Equations

Distance

Use the below equation known as the Euclidean Distance (or the L2 Norm) to find the distance of a point $x^{(i)}$ from a centroid μ_k

$$||x^{(i)} - \mu_k||_2 = \sqrt{\sum_{i=1}^{m} (x_j^i - \mu_{(k,i)})^2}$$

Loss

The loss function is the average of the sum of the squared distances for all m samples $x^{(i)}$ from their assigned centroids $\mu_{c^{(i)}}$

$$\frac{1}{m} \sum_{i=1}^{m} ||x^{(i)} - \mu_{c^{(i)}}||_2^2$$

Resources

Consult the following videos learn more about the above equations and pseudo code for the kmeans algorithm

intro to kmeans: https://www.youtube.com/watch?v=Ev8YbxPu_bQ

pseudo code: https://www.youtube.com/watch?v=hDmNF9JG3lo

loss function: https://www.youtube.com/watch?v=LvgcfMOyREE&t=207s

centroid initialization: https://www.youtube.com/watch?v=PpH_hv55GNQ&t=4s

choosing K: https://www.youtube.com/watch?v=lbR5br5yvrY

stopping criteria: https://stats.stackexchange.com/questions/260917/stopping-condition-of-k-means