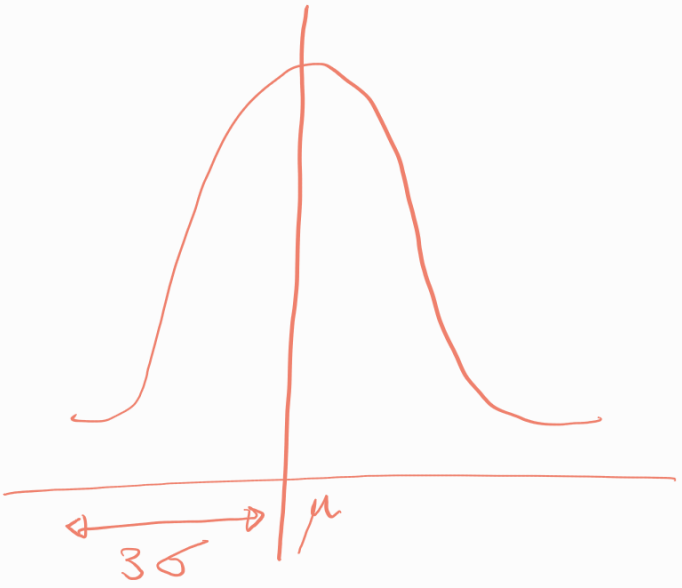
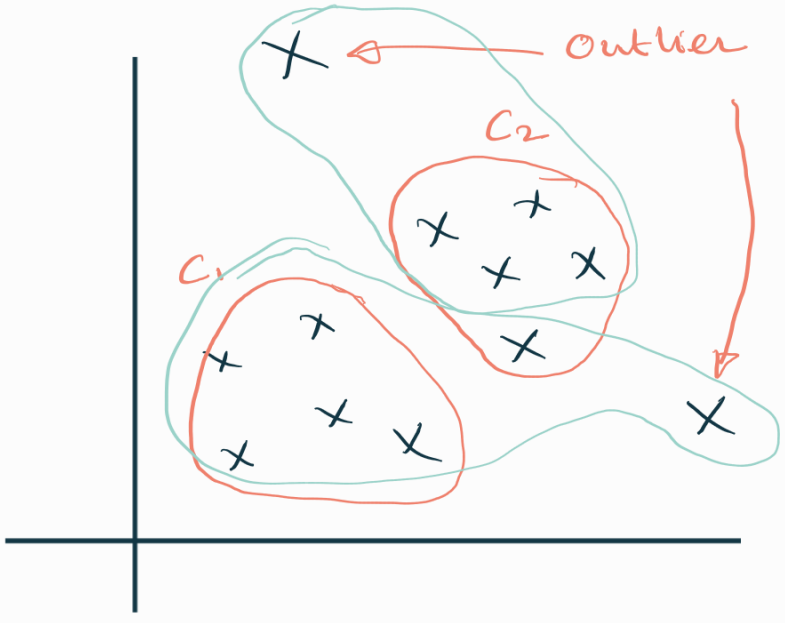
Limitations of K-means Clustering

