

K-means: Limitations

- Makes **hard assignments** of points to clusters
 - A point either completely belongs to a cluster or not belongs at all
 - No notion of a **soft assignment** (i.e., **probability** of being assigned to each cluster: say $K = 3$ and for some point \mathbf{x}_n , $p_1 = 0.7$, $p_2 = 0.2$, $p_3 = 0.1$)
 - **Gaussian mixture models** and **Fuzzy K-means** allow soft assignments
- Sensitive to **outlier examples** (such examples can affect the mean by a lot)
 - **K-medians** algorithm is a more robust alternative for data with outliers
 - Reason: Median is more robust than mean in presence of outliers
- Works well only for **round shaped**, and of **roughly equal sizes/density clusters**
- Does badly if the clusters have **non-convex shapes**
 - **Spectral clustering** or **kernelized K-means** can be an alternative