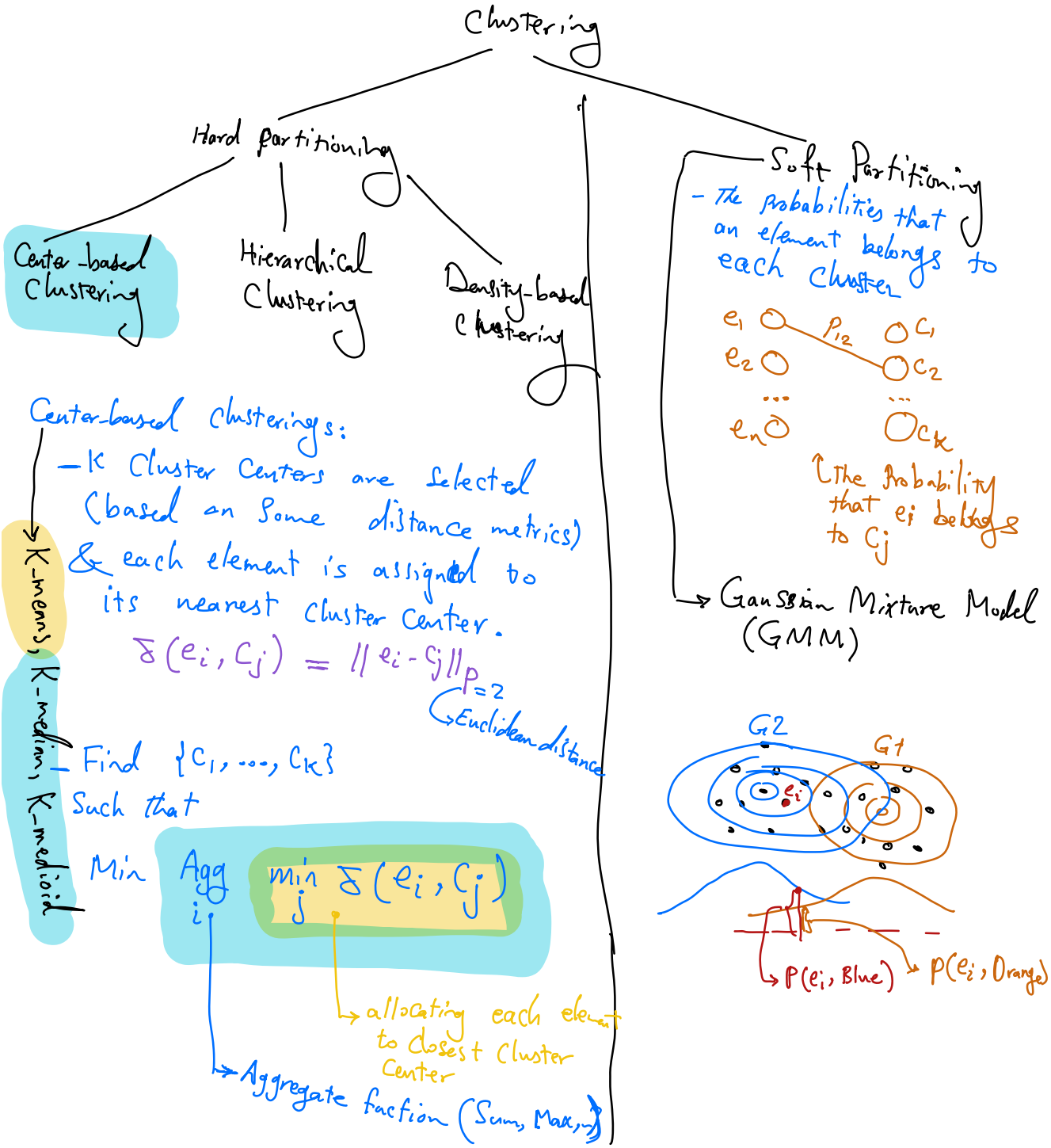
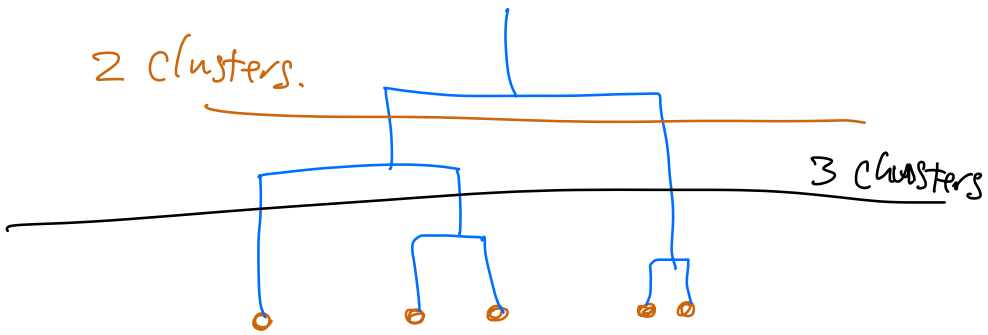
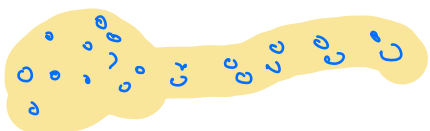
Clustering: Given a Universe of elements $U = \{e_1, \dots, e_n\}$ a Clustering is a partitioning of U into K Subsets.

