

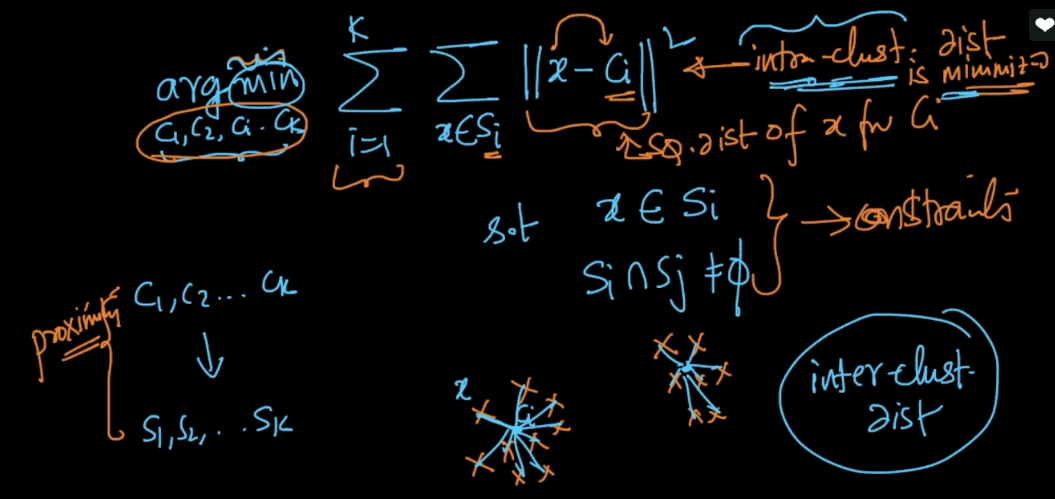
So in K-means clustering, the task is to find the k-centroids for k sets, where the constraint is that every point belongs to a cluster. And there is no common point in any of the two clusters.

Given this our objective function is as mentioned in below image.

Which basically says that to minimize the intra cluster distances.

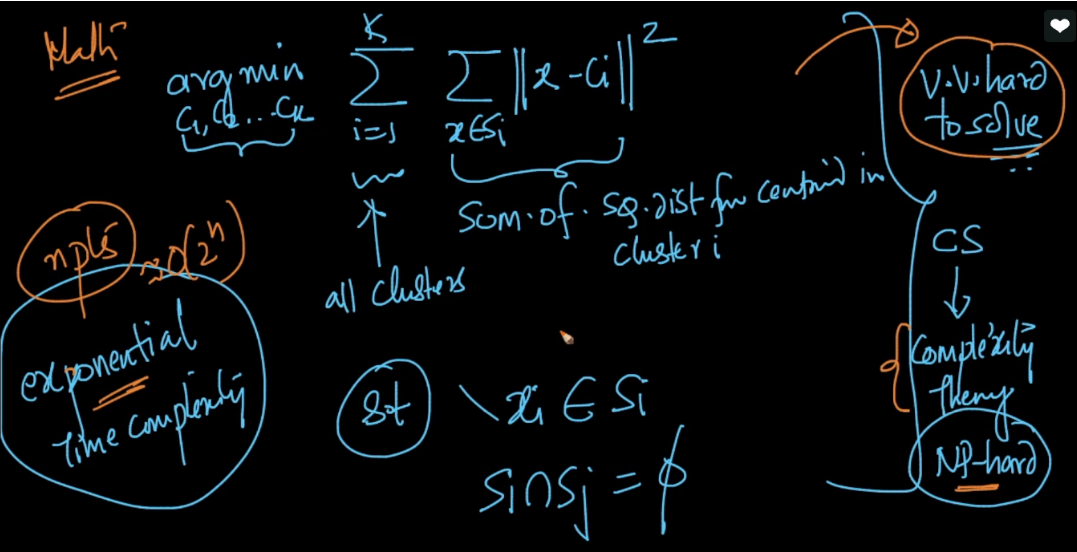
In this objective function we are finding c\_i’s where the distance of each point in that cluster from centroid of that cluster c\_i is minimum.

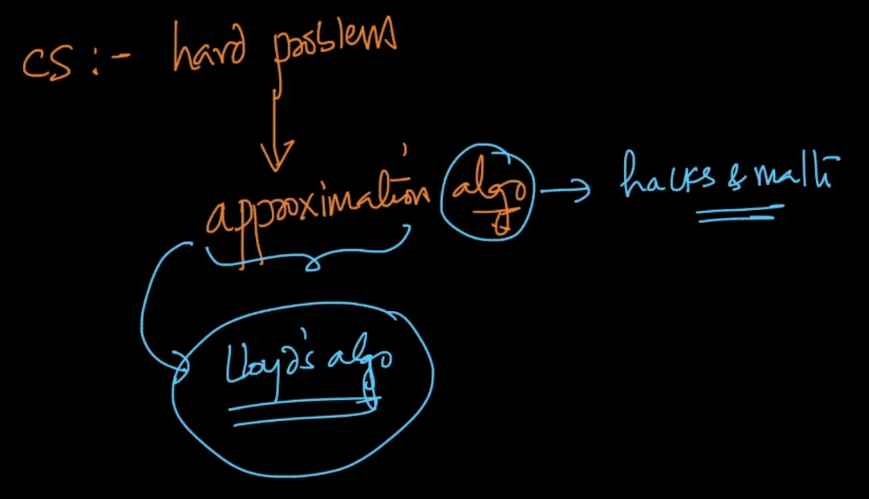
That means we are finding centroids for each sets such that all the points in a set are very close to each other.



This objective function is very hard to compute and it comes under NP-hard problems, for such problems there are no such exact solutions, there are only some approximate solutions and algo which gives this approximate solutions are called approximation algorithm.

There is a algorithm called Lloyd which gives approximation result for this objective function.





Data preparing for k-mean clustering <https://www.edupristine.com/blog/k-means-algorithm>