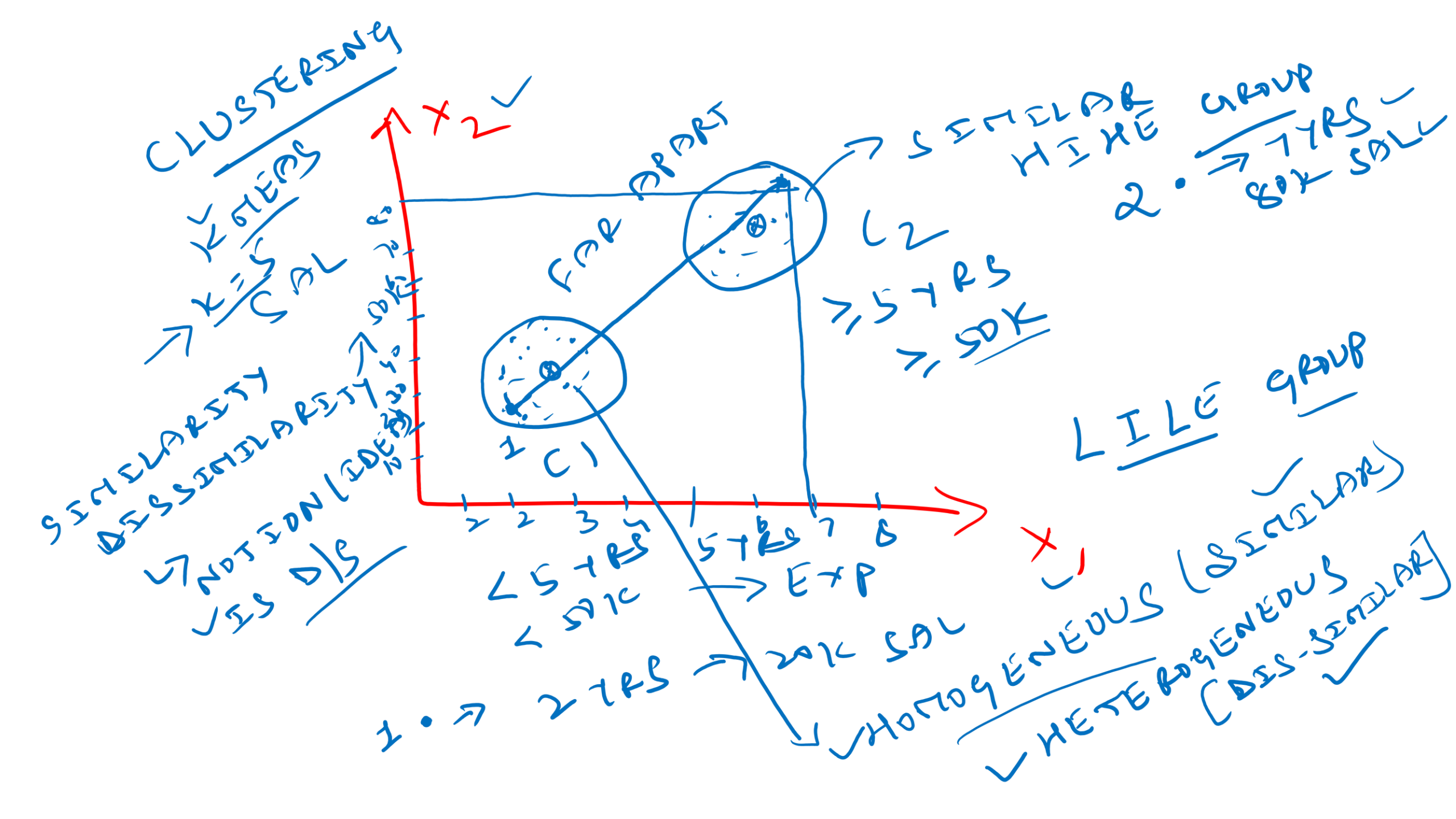
AGENDA – DAY 7 – 06-DEC-2025 (SAT)

