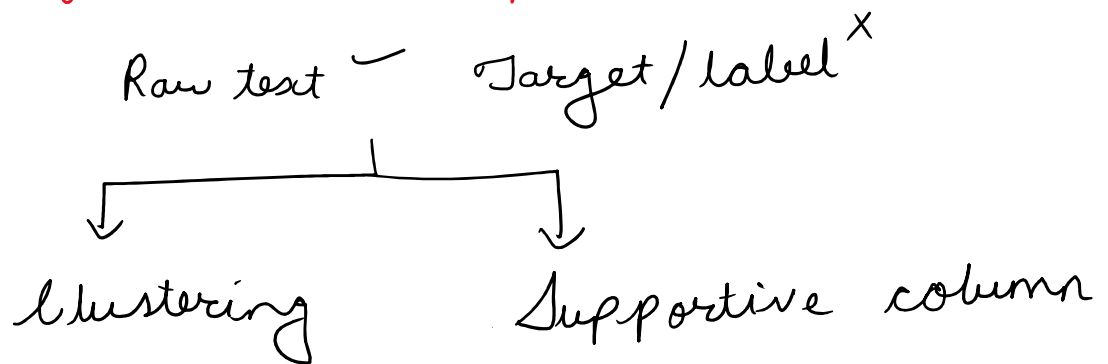
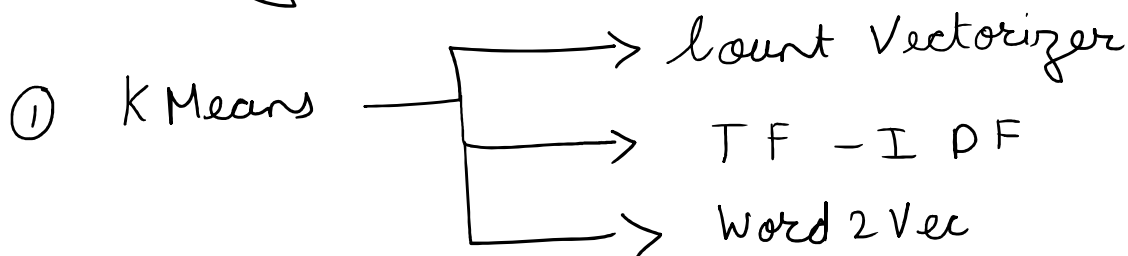


What is unlabelled data?



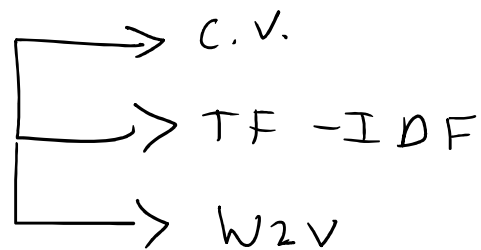
① clustering



② Hierarchical clustering

③ DB Scan

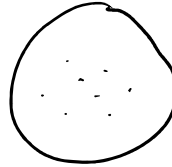
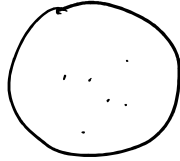
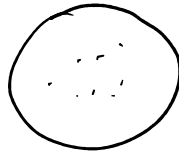
④ SOM



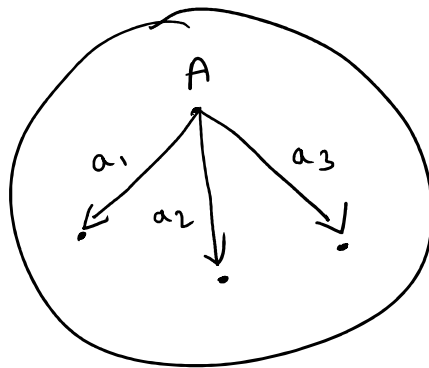
evaluation of clustering:

