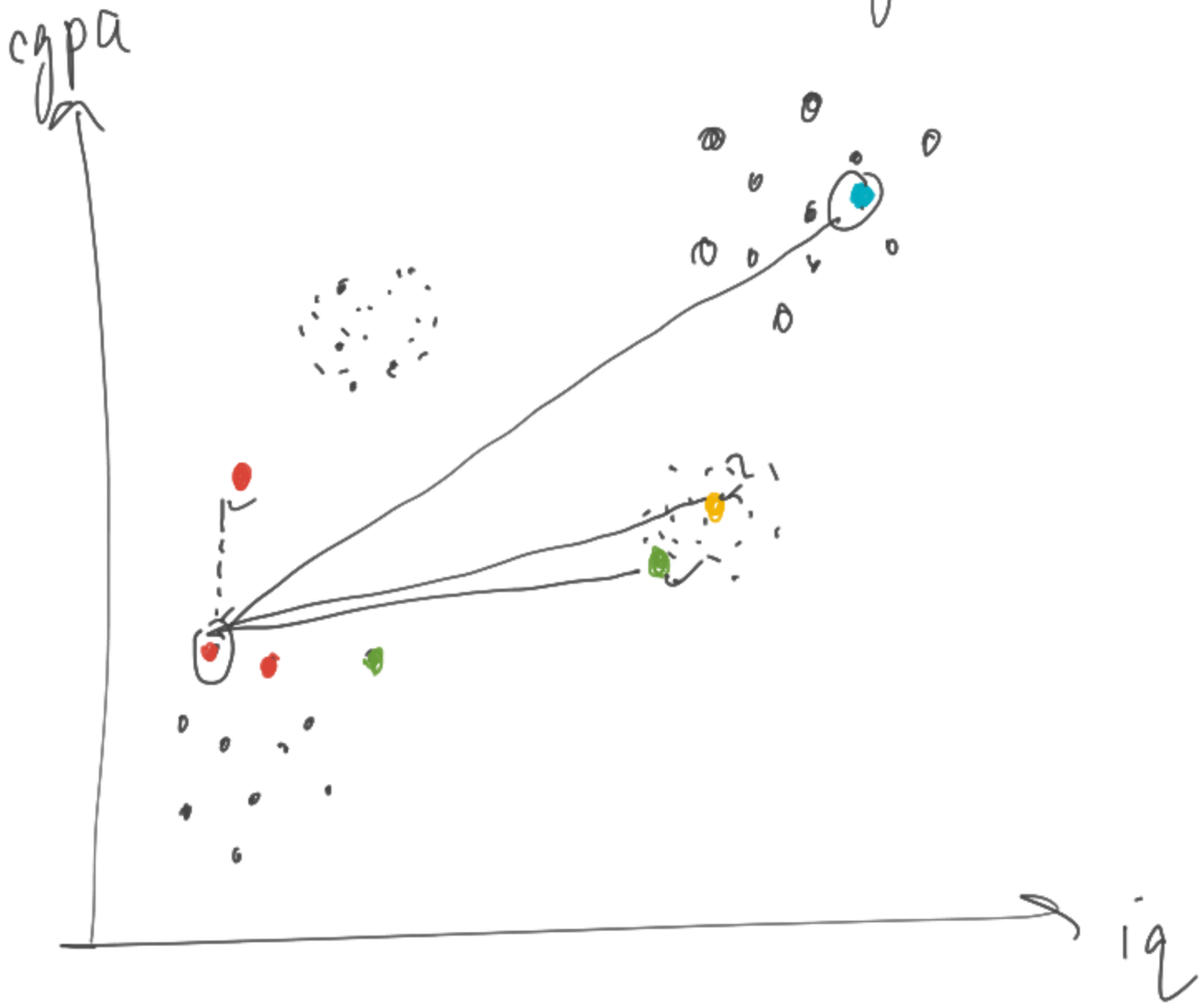
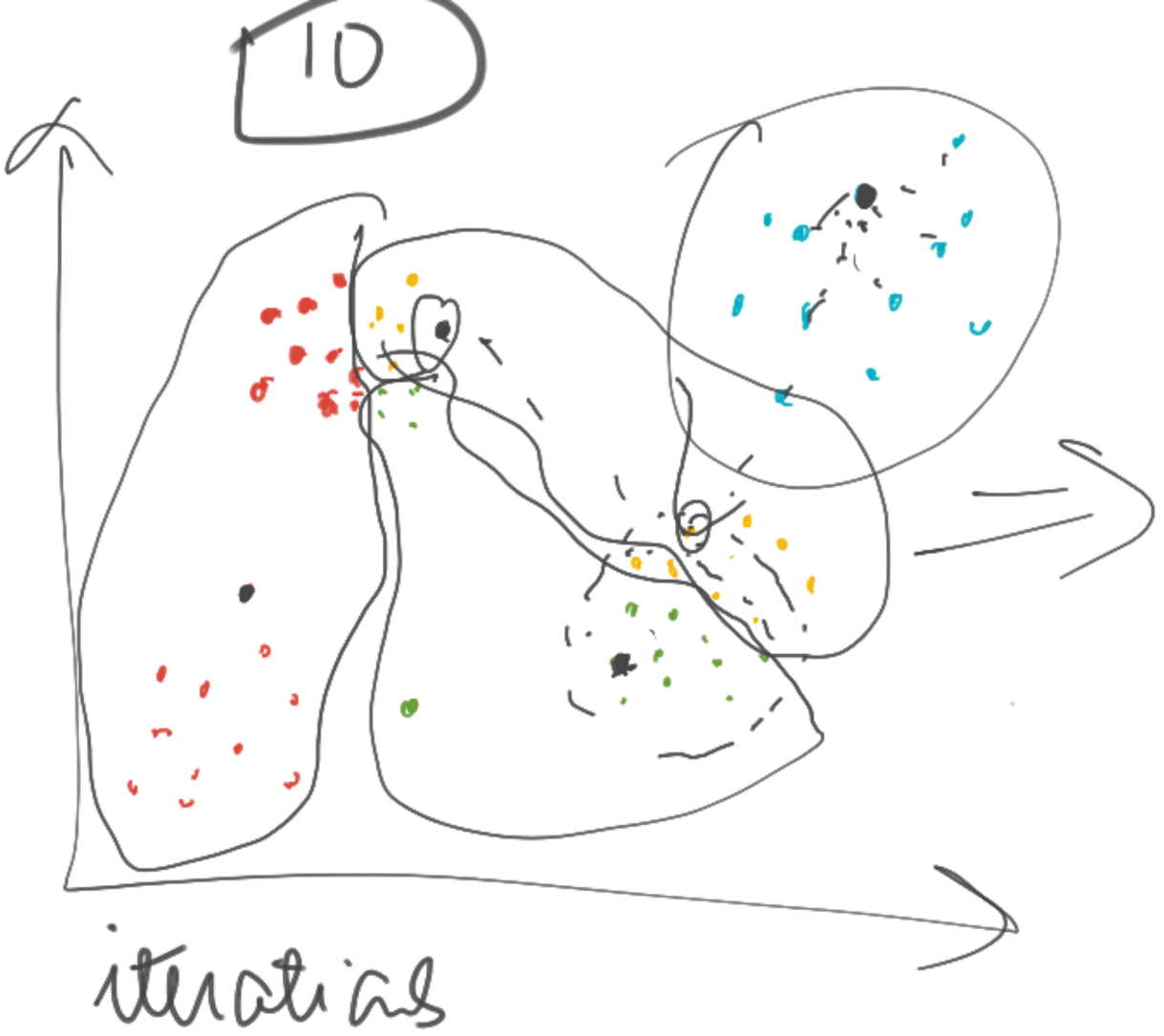
