K- Means

+ve −ve ⇒ 2 grps

+ve −ve Neutral ⇒ 3 grps

Objective fn

$$\text{Wcss} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_i - c_j)^2$$

Within cluster Sum of Squares

→ Minimized.

Task :

Find the centroids (C_ks) which minimize the Wcss.

— NP Hard problem.

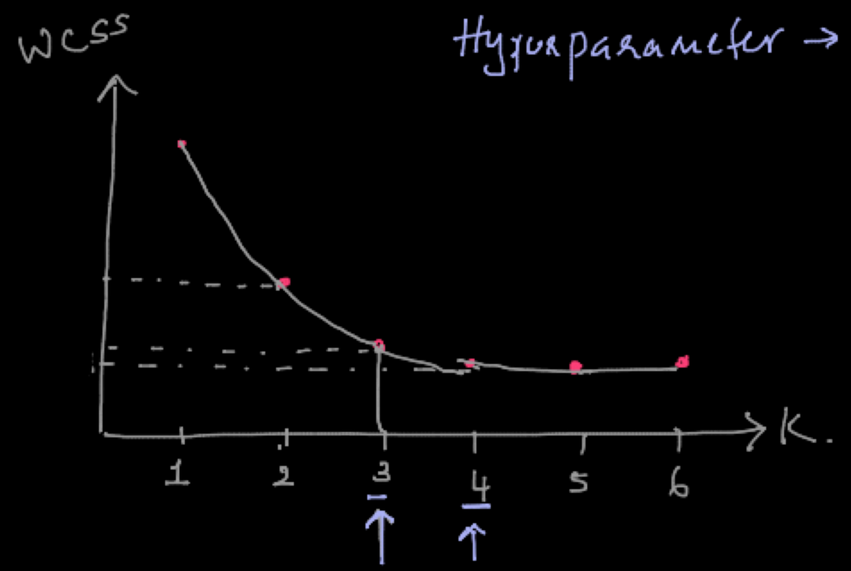
Lloyd's Approximation

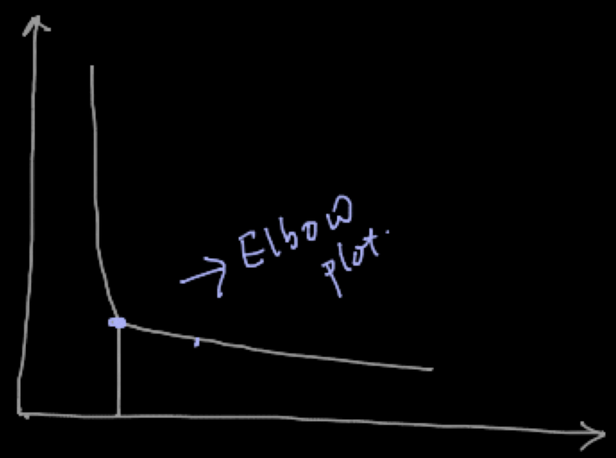
→ Iterative process

→ Repeat until centroids stop moving.

Elbow plot (Scree plot, Knee plot).

Hyperparameter → tuning



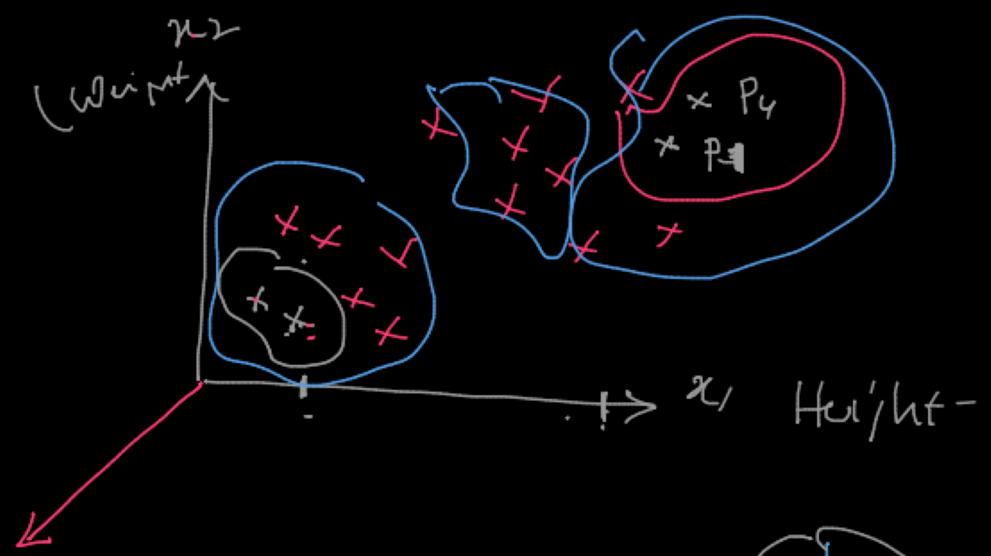


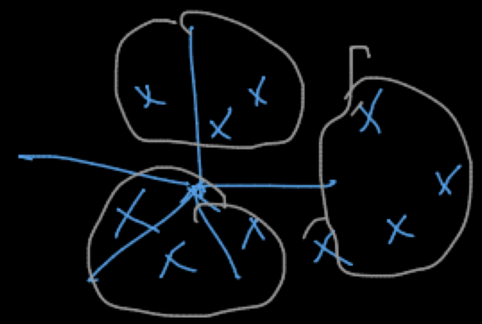
GridSearch →

K - Means.