- **REACP – DAY 6 – MAX 10 MINUTES**

- **DAY 7**
- Introduction to Unsupervised Learning
  - What Is Unsupervised Learning?
  - Approaches to Unsupervised algorithm (Clustering, Dimensionality Reduction, Association rule)
- Clustering Techniques
  - Overview of Clustering
  - K-Means Clustering:
    - Algorithm and Implementation
    - Choosing the Number of Clusters (Elbow Method, Silhouette Score)
  - Hierarchical Clustering:
    - Agglomerative vs. Divisive Methods
    - Dendrograms and Linkage Criteria
  - DBSCAN (Density-Based Spatial Clustering of Applications with Noise)
- Dimensionality Reduction Techniques:
- Association Rule Learning
- Anomaly Detection Techniques
- Model Evaluation in Unsupervised Learning
- **Q & A**
- **SUMMARY, HEADS-UP FOR DAY 7 & CLOSURE**

REACP – DAY 6 – MAX 10 MINUTES

- Decision Tree
- KNN
- SMOTE
- Random Under sampling
- Multi-label classification
- Random Forest
- Leave one out


CLUSTERING
K GROUPS
K = 5
SIMILARITY
DISSIMILARITY
NOTION (IDEA)
YES D/S
X_2
X_1
FAR APART
C1
C2
1
2
3
5
6
7
8
20K
30K
40K
1 → 2 YRS
< 5 YRS
< 30K
5 YRS
EXP
20K SAL
> 5 YRS
> 50K
SIMILAR
MIXED
2 → 7 YRS
50K SAL
LIKE GROUP
HOMOGENEOUS (SIMILAR)
HETEROGENEOUS
(DIS-SIMILAR)

CLUSTERING:

1000 CUSTO → 500K

10K → PREM 250K

SL → RF → "ACCURACY"

↳ UNSUP LRNG → NO LABELS

↳ PATTERNS, SIMILARITIES & DIS-SIMILARITIES

↳ "SUBJECTIVE SEGMENTATION"

DECIDING # OF CLUSTERS

↳ "DIVIDE & CONQUER STRATEGY"

MRKTG → POSITION YOUR PRODUCT

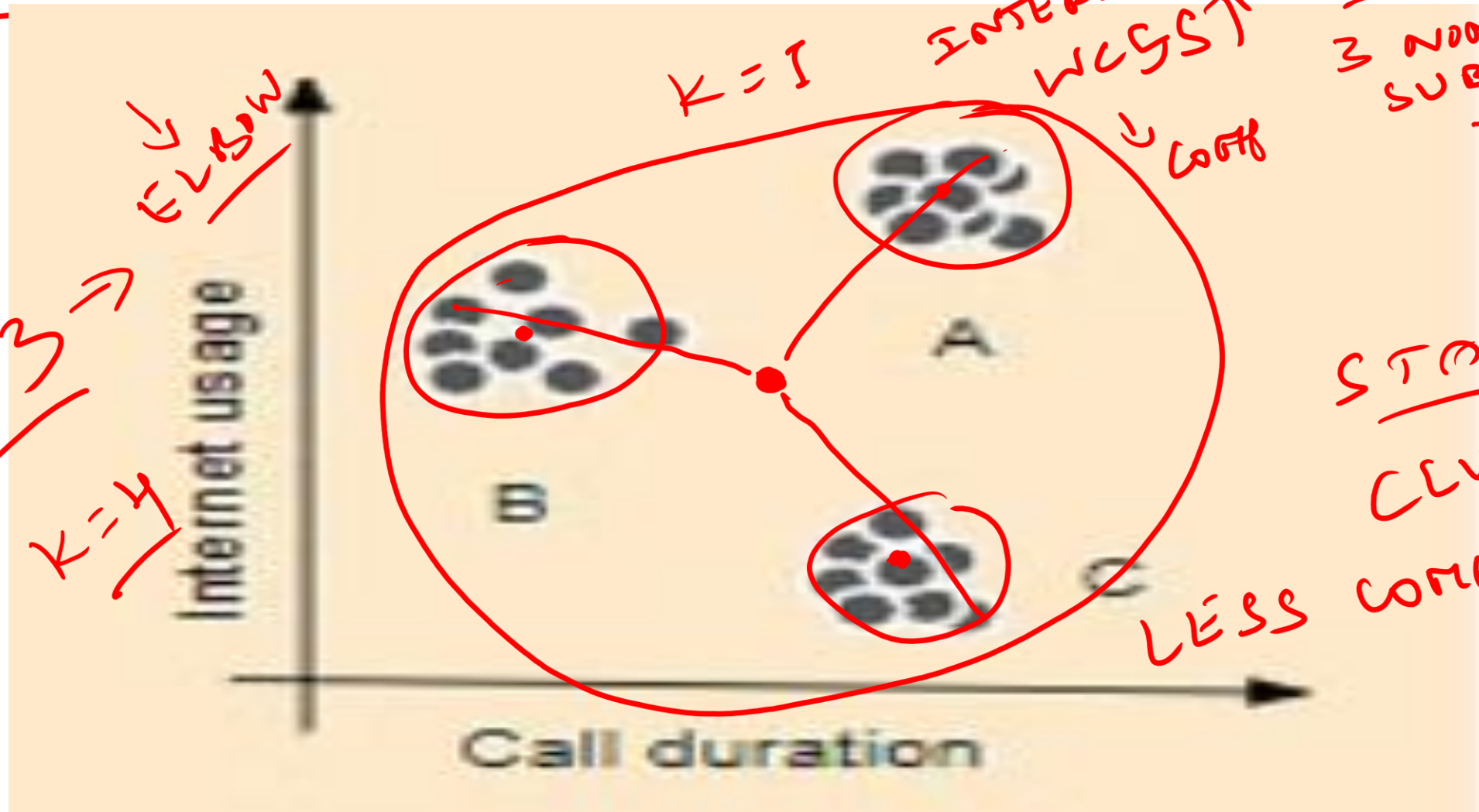
ELBOW METHOD (K=5)