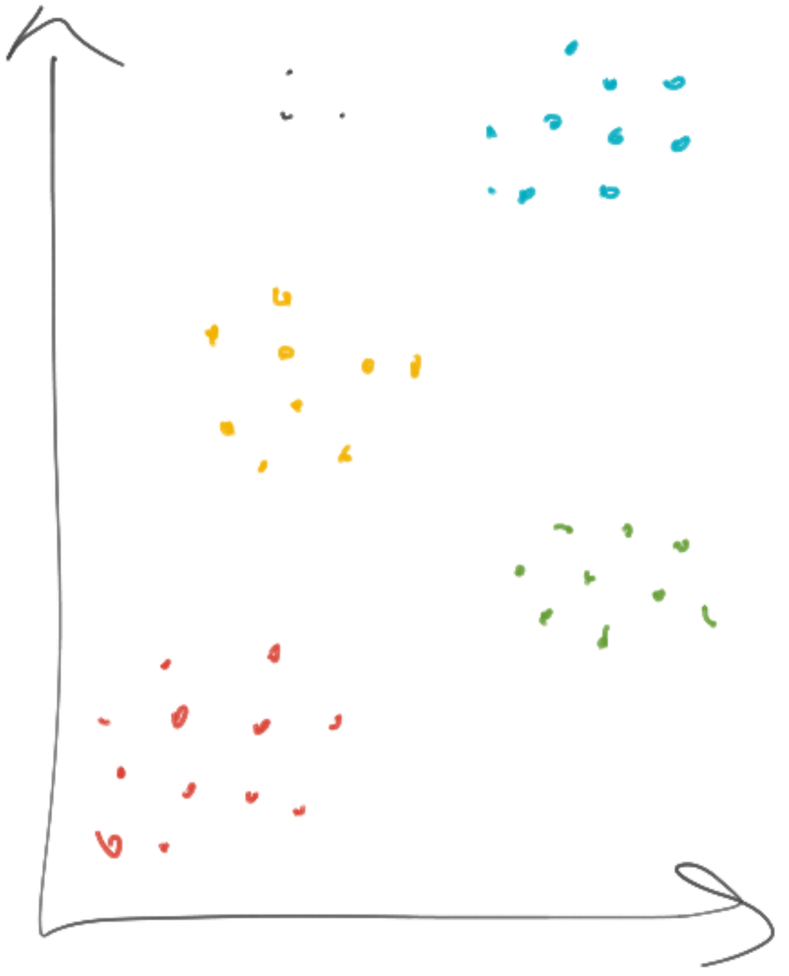