- Number of clusters are fixed in advance
- Outlier can lead to bad clusters





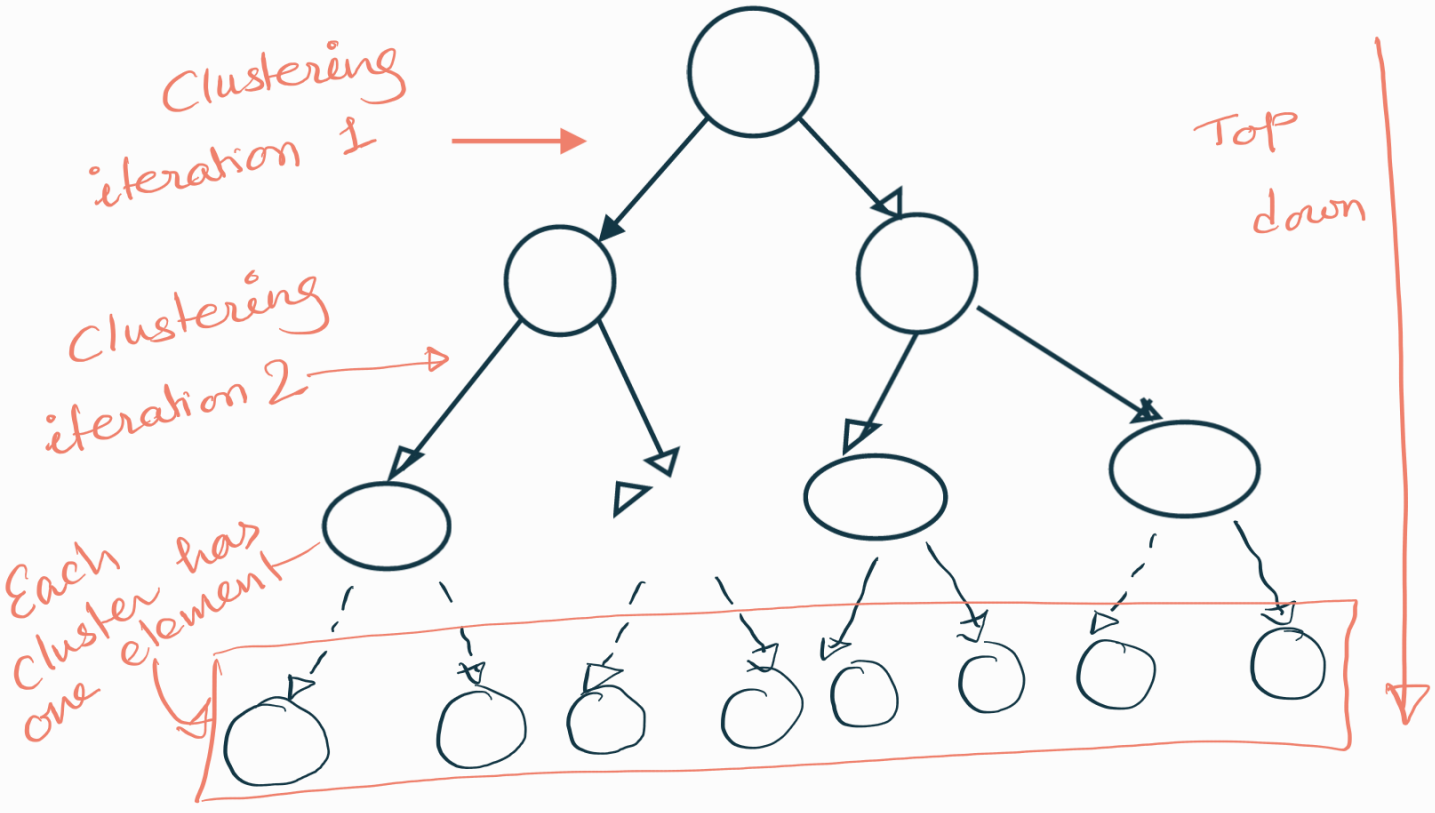
K-Medoid cluster - To handle issue of outliers.

Hierarchical Clustering

- Do not fix the number of clusters in advance
- Instead, stepwise merge or divide data into different clusters.
- Hierarchical Divisive Clustering : (Top-down)
  - Entire dataset is one cluster at the beginning
  - partition the data into different clusters

in each iteration.

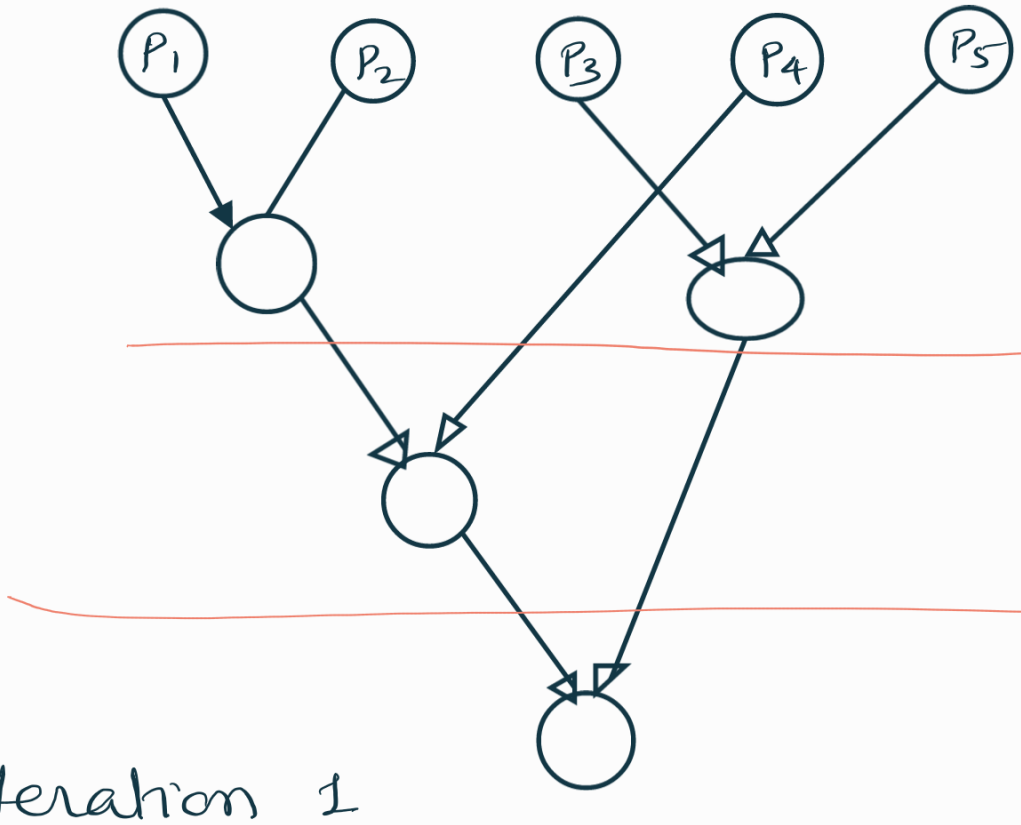
- partitioning is done using other clustering algorithm e.g K-means.



