K-Means Clustering

- Represent each text object as a term vector and assume a similarity function defined on two objects
- Start with k randomly selected vectors and assume they are the centroids of k clusters (initial tentative clustering) ➔ **Initialization**
- Assign every vector to a cluster whose centroid is the closest to the vector **≈ E-step difference?**
- Re-compute the centroid for each cluster based on the newly assigned vectors in the cluster **≈ M-step difference?**
- Repeat this process until the similarity-based objective function (i.e., within cluster sum of squares) converges (to a local minimum)

Very similar to clustering with EM for mixture model!