

Spherical k-means clustering

Spherical k-means clustering the same idea, but with points on a sphere. We investigated a MATLAB implementation by Nguyen[?, ?], which required a mean-and-norm-normalized dataset located on a hypersphere. Important aspects of this implementation include:

- When there exists an empty cluster, the largest cluster is split
- Use the dot product as “negative distance”, which leverages the fact that observations are unit vectors on the hypersphere
- Use the normalized sum of observations as a centroid/mean, which leverages the fact that observations are unit vectors on the hypersphere. Note that this fails on pathological cases where the sum of observations is zero.