DOMAIN UND

6

4

KMEANS CLUSTERING



K = 1

INTERIA WCSS ↑

↓ COST

K = 3

3 NON-EMPTY SUBSETS

3 CENTROID

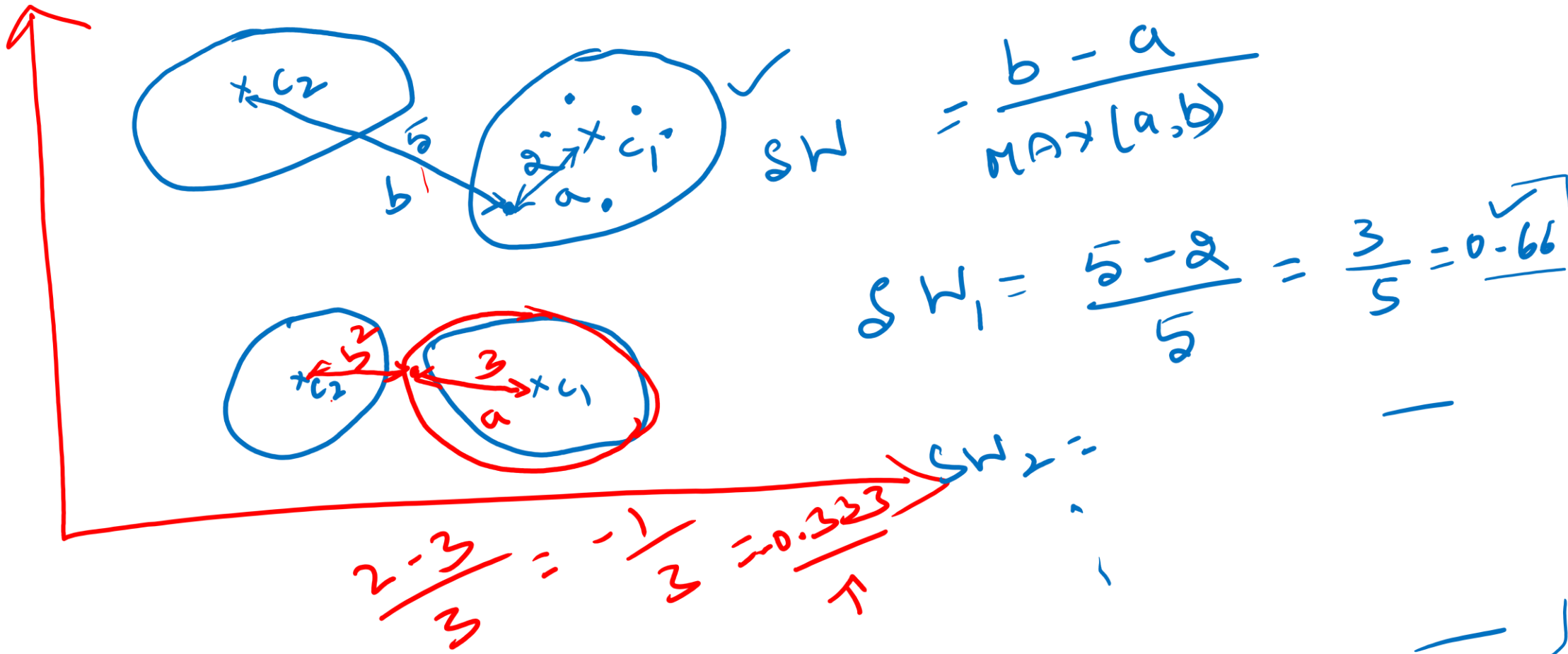
