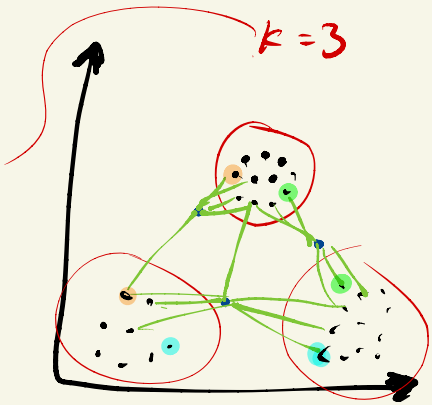
k-means clustering (un-supervised learning)

- computer k cluster from data
- k is pre-defined by # of clusters the user / me!

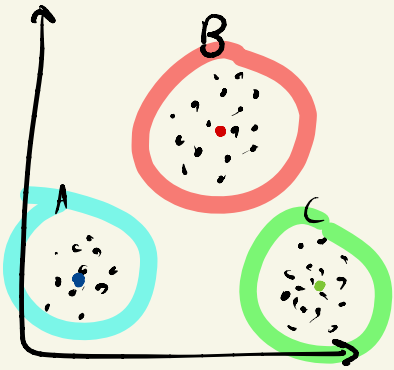


1. Choose 3 points to be centroids of cluster
2. Calculate distance between points and centroid and assign points to closest centroid — expectation step
3. Compute new centroid point and recalculate and redo step 2 (centroid point moves) — Maximization

(thus centroids drift into clusters and achieves lowest distance between all the points, thus finds cluster)

Expectation - Maximization

4.



Principal Component Analysis (PCA)

- dimensionality reduction
- component (direction in space with largest variance)
- the larger the variance, the further spread out the points will likely be
  - > good side effect: is also smallest residuals