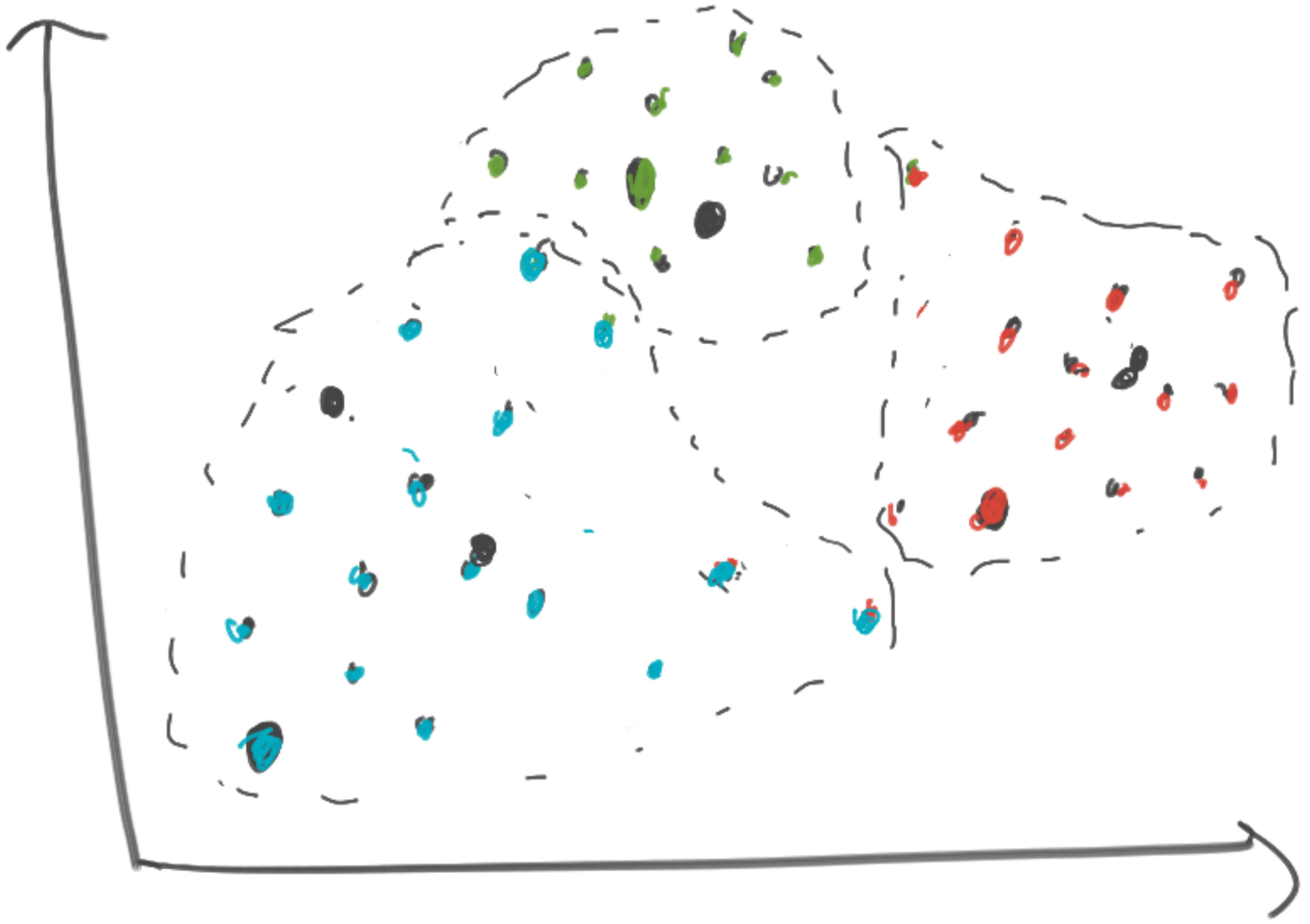
K-means Clustering