↓ ELBOW

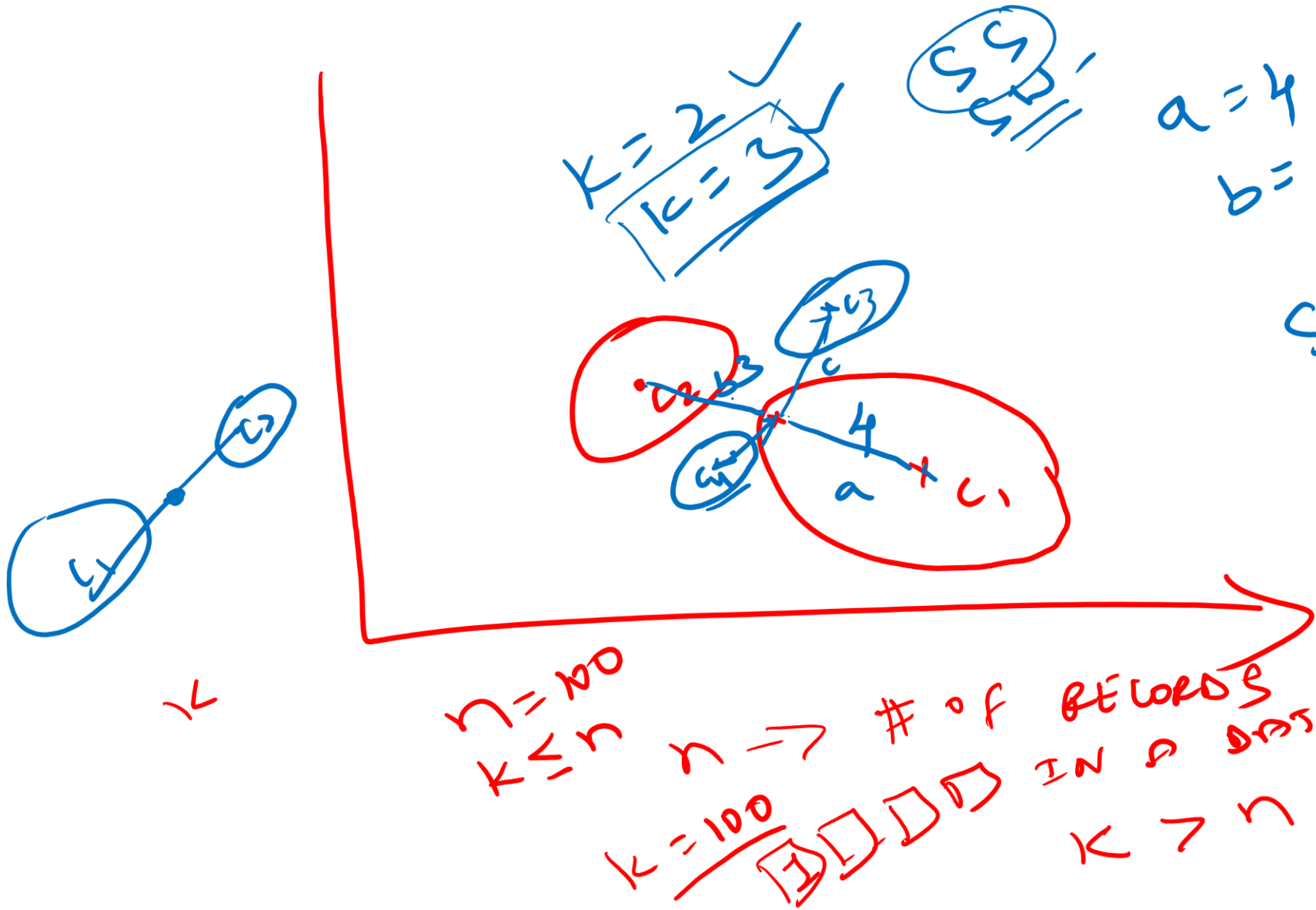
✓ K = 3 →

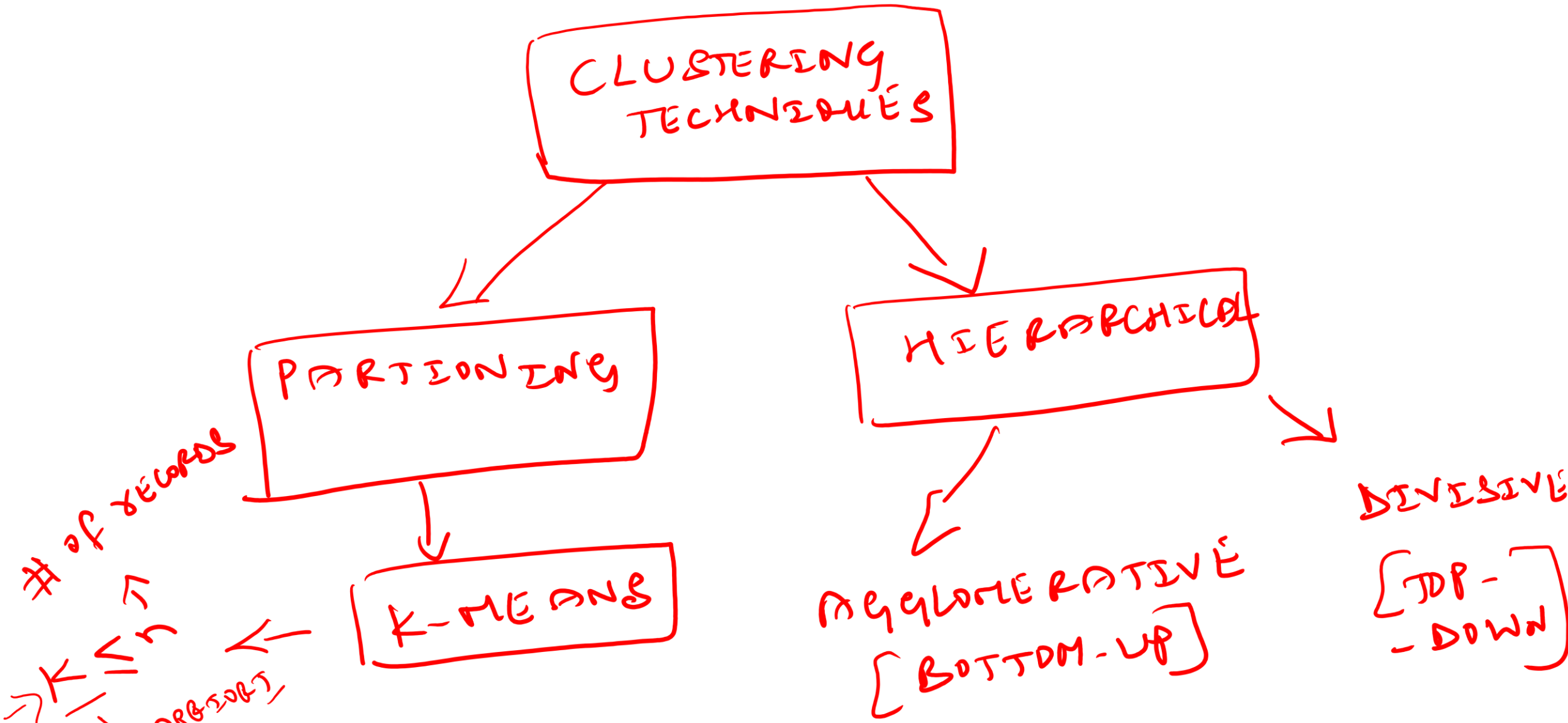
K = 4

STABLE CLUSTERS