Silhouette score \Rightarrow To evaluate clustering

$$\text{Silhouette coefficient} = \frac{b-a}{\max(a, b)}$$

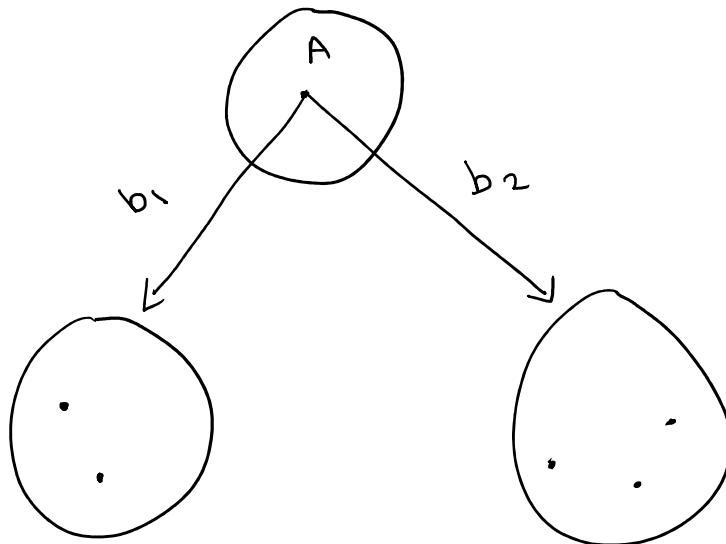


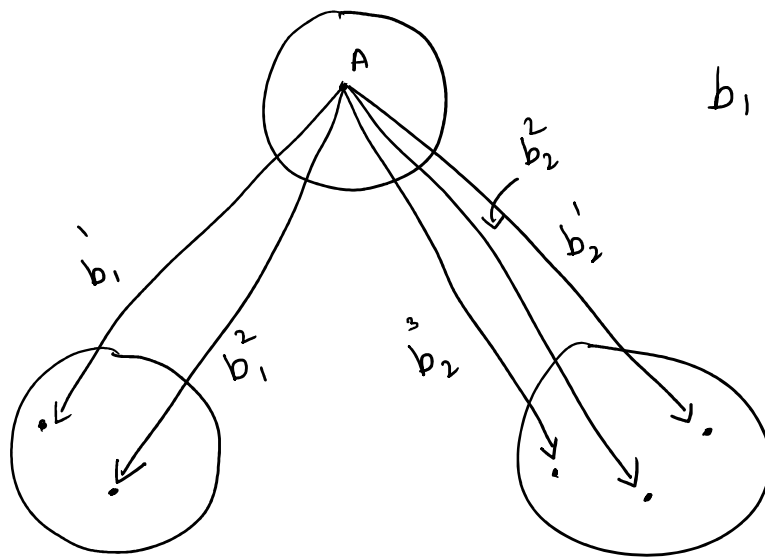
what is a ?



$$a = \frac{a_1 + a_2 + a_3}{3}$$

what is b ?





$$b_1 = \frac{b_1^1 + b_1^2}{2}$$

$$b_2 = \frac{b_2^1 + b_2^2 + b_2^3}{3}$$

$$b = \min(b_1, b_2)$$

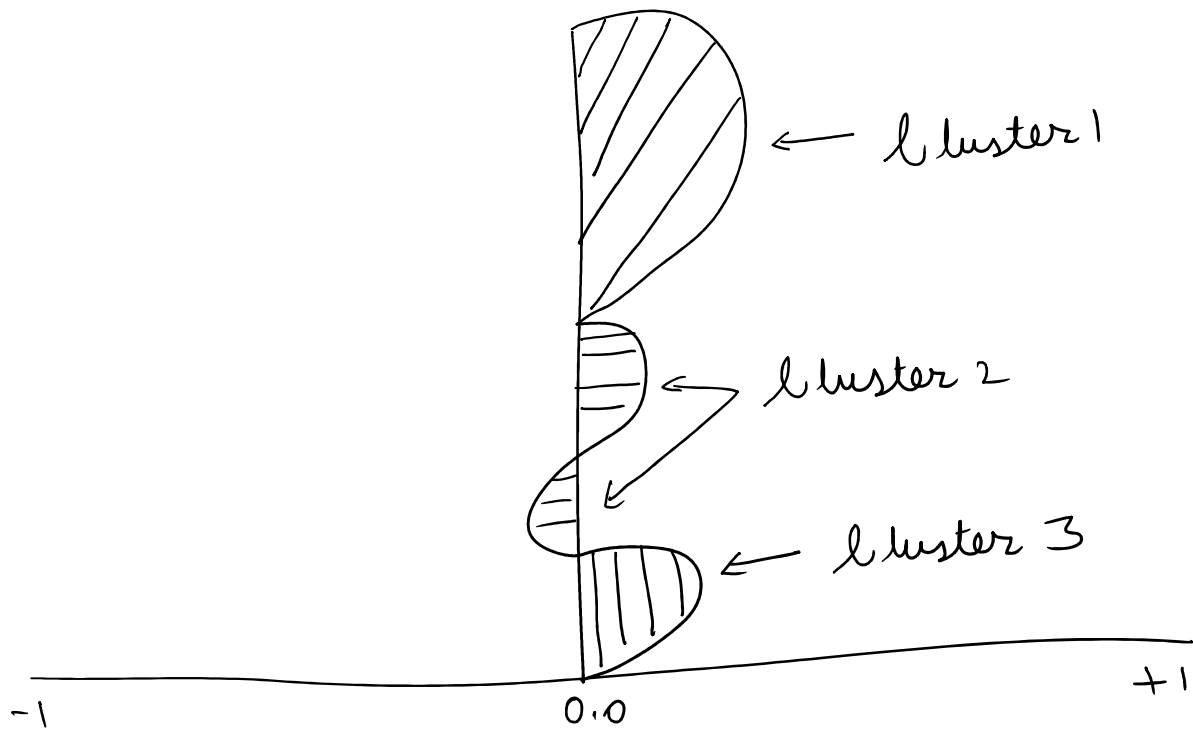
Silhouette score \Rightarrow Average of all the silhouette coefficients

-1 to +1
 \downarrow \downarrow

clusters are
not formed
properly

clusters are
formed
properly

Silhouette visualizer \rightarrow It visualizes distribution of silhouette coefficients



② Supportive column

Stars / Rating column (1-10)	Target / label
1 - 3	Negative Review
4 - 6	Neutral Review
7 - 10	Positive Review