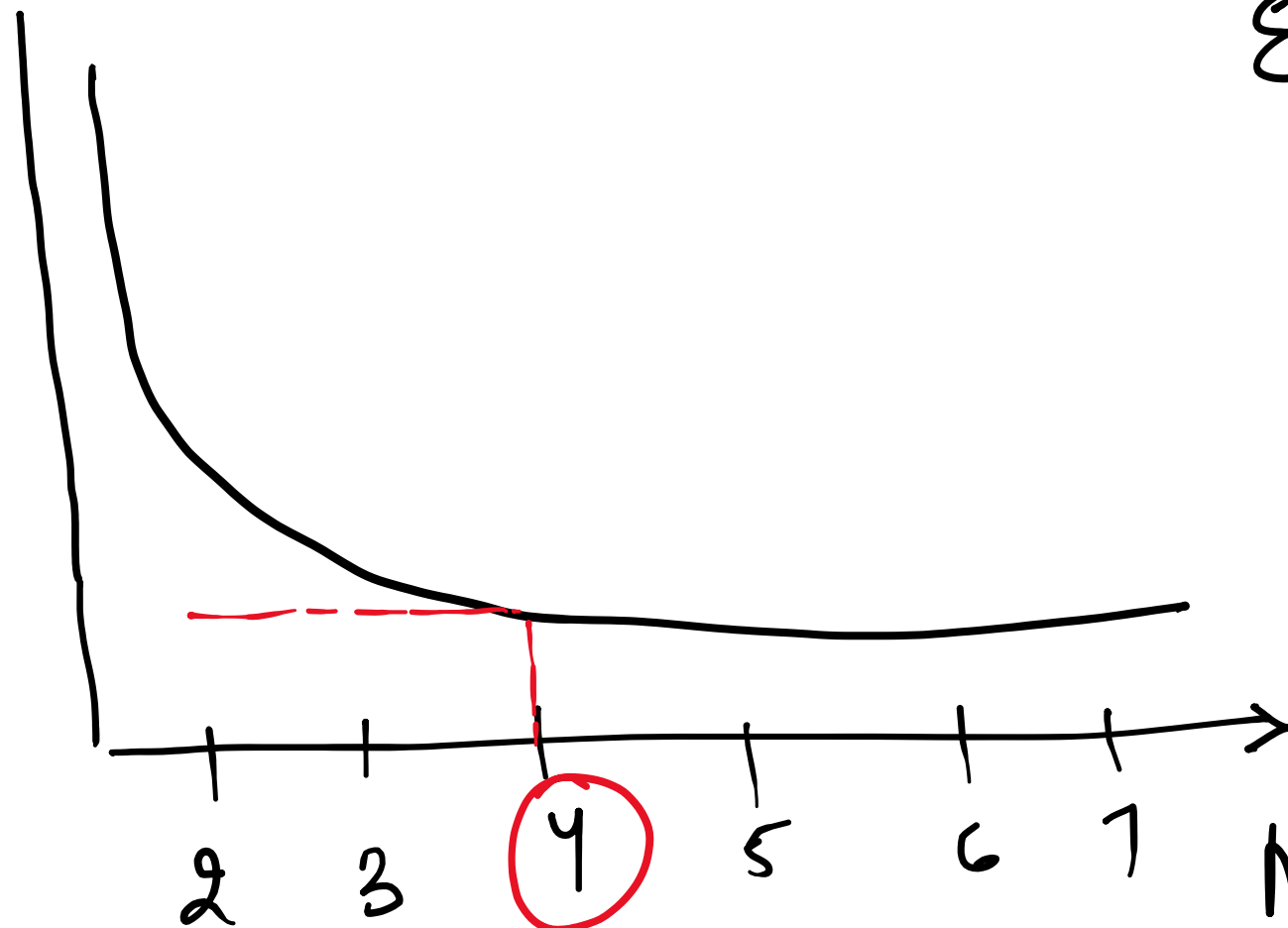


$K=3$
 $AB/C/DE$

$K=4$
 $AB/C/D/E$

Elbow method

Error



No. of Clusters

optimal no. of clusters

