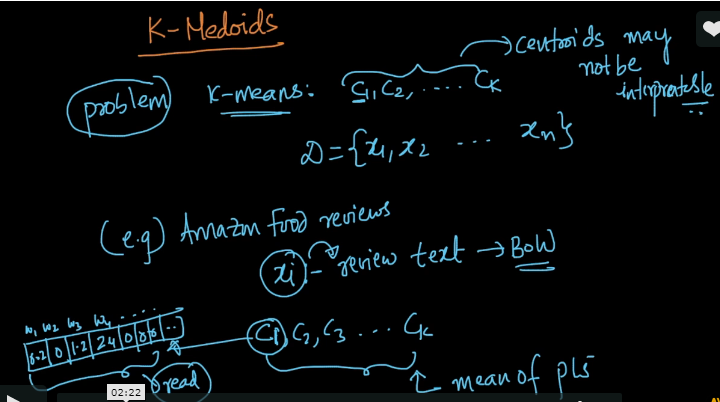
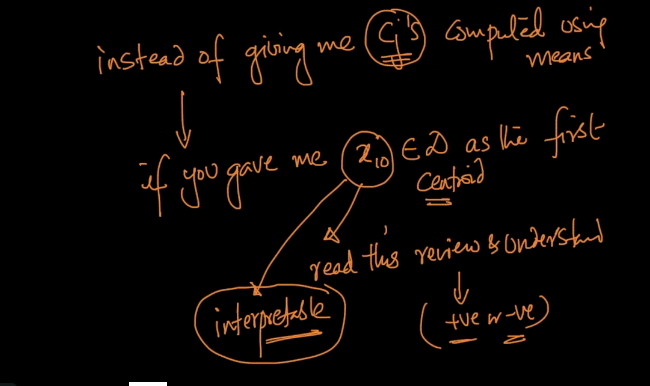
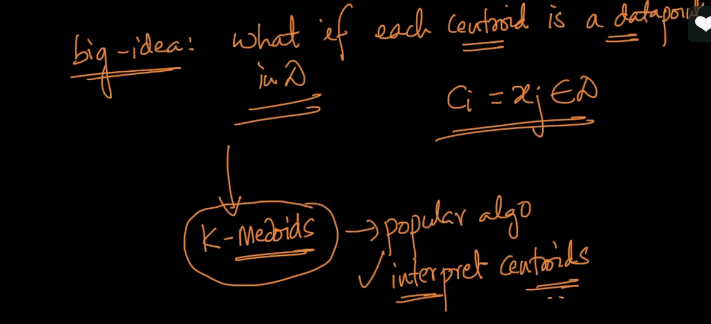
Now in k-means the problem we have is the centroid may not be interpretable i.e. if we take example of Amazon Food Reviews dataset and say we have Xi review text and it is BOW vectorized so it will be not be interpretable to read.



So the idea is instead of giving Cj which is computed using means of points if we get index or point which belong to Dataset as the first centroid so it will be more interpretable.



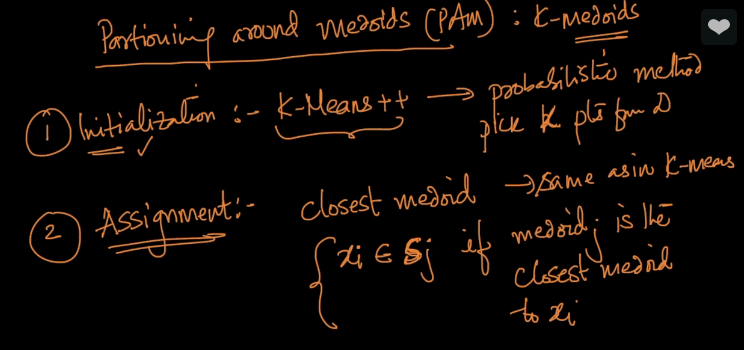
Now it all means is what if instead of Centroid being a vector which is unterpretable what if we get a datapoint for each centroid and that’s what k-medoids do.



So the algorithm for K-medoids is PAM (partitioning around medoids).

It is very similar to Lloyds algorithm i.e. in first step we initialize but in some different manner i.e. instead of Cj we get datapoint as Cj.