Hierarchical Agglomerative Clustering

(Bottom-Up)

- Each datapoint is a cluster at the beginning of the algorithm.
- In every iteration merge nearest/most similar clusters.

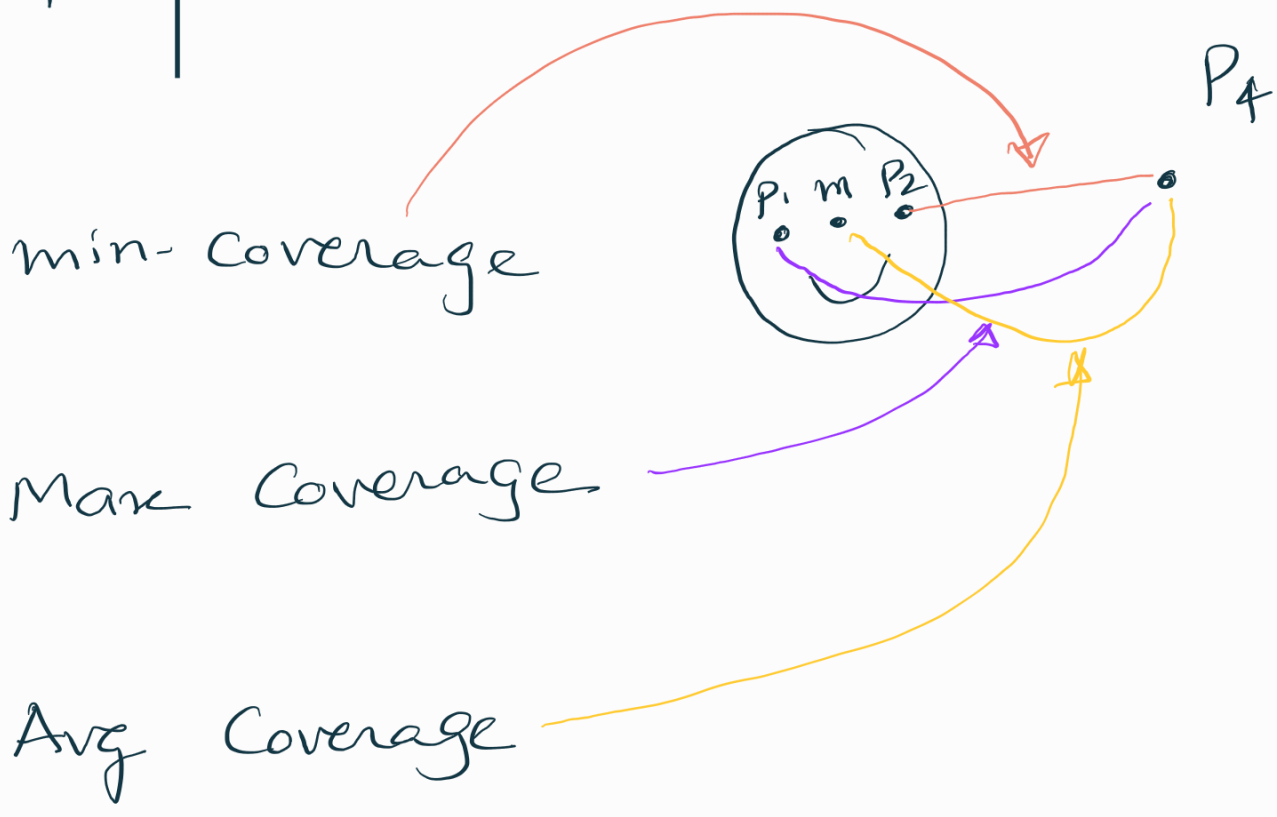


Iteration 1

	P_1	P_2	P_3	P_4	P_5
P_1					
P_2	✓				
P_3					
P_4					
P_5			✓		

