Note on GMM, EM & KMeans
(follows after K-Means)

K Means is a variant of EM, with the assumptions that clusters are spherical → spherical ⇒ identical variance-covariance matrices for each cluster.

K Means assumes the clusters are more or less round and solid (not heavily elongated or ringed) clouds in euclidean space, not required to come from normal distributions.

K Means

Hard assignment of data points to one particular cluster for convergence.

uses L_2 norm for optimization

EM

soft assigns a point to a cluster, based on probability of any point belonging to a centroid.

No L_2 norm, Expectation based ie prob (a point in a cluster)

EM → repeatedly expecting the likelihoods & then maximizing the model in 2 steps

- E-step, where each object is assigned to the centroid such that it's the most likely cluster.
- M-step, where the model (centroids) are recomputed by LS optimization

Difference from GMM:

- ✓ doesn't use hard-partitioning
- ✓ in GMM, the models are centroids only, no co-variances/variances are taken into account.