Page :
Date: / /

Unit-3

Types of clustering are :-

- Partitioning Clustering
- Density-Based Clustering
- Distribution Model-Based "
- Hierarchical "
- Fuzzy "

Partitioning →

⇒ K-Means :- It groups the unlabeled dataset into different clusters. Here K defines the number of pre-defined clusters that need to be created in the process. It is a centroid based algo.

$$\text{Euclidean Distance} = \sqrt{(X_0 - X_c)^2 + (Y_0 - Y_c)^2}$$

(ED is used to solve k-means with help of centroids and the lowest value (new data) in b/w clusters will go there)

⇒ K-Medoid :- Same as K-Means but after dividing the data into two (if K=2) it calculates total cost and then again select a new cluster then K-Means is repeated then total cost is calculated. In step 3 both total cost are subtracted and In case of negative value we again have to ~~to~~ repeat K-Means with a new value. (Final value should be greater than 0).

Hierarchical →

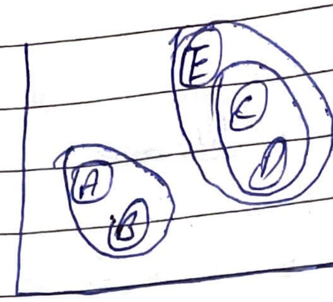
⇒ Heirarchial :- We ~~deta~~ develop the hierarchy of cluster in the form of a tree and this tree-shaped structure is known as ~~dendogram~~ dendrogram.

Approaches :-

Date: / /

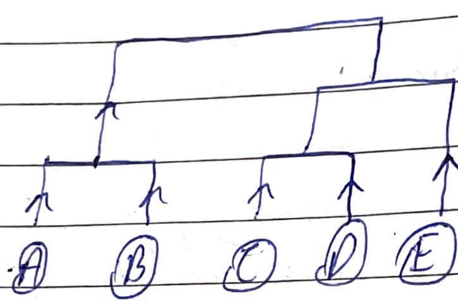
(Model paper-1 2 marks)

1) **Agglomerative** :- It is a bottom-up approach, in which the algorithm starts with taking all data points as single clusters and merging them until one cluster is left. Eg :-



Circle represent a cluster.

Dendrogram :-



↑ - Bottom-up approach.

2) **Divisive** :- Reverse of agglomerative as it is a top-down approach.

Imp

⇒ **Principle Component Analysis (PCA)** :- Used to reduce Overfitting. It is a statistical process that converts the observations of correlated features into a set of linearly uncorrelated features with the help of orthogonal transformation. PCA generally tries to find the lower-dimensional surface to project the high-dimensional data.

⇒ Association Rule learning :- It checks for the dependency of one data item on another data item and maps accordingly so that it can be more profitable. It tries to find some interesting relations among the variable

Apriori Algorithm :- This algo uses frequent datasets to generate association rules. It is designed to work on the databases that contain transaction.

Support :- It is frequency of A or how frequently an item appear in the dataset.

$$Support = \frac{Freq(X)}{T}$$

Confidence :- It indicates how often the rule has been found to be true.

Lift :- It is the strength of any rule.

$$Lift = \frac{Support(X,Y)}{S(X) \times S(Y)}$$

Date: / /

FCFS - (Non-Preem)

	At	Bt	Ct	TAT	WT	RT
P_1	0	2	2	2	0	0
P_2	1	2	4	3	1	1
P_3	5	3	8	3	0	0
P_4	6	4	12	6	2	2

P_1	P_2		P_3	P_4

0 2 4 5 8 12

TAT = CT − At

Wt = TAT − Bt

Rt = CPU first time − At

SJF (Non-Preem)

	At	Bt	Ct	TAT	Wt	Rt
P_1	1	3	6	5	2	2
P_2	2	4	10	8	4	4
P_3	1	2	3	2	0	0
P_4	4	4	14	10	6	6

	P_3	P_1	P_2	P_4

0 1 3 6 10 14

RR - (Preemp)

Time Quantum = 2

	At	Bt	Ct	TAT	Wt	Rt
P_1	0	5	12	12	7	0
P_2	1	4	11	10	6	1
P_3	2	2	6	4	2	2
P_4	4	1	9	5	4	4

Ready Queue -

~~P_1~~	~~P_2~~	~~P_3~~	~~P_1~~	~~P_4~~	~~P_2~~	P_1	

Running Queue -

P1	P2	P3	P1	P4	P2	P1

0 2 4 6 8 10 11 12

$\longrightarrow$

Context Switching.

