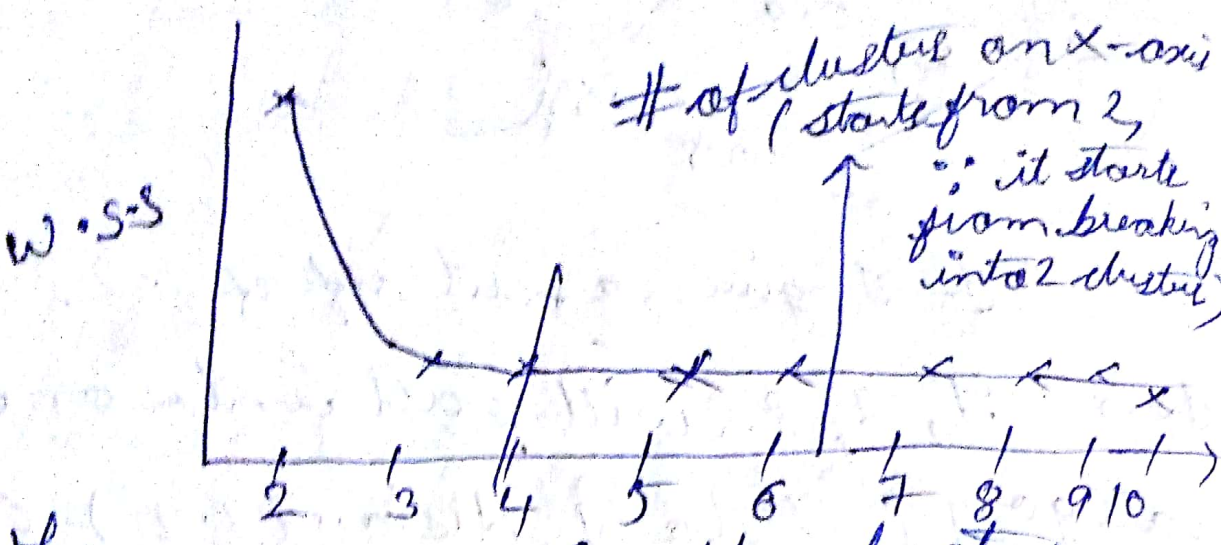
Spree graph / elbow curve:



when we are ↑s the cluster,
there is a large decrease in
Sum of squares.

(group) (they have to be similar)
→within cluster, they have to be homogenous
→across group
after cluster they have to be heterogeneous
(different from each other)

[As the graph starts becoming
flat that is where we have
to decide to stop further clustering, in
the above case it ~~stands~~ stops after 4
As we breaks the observations
into clusters, the observations

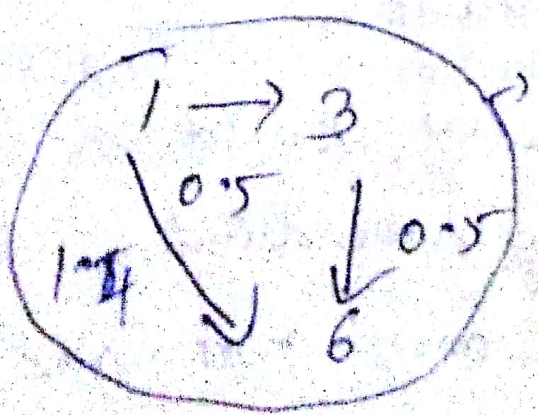
keep getting closer to the mean and in that way the variance would be ~~dec~~ decreasing and WSS too.

Q. How to decide the value of k? (we look ~~oth~~ at the Q. Spree graph)

[In clustering, most of the times to know similarities, distance function is used]

[When Z score is used, the data is broken b/w -3 to +3]

cut off = 1.5



taken as one cluster (for this the method used is ward (WARD) method)