1 Clustering ?

The process of grouping a set of Physical or abstract objects into classes of similar objects is called Clustering.

2 Define Cluster ?

A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in one clusters.

3 Clustering techniques

1) Partitioning clustering.
2) Density Based Clustering
3) Distribution model based clustering
4) Hierarchial clustering
5) Fuzzy clustering.

4 Applications of clustering.

1) amazon
2) Netflix
3) Social N/w analysis.
4) Statistical data analysis
5) Image segmentation.

Unit - 3 2m.

1 Clustering ?

The process of grouping a set of Physical or abstract objects into classes of similar objects is called Clustering.

2 Define Cluster ?

A Cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in one clusters.

3 Clustering techniques

1) Partitioning clustering.
2) Density Based Clustering
3) Distribution model based clustering
4) Hierarchial clustering
5) Fuzzy clustering.

4 Applications of clustering.

1) amazon
2) Netflix
3) Social N/w analysis.
4) Statistical data analysis.
5) Image segmentation.

5 3 distance measures.

1) minkowski

$$d(i,j) = \left(|x_{i_1} - x_{j_1}|^p + |x_{i_2} - x_{j_2}|^p + \cdots |x_{i_n} - x_{j_n}|^p\right)$$

2) Manhattan.

$p = 1$

$$d(i,j) = |x_{i_1} - x_{j_1}| + |x_{i_2} - x_{j_2}| + \cdots |x_{i_n} - x_{j_n}|$$

3) Euclidean

$$d(i,j) = \sqrt{(x_{i_1} - x_{j_1})^2 + (x_{i_2} - x_{j_2})^2 + \cdots (x_{i_2} - x_{j_2})^2}$$

6 Single linkage clustering.

For two clusters R and S, the single linkage returns the minimum distance b/w two points i and j such that i belongs to R and j belongs to S.

$$L(R,S) = \min (dis(i,j)), i \in R, j \in S$$

7 Complete linkage.

For two clusters R and S, the complete linkage returns the maximum distance b/w two points i and j such that i belongs to R and j belongs to S.

$$L(R,S) = \max (dis(i,j)), i \in R, j \in S.$$

8 Average linkage clustering

For two clusters R and S, first for the distance b/w any data point i in R and any datapoint j in S and then the arithmetic mean of these distances are calculated. Average linkage returns the value of arithmetic mean.

$$L(R,S) = \frac{1}{n_R + n_S} \sum_{i=1}^{n_R} \sum_{j=1}^{n_S} D(i,j)$$

where

n_R → no of datapoints in R

n_S → no of datapoints in S.

9 Dendogram

We develop the hierarchy of clusters in the form of tree and tree shaped structure is called dendogram.

10 Formulas of Decision Tree Induction.

$$Info(D) = -\sum_{i=1}^{c} P_i \log_2 P_i$$

$$Gain_x(D) = Info(D) - Info_x(D).$$

11 KNN algorithm

1) KNN stands K - nearest neighbour

2) It stores all the available data and classifies a new data point based on the similarity.

Unit-4. 2m

1 Drawback of Single layer perceptron.
It cannot support XOR gate properly.

2 Which are gates that are supported by single layer perceptron.
AND, OR, NOT.

3 Multilayer perceptron
Problem can be solved by XOR gate.

4 Activation function.

Range

1) Step function $f(z) = \begin{cases} 0 & z < 0 \\ 1 & z \geq 0 \end{cases}$ $\{0, 1\}$.

2) Signum function $f(z) = \begin{cases} -1 & z < 0 \\ 0 & z = 0 \\ 1 & z > 1 \end{cases}$ $\{-1, 1\}$

3) Linear function $f(z) = x$ $(-\infty, \infty)$

4) Relu function $f(z) = \begin{cases} 0 & z < 0 \\ z & z \geq 0 \end{cases}$ $(0, \infty)$

5) Sigmoid function $f(z) = \frac{1}{1+e^{x}}$ $(0, 1)$

6) Hyperbolic tan $\tanh(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$ $(-1, 1)$

5 ANN

1) It stands for Artificial Neural Network.

2) It is a Computational n/w based on biological neural n/w that construct the structure of human brain.

6 3 types of nervous system.

1) Sympathetic nervous system.
2) The Parasympathetic nervous system.
3) The enteric nervous system.