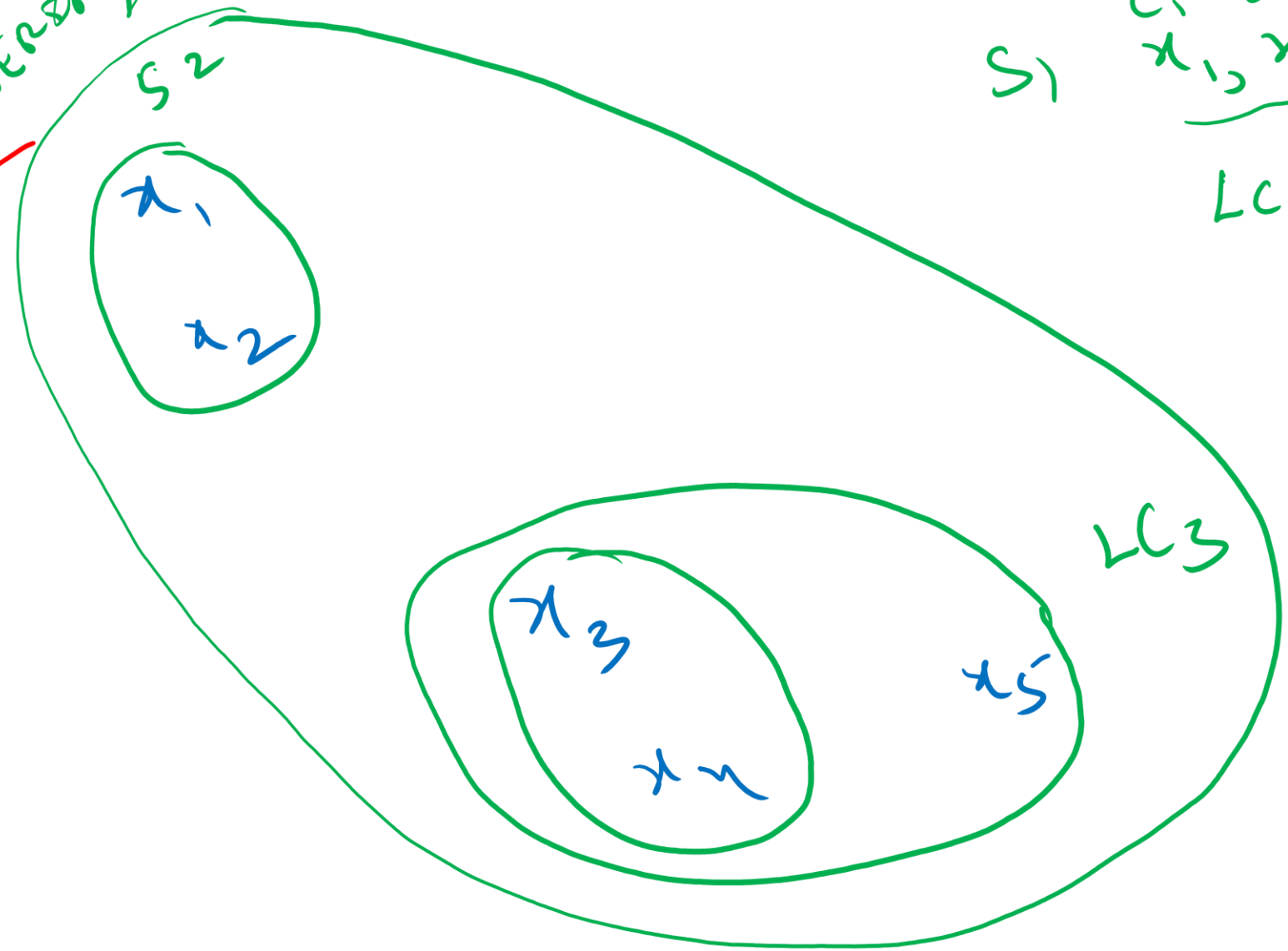
LESS COMPACT


c2
5
b
2
c1
a
SW = (b - a) / MAX(a,b)
SW1 = (5 - 2) / 5 = 3/5 = 0.66
2
3
c2
3
c1
a
SW2 =
2-3/3 = -1/3 = -0.333
SW100 70 ...


K=2
K=3
a=4
b=3
SW = b-a / max(a,b)
= 3-4 / max(3,4)
= -1/4
= -0.25
-VE
K
n=100
K≤n
n → # of RECORDS IN A DATA SET
K=100
K > n ?


