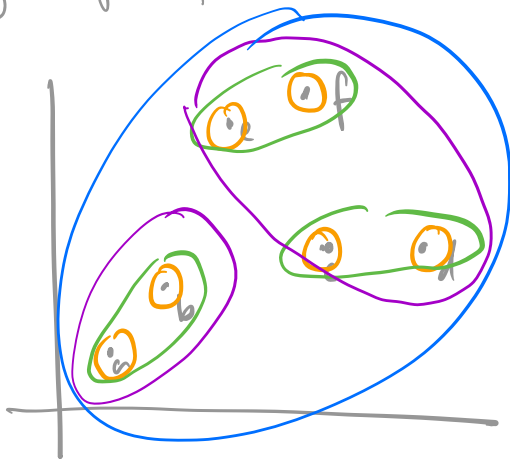


Friday 10/22/21

Clustering and Heat Maps

In the breast data: rows (samples) columns (genes)
we saw a DENDROGRAM: diagram representing clusters. Idea is to identify patterns & groupings within data.

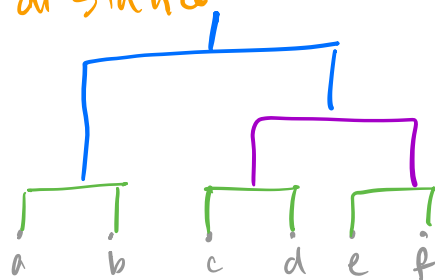


see the height
at which
they are
joined

Agglomerative clustering

(Bottoms up)

Need a notion of
distance



Type of HIERARCHICAL clustering

- Begin with every data point in its own cluster
- Iteratively create clusters
 - Need ① Notion of distance between points
 - ② Notion of distance between groups

For distance b/w points, easiest to use Euclidean distance in \mathbb{R}^n .

$$d_{ij} = |x_i - x_j|$$

$$d_{ij}^2 = |x_i - x_j|^2$$

$$x = (x_1, \dots, x_n)$$