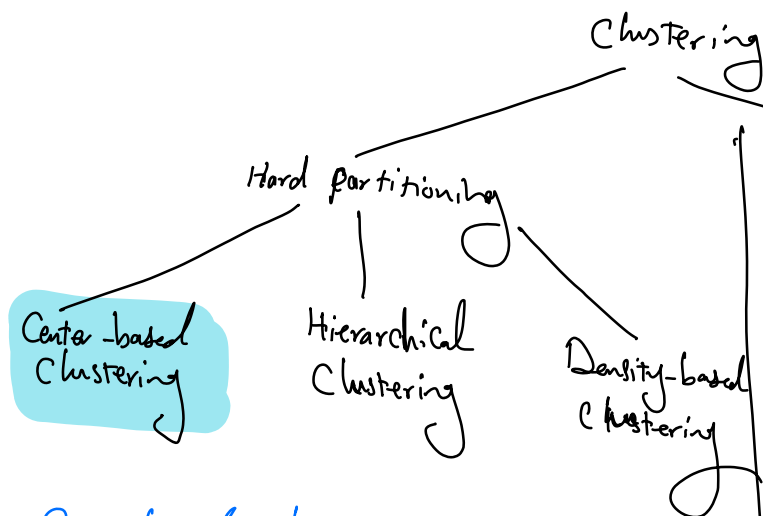


Clustering: Given a Universe of elements $U = \{e_1, \dots, e_n\}$
 a clustering is a partitioning of U into K subsets.



Center-based clusterings:

- K Cluster Centers are selected (based on some distance metrics) & each element is assigned to its nearest cluster center.

$$\delta(e_i, c_j) = \|e_i - c_j\|_{p=2}$$

↪ Euclidean distance

Find $\{c_1, \dots, c_k\}$
 Such that

Min

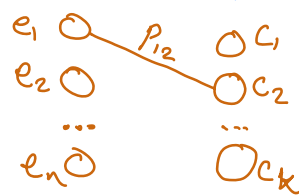
Agg

$$\min_j \delta(e_i, c_j)$$

↪ allocating each element to closest cluster center

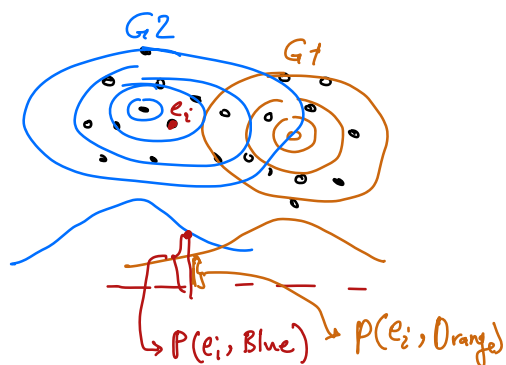
↪ Aggregate function (Sum, Max, ...)

Soft Partitioning
 - The probabilities that an element belongs to each cluster

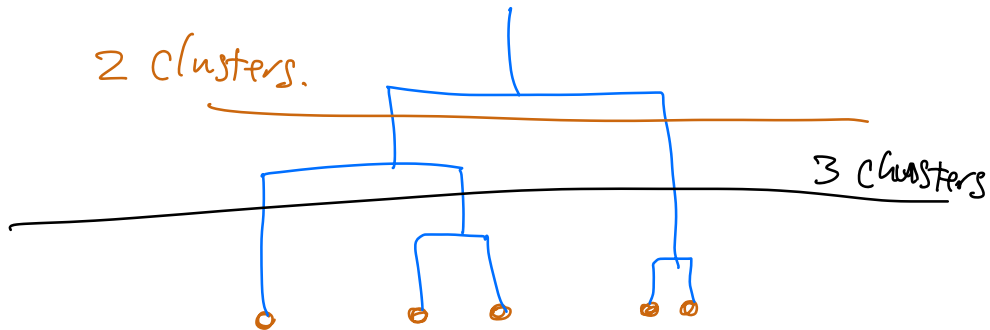


↪ The probability that e_i belongs to c_j

↪ Gaussian Mixture Model (GMM)