↳ Hyperparam.

	x_1 Height	x_2 Weight	$w_i x_i$	
P_1	190	95	0	←
P_2	130	50	1	
P_3	125	45	1	
P_4	200	100	0	





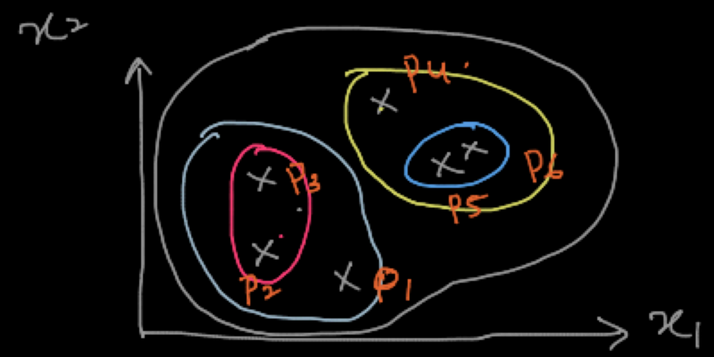
x_1	x_2	x_3
SAT	Top 10	SFF

HAC → Agglomerative

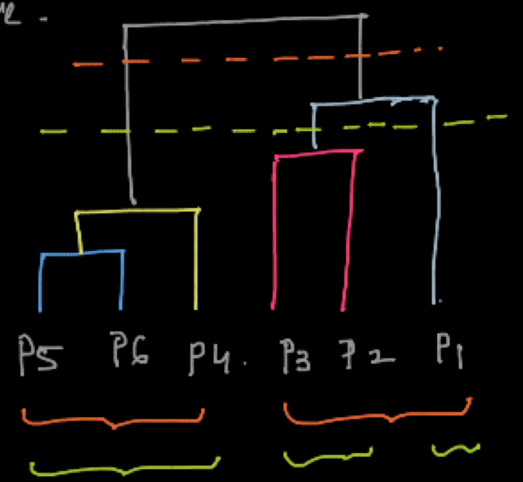
Hierarchical Clustering.

— Agglomerative.

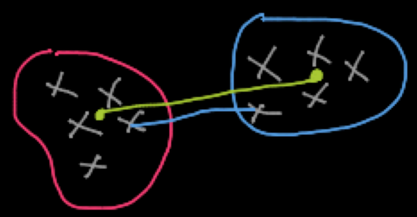
— Divisive.



→ Dendrogram



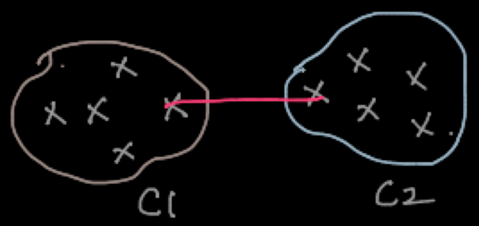
Find the distance between cluster → Linkage



1. Single Linkage.
2. Complete linkage.
3. Centroid linkage
4. Avg linkage.

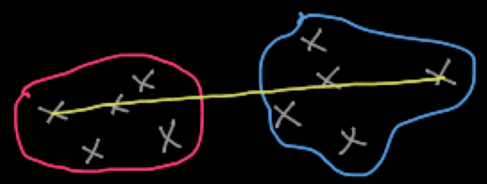
Linkages.

Single linkage



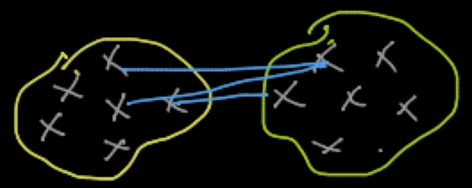
— distance both the closest points in two clusters

Complete linkage.



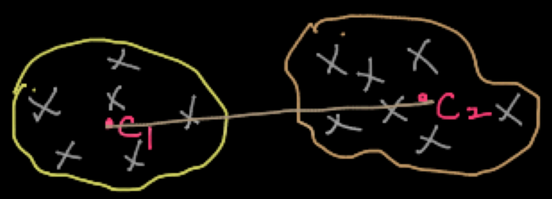
— distance between the farthest points in two clusters

Avg linkage.



— Average of the distances between points in two clusters.

Centroid linkage.



— distance between the centroids of two clusters.