10
iterations



K = 3



$K = 1 \Rightarrow$ worst case

$K = 2 \ldots \quad 10$

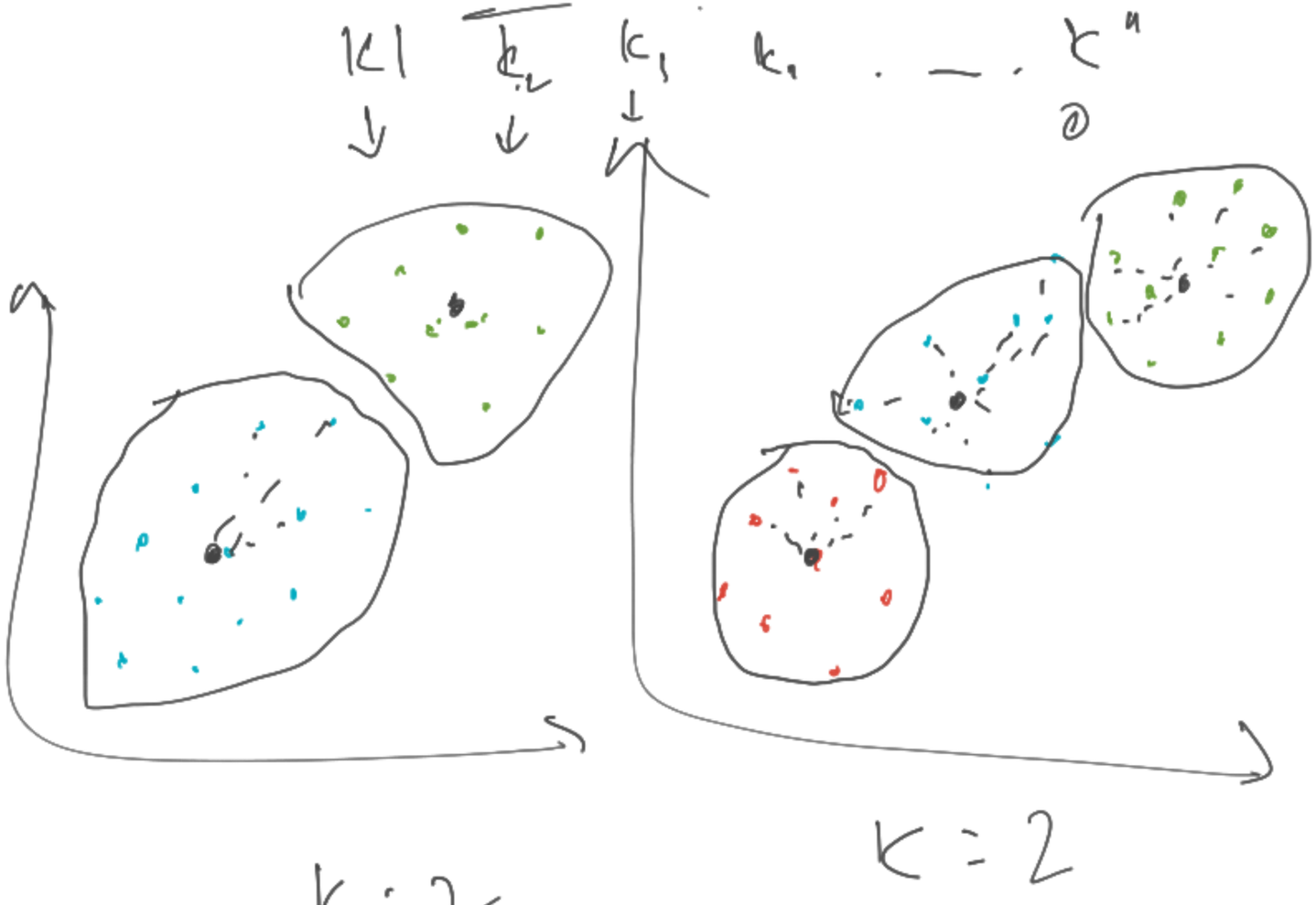
Sum of distance $\{S_1\}$ = huge