	(P_1, P_2)	(P_3, P_5)	P_4
(P_1, P_2)		✓	✓
(P_3, P_5)			
P_4			

$$d((P_1, P_2), (P_3, P_5)) < d((P_1, P_2), P_4)$$

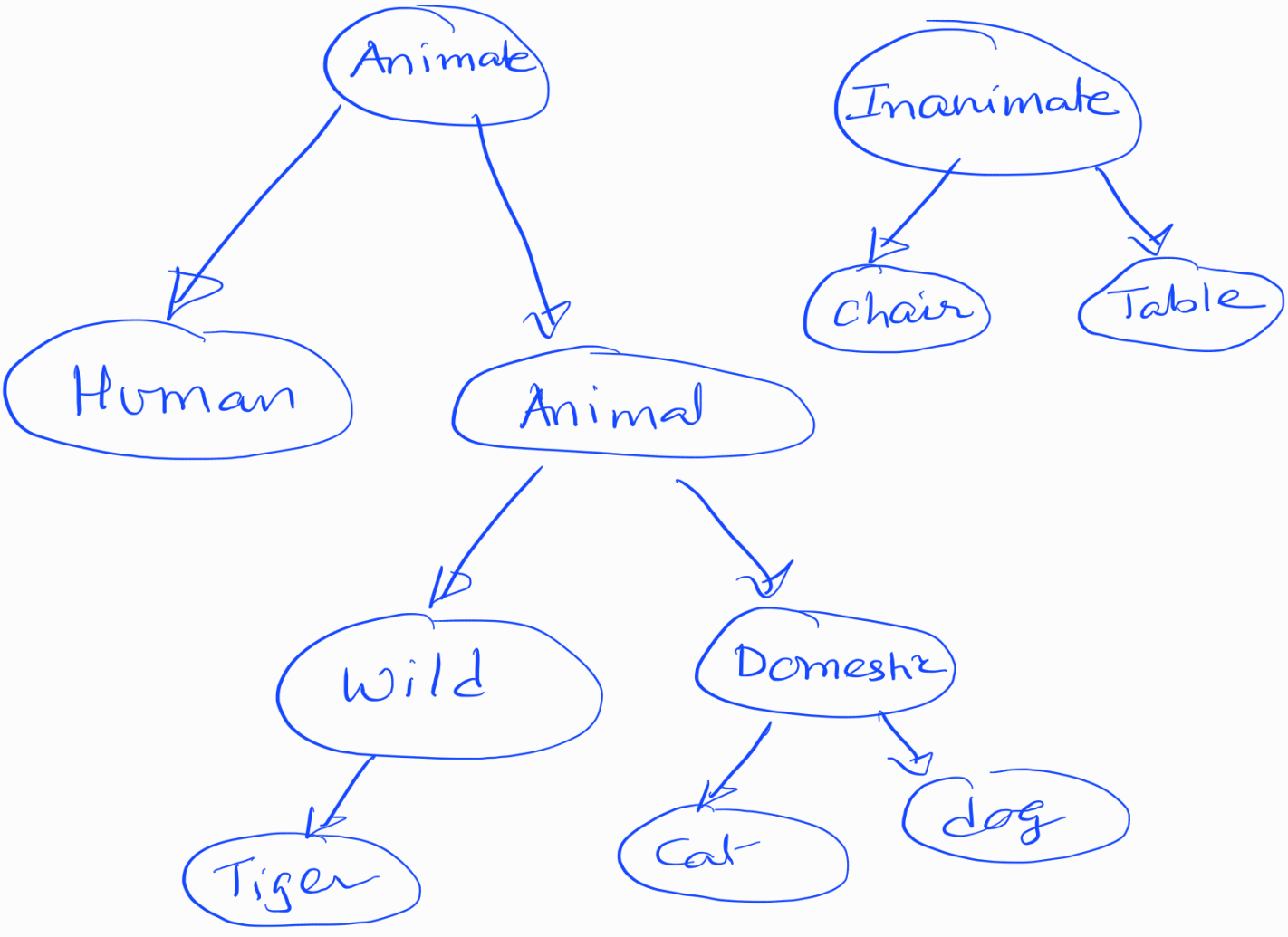


---

* Find Semantic Similarity between words.

Cat, dog, table

Taxonomy



Object	# legs	# tail	# speak
Cat	4	1	Yes
dog	4	1	Yes
Table	4	0	No