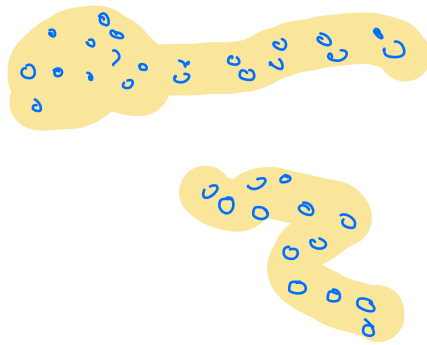


- Hierarchical Clustering: Partition U , Using a Tree data Structure



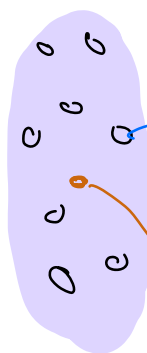
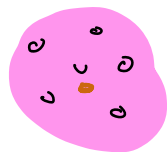
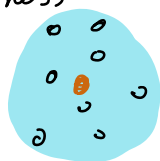
- Density-based Clustering: to identify connected dense regions as clusters.

↳ DBSCAN



Fairness in (Center-based) Clustering

↑
Group Fairness



$e_i \in g_j$

Demographic group

$c_i \in g_k$

- when belonging to a specific cluster has advantage/disadvantage.

→ Demographic Parity (Balance): make sure that the ratios of each group in each cluster is proportional to the original ratios.

$$\frac{|g_i|}{|U|} = \frac{|g_i \cap C_j|}{|C_j|} \quad \forall C_j, g_i$$

→ Relaxed Balance:

$$\left| \frac{|g_i|}{|U|} - \frac{|g_i \cap C_j|}{|C_j|} \right| < \alpha$$

- **Socially Fair Clustering**: Make sure the average distance of each group to their cluster centers is the same

$$\frac{1}{|g_1|} \sum_{l=1}^K \sum_{\substack{e_j \in C_l \\ e_j \in g_1}} \delta(e_j, c_l) = \frac{1}{|g_2|} \sum_{l=1}^K \sum_{\substack{e_j \in C_l \\ e_j \in g_2}} \delta(e_j, c_l)$$

- **Equitable Sharing of Resources**

$$u_j = |\{e_i \in U \mid e_i \in C_j\}|$$

$$\frac{1}{|g_1|} \sum_{e_i \in g_1} u(e_i) = \frac{1}{|g_2|} \sum_{e_i \in g_2} u(e_i)$$

- **Fair Center Selection**: Make sure the representation of each group in the selected centers is proportional to their overall distribution

* Socially Fair K-means.

Liyod Alg.

- Select k Centers arbitrarily
- Repeat

- assign each point to the nearest Center

- Update the Center of each Cluster

Min the Avg distance to the Cluster Center
(Select the mean as the center)

- ~ ~ ~ ~

- Repeat

- ~ ~ ~ ~

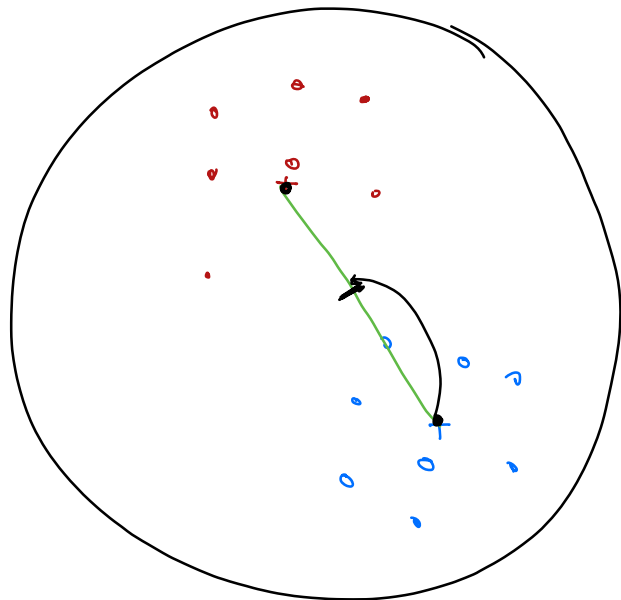
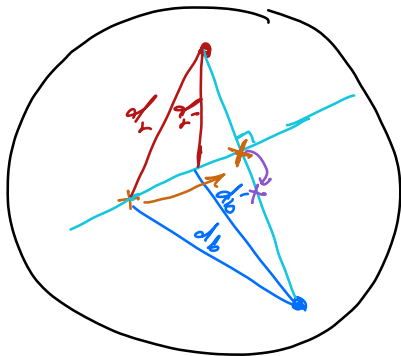
~ ~ ~

- ~ ~ ~ ~

~ ~

- ? ?

?? min the maximum avg-distance of each group to the Cluster Center



Binary-Search for Finding the Fair Cluster-Center