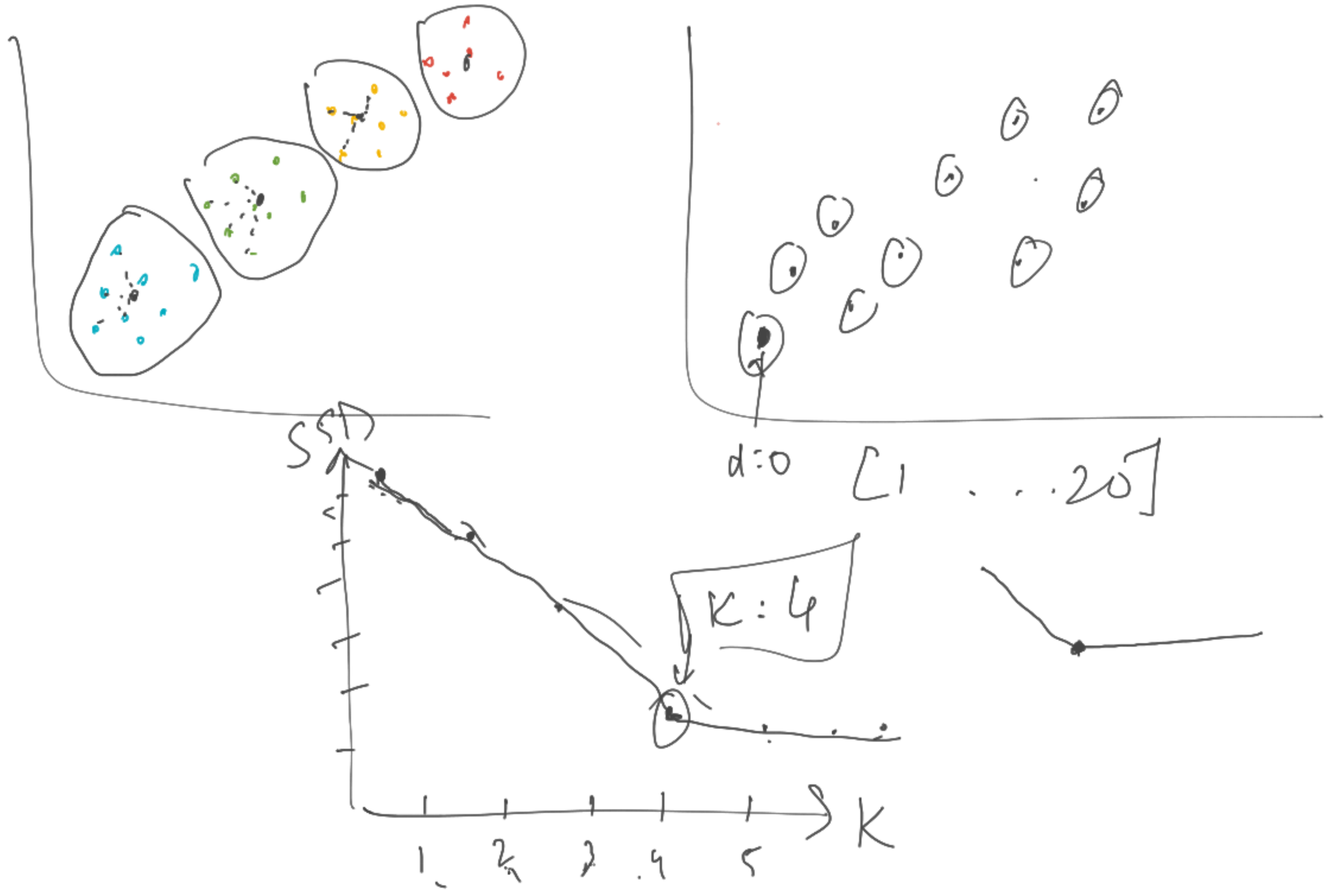
$\{S_1 + S_2\}$

$\{S_1 + S_2 + S_3\}$

$\{S_1 + S_2 \ldots S_n\} = 0$

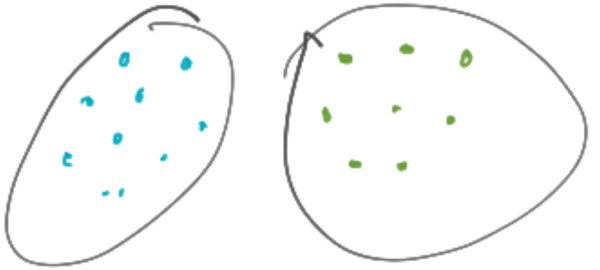
100 ⇒ 100




S(d)
d:0
[1 . . . 20]
K:4
K