Priority Scheduling - (Preempt)

Priority		At	Bt	Ct	TAT	Wt
10	P_1	0	5	12	12	7
20	P_2	1	4	8	7	3
30	P_3	2	2	4	2	0
40	P_4	4	1	5	1	0

P1	P2	P3	P3	P4	P2	P1

0 1 2 3 4 5 8 12

Bankers Algo

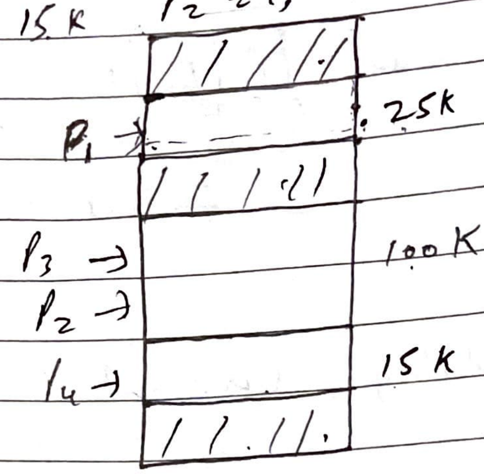
Total A = 10, B = 5, C = 7

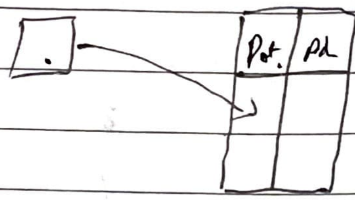
	Allocation A B C	Max Need A B C	Availble A B C	Remaining Need A B C	
P_1	0 1 0	7 5 3	3 3 2	7 4 3	— 4
P_2	2 0 0	3 2 2	5 3 2	1 2 2	— 1
P_3	3 0 2	9 0 2	7 4 3	6 0 0	— 5
P_4	2 1 1	4 2 2	7 4 5	2 1 1	— 2
P_5	0 0 2	5 3 3	7 5 5	5 3 1	— 3
	7 2 5		10 5 7		

Sequ…

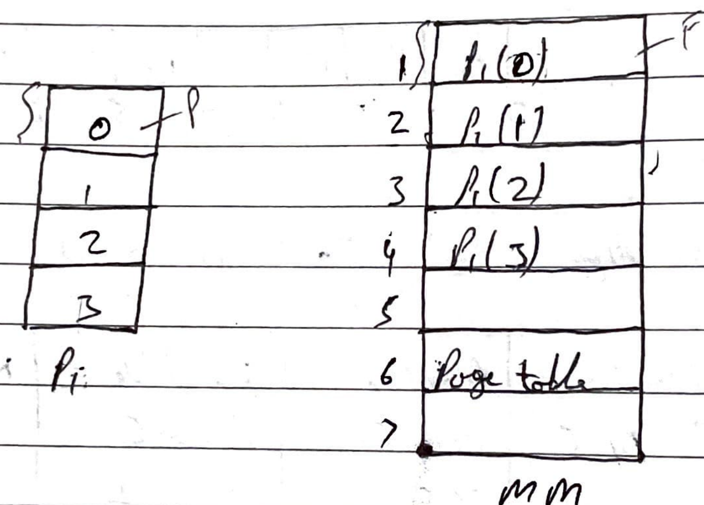
$P_1 = 15K$ $P_2 = 15$ $P_3 = 15$ $P_4 = 15$

P_4 Best fit
P_3 worst fit
P_2 Next fit
P_1 First fit.





Paging



0	Frame 1
1	Frame 2
2	Frame 3
3	Frame 4

Page table

Frame 6

Page Register