CLUSTERING TECHNIQUES
PARTIONING
HIERARCHICAL
# of records
K-MEANS
AGGLOMERATIVE
[BOTTOM-UP]
DIVISIVE
[TOP-
-DOWN

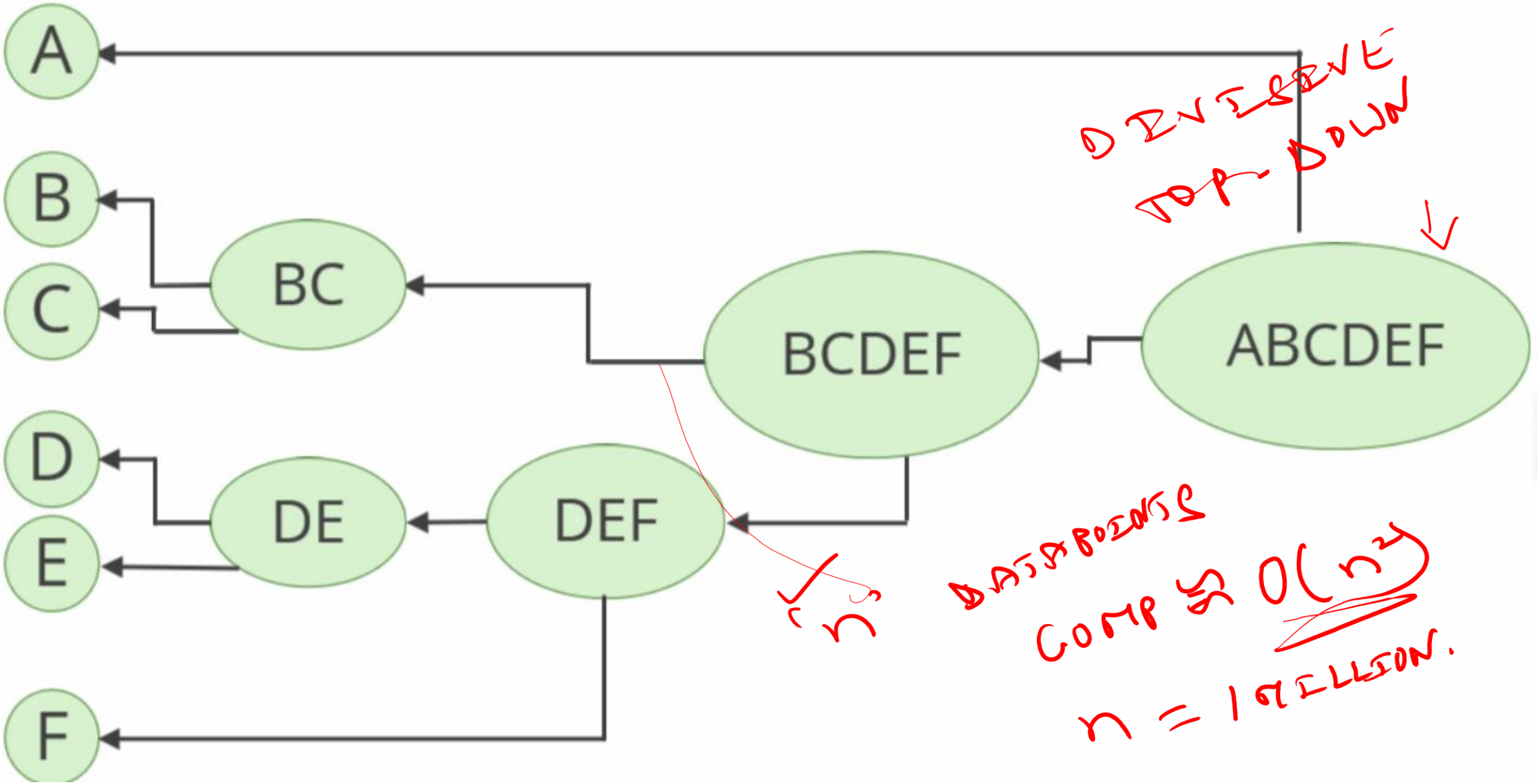
Agglomerative Clustering

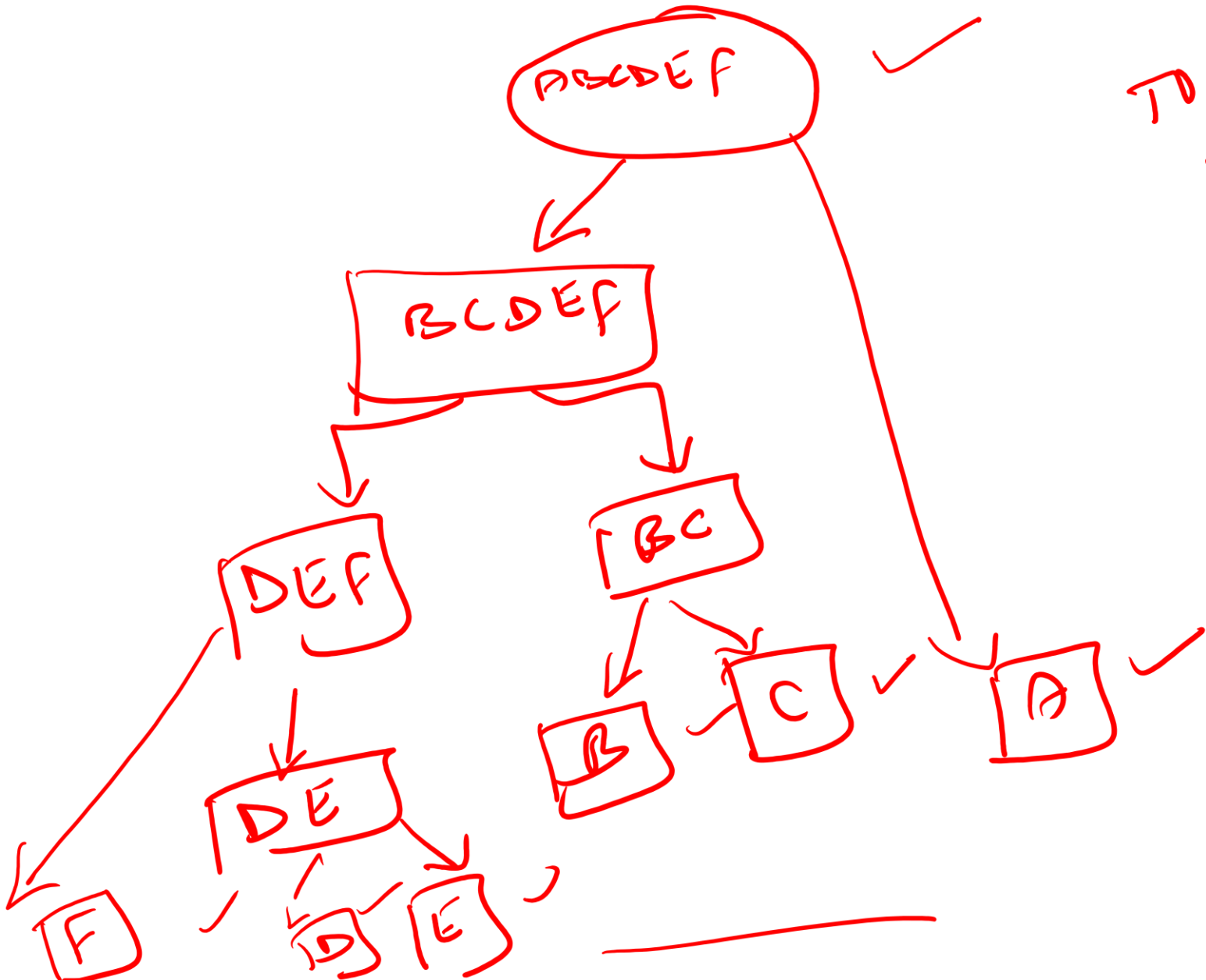


S1 x_1, x_2 , x_3, x_4, x_5

C_1 C_2 C_3 C_4 C_5

LC_1 LC_2


A
B
C
BC
BCDEF
ABCDEF
D
E
DE
DEF
F



TOP-DOWN
DIVISIVE
HIERARCHICAL
CLUSTERING

$O(n^2)$

PARTY → 100 PEOPLE.

AGGLOM BOTTOM-UP

✓ PAIRING PEOPLE → UNTIL ALL 100 ARE GROUPED SO GR

DIVISIVE.