Spherical k-means clustering

Spherical k-means clustering the same idea, but with points on a sphere. We investigated a MATLAB implementation by Nguyen[2, 1], 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.

References

[1] Vinh Nguyen. Matlab central file exchange: The spherical k-means algorithm. *http://www.mathworks.com/matlabcentral/fileexchange/32987-the-spherical-k-means-algorithm*. Accessed: 2016-05-18.

[2] Vinh Nguyen. Gene clustering on the unit hypersphere with the spherical k-means algorithm: coping with extremely large number of local optima. In *World Congress in Computer Science, Computer Engineering, and Applied Computing (Hamid R. Arabnia and Youngsong Mun 14 July 2008 to 17 July 2008)*, pages 226–233. CSREA Press, 2008.