1
2
3
4
5

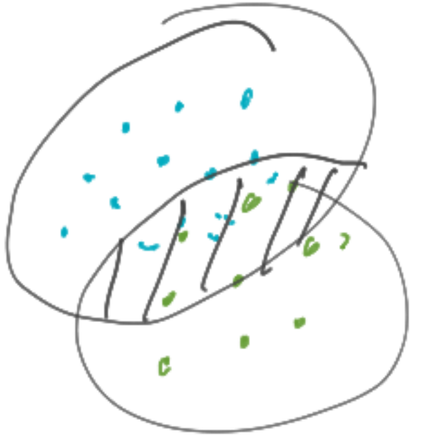
Hard clustering

- each pt belongs to only one cluster

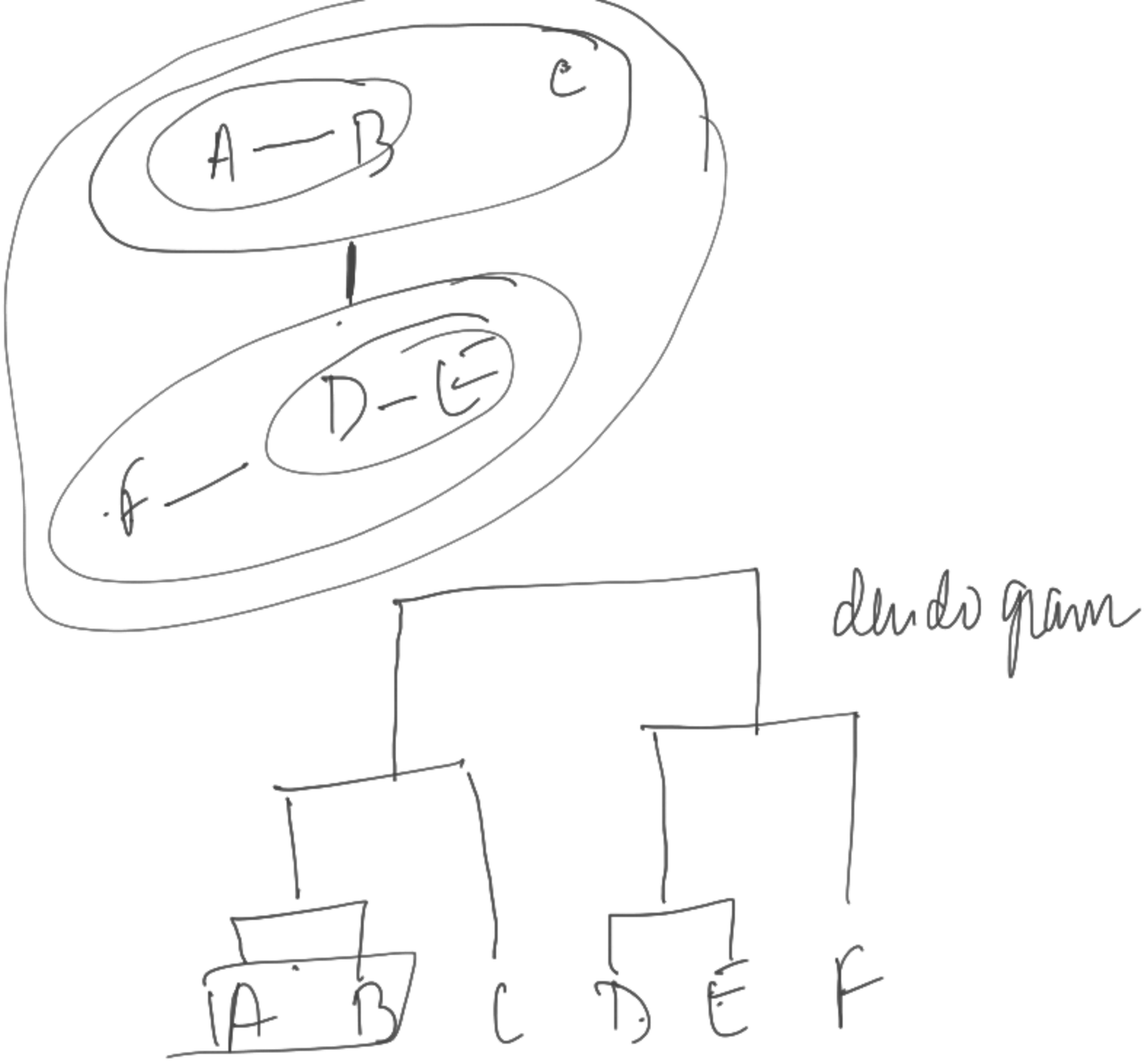


Overlapping

