Sr no :- 1620BECE300 80

Page :- 1

IR

Course :- BE-CE

Semester :- 8th

Date _____

Page _____

(1) The agglomerative clustering is the most common type of hierarchical clustering used to group objects in clusters based on their similarities. It is also known as AGNES. This algorithm start by treating each object as a singleton cluster. Next pair of cluster are successively mearged untill all cluster have been merged into one big cluster containing all objects. The result is a tree - based representation of the object named dendogram.

- Algglomerative clustering works in a bottom - up manner. That is, each object is initially considered as a single - element cluster (leaf). At each step of the algorithm the two cluster that are the most similar are com -bined into a bigger cluster (nodes). This procedure is iterated until all points are member of just one single big cluster (root)

- similarities between the cluster are measured using matrix generation

(3)   KNN (K-nearset neighbour) algorithm
is an type of supervised learning
algorithm. It is implemented using
vector space modat.

- KNN classifier main ftet function
is to find that a new data point
in a vector space is belonging
to which contengency region or
which class

- Steps used by KNN classifier are

(i) For implementing any algorithm
we need data set. So during the
first step of KNN we must load
the training as well as test data

(ii) Next, we need to choose the
value of k :- k, nearest neighbour

(iii) For each data points,

    (a) calulate the euclidian distance
        between each data point
        and new point.

(iv) find k nearset points to the
    test point.

(v) Assign class to test point based on majority class of k points

(5)   **HITs Algorithm :-**

Hyper-link - Induced Topic search (also known as hubs and authorities) is a link analysis algorithm that rates web pages. It determines two values for a page :-

(i) Its authority, which estimates the value of the content of the page,⊕

(ii) Its hub value, which estimates the value of its links to other pages.

=)   Steps involved in HITs algorithm

(i)   Starting from the user supplied query, HITs assembles intial sets of pages :-

The intial set of pages is called root set. These pages are then expanded to a larger root set T by adding any pages that are linked to or from any page in the intial sets.

(ii)

(2) HITs then associates with each page
p a hub weight $h(p)$ and an autho
-rity weight $a(p)$, all intialized to 1.

(iii)

(3) HITs then iteratively updates the hubs
and authoritiy weight of each page
let $p \to q$ denotes " page p has an
hyper link to page q ". HITS updates
the hubs and authorities as follow

$$a(p) = \sum_{p \to q} h(q)$$

$$h(p) = \sum_{q \to p} a(q)$$