

Wiki Activity: Clustering (Personal reflections)

Watching the first animation of K-means clustering by starting the process with 4-left most points, and then re-watching it by starting with 4 random points in one cluster, 4 bottom points, 4 top points, 4 right-most points, I was able to see how the process evolves. Through this animation, it was presented the end-to-end process from the initial randomly assigned centroids, to the continuous assignment of points to their nearest centroid, to recalculation and repositioning of centroids till no further reassign points and update centroids steps were needed, as the final clusters have been formulated.

Watching the second animation and being able to do the initial random selection of centroids gave me the opportunity to understand the impact both of the distance between the selected centroids and of the number of centroids on the clustering process. Placing centroids close to each other or having few centroids made clustering cumbersome. On the contrary, placing centroids far from each other or having more centroids made clustering effective and there was no need for continuous reassignment of points and update of centroids.

To sum up, K-means is an easy-to-use clustering approach, and considering the importance of the initial selection of centroids for the performance of the clustering, k-means++ algorithm could be of added value for selecting the appropriate initial values.