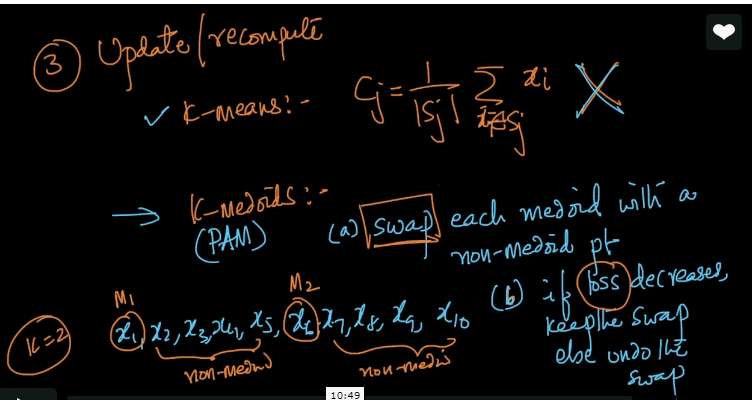
Second stage is all same Assignment which is we assign points to set where they are getting closest medoid.



Now comes the third stage which is bit different than K-means.

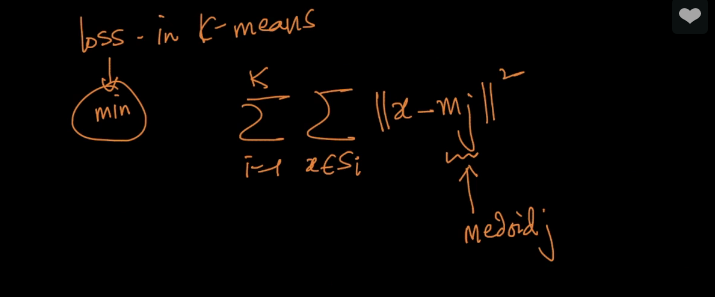
Instead of calculating Cj by calculating means of every point we use Swapping method.

i.e. we swap ever medoid point with a non – medoid point ,say m1 is medoid and m2,m3,m4 are non medoid so we swap m1 by all one by one and check if loss reduces or not and if loss reduces than we keep the swap else we undo it. So we need to perform multiple swap for all the medoids and check if loss reduces or not.



Now the question here is what is loss in k-medoids?

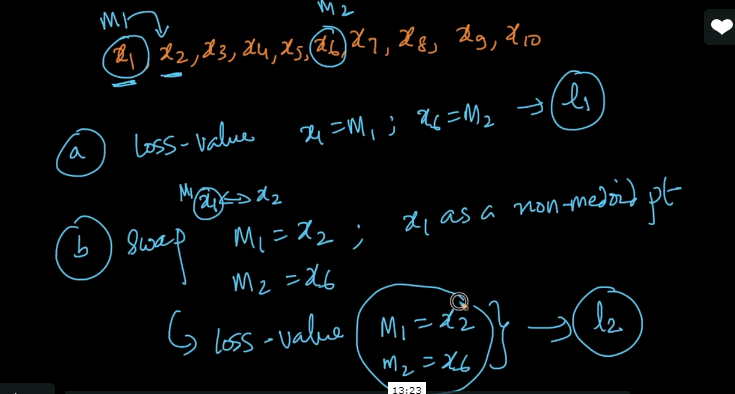
So here sum of all {i} for all sets ||x-mj||^2.



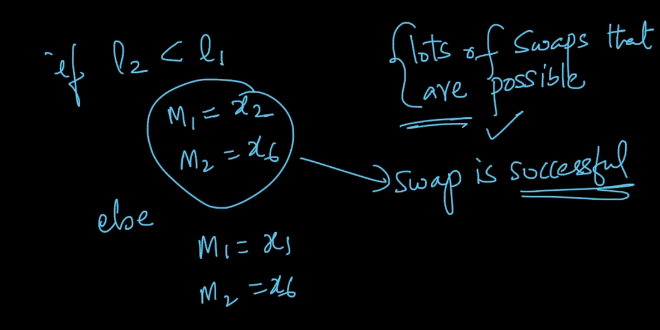
Now lets take an example to understand the concept of swapping.