- each pt can belong to multiple cluster



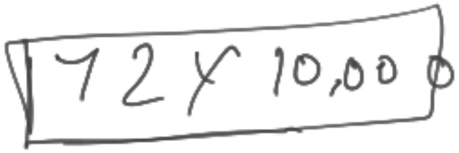
Hierarchial



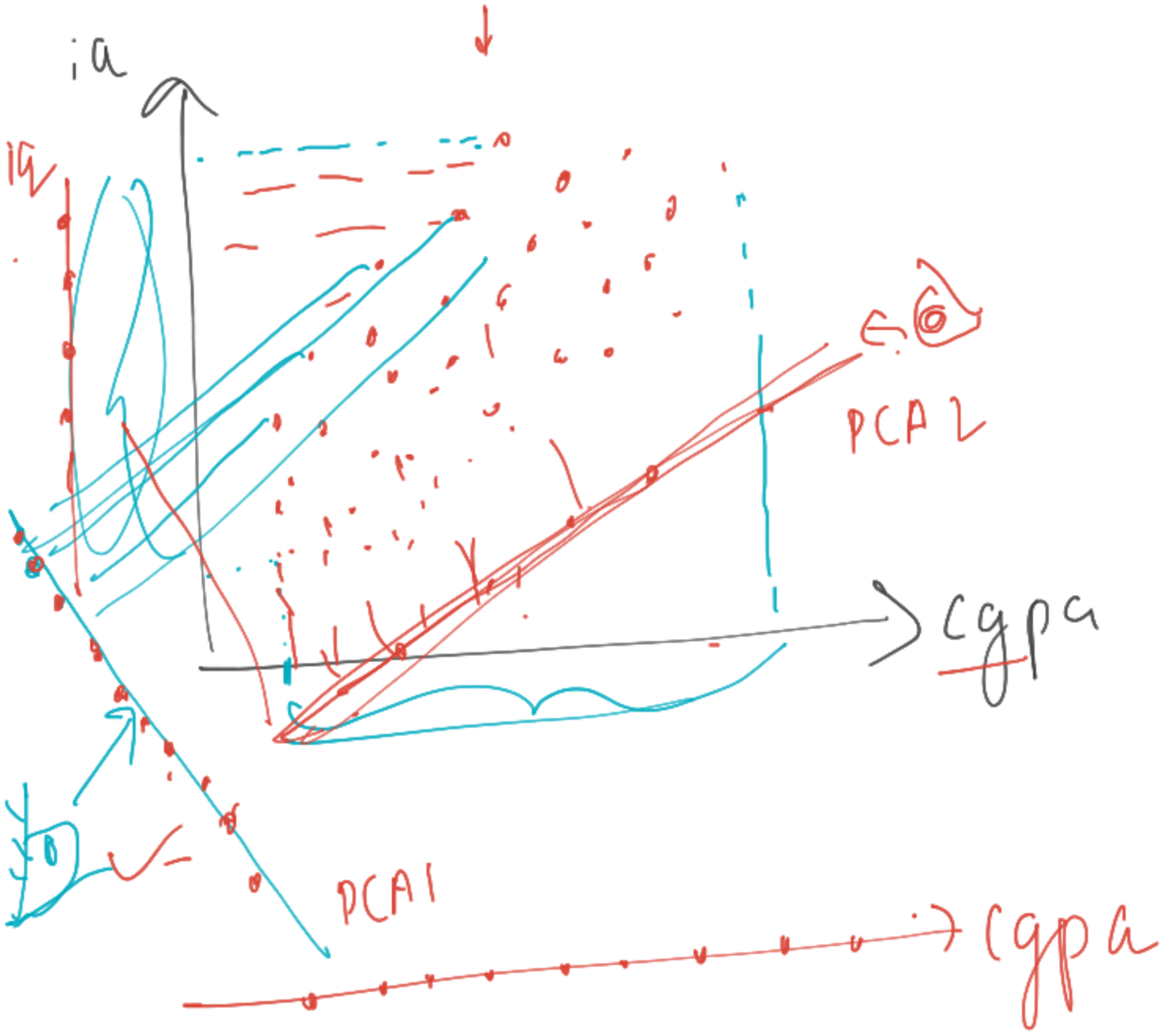
dendrogram

PCA (Principal Component Analysis)