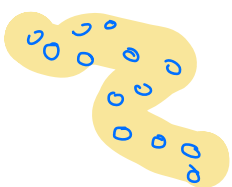


- Hierarchical Clustering: Partition U, Using a Tree data structure



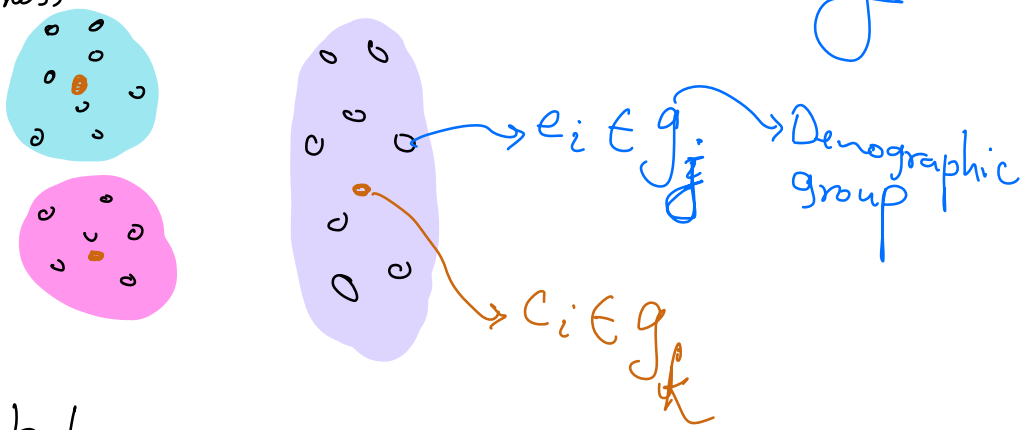
- Density-based clustering: to identify connected dense regions as clusters.
  - ↓ DBSCAN





Fairness in (Center-based) Clustering

↑ Group Fairness



- 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

$$\frac{1}{|g_1|}\sum_{\ell=1}^{K}\sum_{\substack{e_j\in C_\ell \\ e_j\in g_1}}\delta(e_j,c_\ell)=\frac{1}{|g_2|}\sum_{\ell=1}^{K}\sum_{\substack{e_j\in C_\ell \\ e_j\in g_2}}\delta(e_j,c_\ell)$$

- 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.

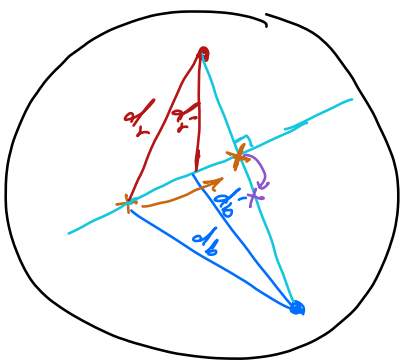
Lloyd Alg.

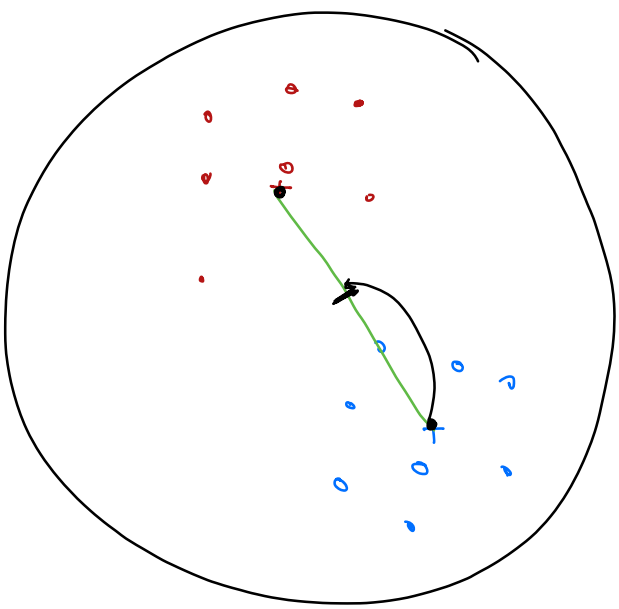
- Select k centers arbitrarily
- Repeat
  - assign each point to the nearest center
  - Update the center of each cluster

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