So in below example we need to swap x1 and x6 by every other point and check if loss decreases or not.

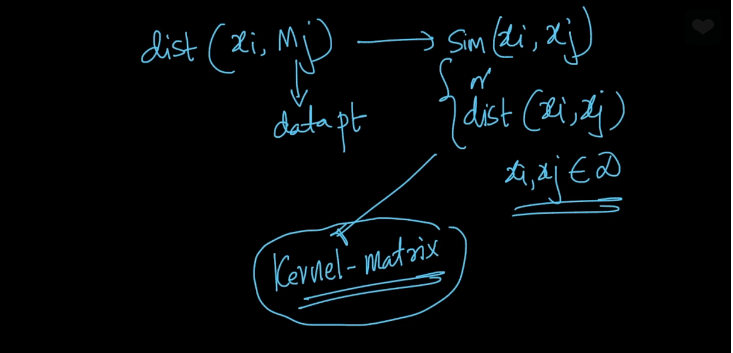
If yes than keep the swap else undo it.



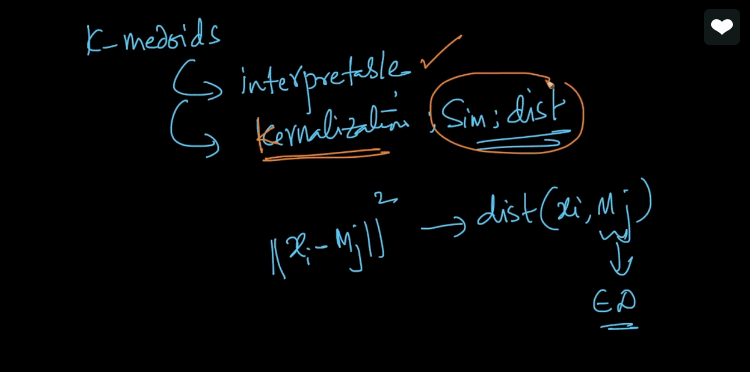
Condition which shows if swap is successful or not.



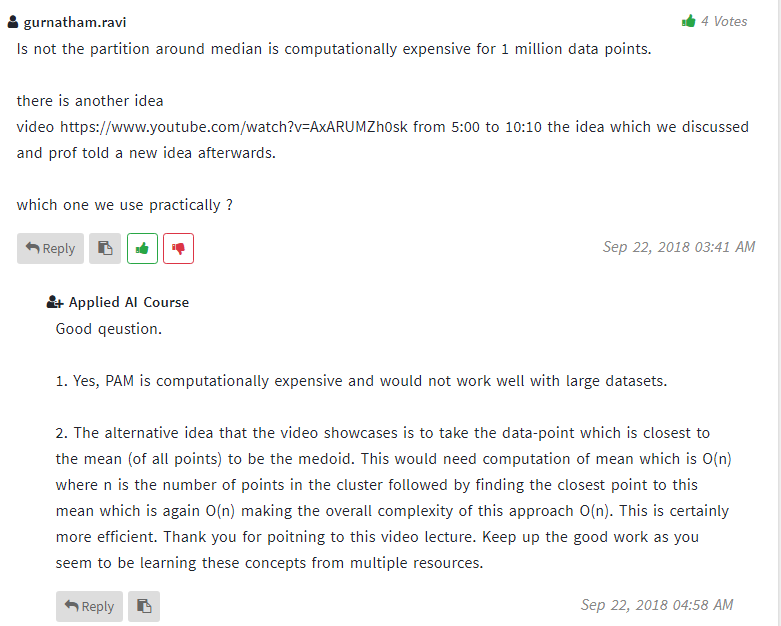
What is someone gives us kernel matrix for all points ,life will be so easy (😊)



The advantages of K-medoids is that it makes all the points interpretable and one thing is kernelization i.e. we get Simlarity and Distances so we just need to replaces ||Xi-Mj||^2 with dist(Xi,Mj) belonging to Dataset.



**Comments:**



<https://www.youtube.com/watch?v=AxARUMZh0sk>