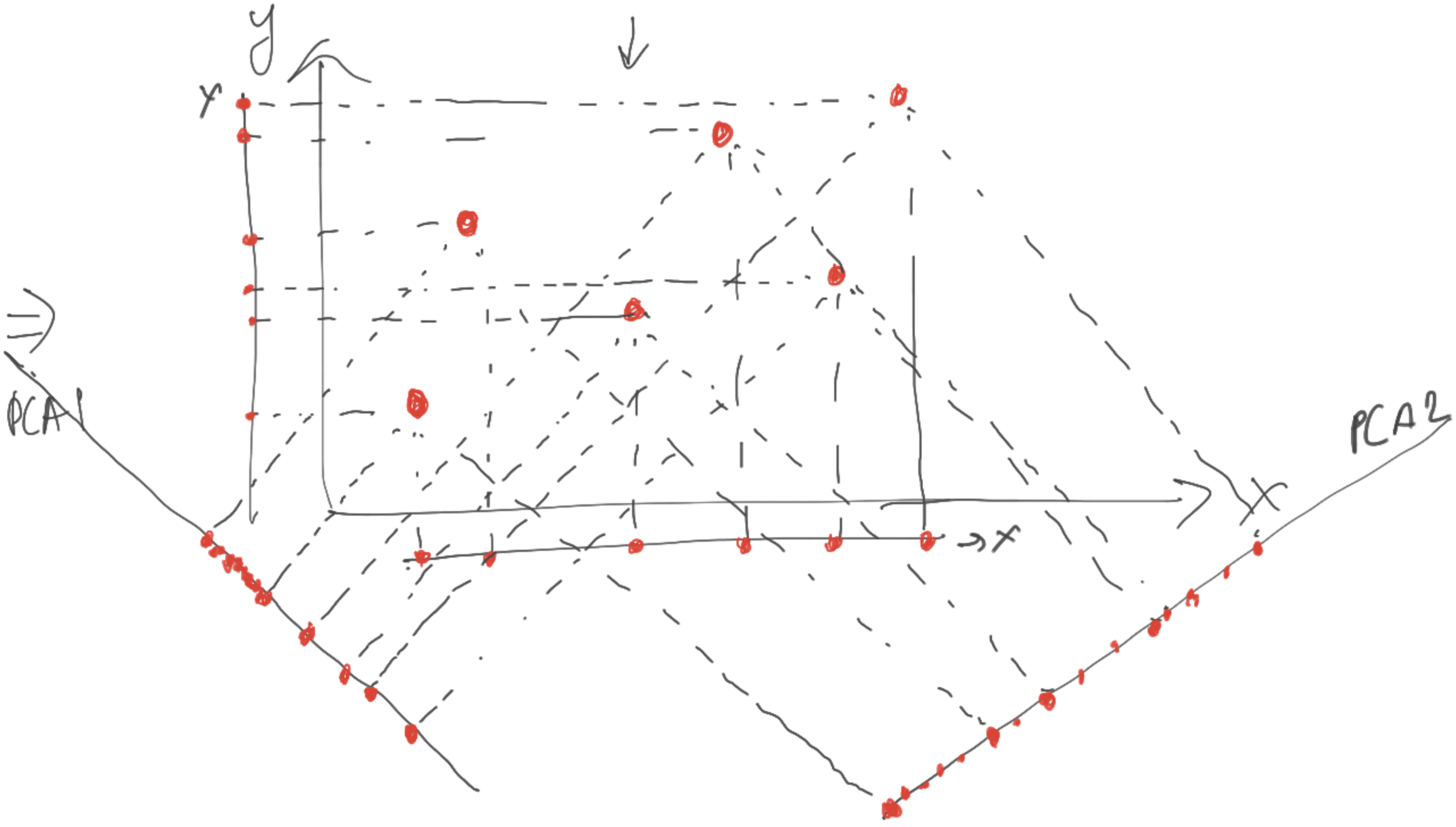
- Dimensionality Reduction



CV
hyper


ia
iq
PCA 2
cgpa
PCA 1
cgpa


iq cgpa is cost
iq is less cgpa


y
x
PCA1
PCA2
x