## A BRIEF REPORT ON CLUSTERING , FEATURE SELECTION / FEATURE REDUCTION , MULTIOBJECTIVE OPTIMIZATION AND VALIDITY INDICES

Clustering is the process of placing similar objects i.e. objects having the same properties in one group / same group. These properties are also referred to using the terms such as attributes or dimensions. Moreover the groups into which the objects are segregated are called clusters and this entire process of separating the objects into groups of similar properties is called clustering .The process of clustering is of great usefulness because through the process of clustering the data, we can obtain the data distribution, observe the characteristics of each cluster and make further studies on any particular clusters. From all the above mentioned information , it is also clear that clustering will be much better when there is greater similarity within a group i.e. when the objects one particular group are are very much similar to each other and greater the difference amongst the groups.

## Furthermore , there are various other classifications of clusters , which is generally based on the way clustering is performed. Some common ones include :

## 🡪 A well-separated cluster in which a cluster is a set of points such that any point that is in a cluster is closer (or more similar) to every other point in the cluster than to any point which is not in the cluster . Another one is a

## 🡪Centre-based cluster a cluster is a set of objects such that an object in a cluster is closer (more similar) to the “centre” of a cluster, than to the centre of any other cluster. The centre of a cluster is often a the average of all the points in the cluster or the “most representative” point of a cluster.

The process of clustering follows a simple process which includes gathering the raw data and applying any clustering algorithm to it which in turn converts it into clusters.

Very often , it is the case that in high dimensional data i.e. the data containing a large number of attributes or features , many features or attributes are irrelevant and not required . That means redundant or irrelevant features are also there to represent a data and which can cause hindrance and inconvenience and can mask existing clusters in noisy data. This redundant data degrades the cluster quality and it becomes very important to get rid of them .Here comes the vital use of Feature selection and Feature reduction that identifies and removes irrelevant and redundant dimensions by analysing the entire dataset.

As far as multiobjective optimization is considered , generally we tend to represent a cluster in a typical form or shape , which generally is a circular shape and outlines. Here the objectives are compactness and separation . Compactness as in , within a cluster the objects must be closely and compactly placed and spaced with respect to the cluster centre and separation which is the distance between each cluster should be more . Hence our primary aim revolves around the fact that we must minimize the compactness and maximize the separation as much as possible.

Validity Indices

When we talk about clusters so extensively , there must also be a way to determine the quality of the clusters or any of its parameters so as to facilitate further study . In this respect , various validity indices have been proposed evaluate the correctness and goodness of a clustering structure quantitatively . Moreover there are broadly two criteria for evaluation and selection of the optimal clustering from a given set of results of a clustering algorithm: Compactness: Members of each cluster should be as close to each other as possible. Separation: The clusters should be widely spaced from each other. A good clustering obtained using a clustering algorithm should have both high compactness and high separation.

There are a number of validity indices , in this regard , to evaluate the goodness of clustering structure. Two popular ones among them are Pakhira Bandyopadhyay Maulik (PBM) Index and Xie-Beni (XB) Index . The Xie - Beni index focuses on the compactness and separation of clusters Both the methods have predefined formulae which on application gives the goodness of the clustering structures.

Thanking you , Sir ,

Regards,

Ayush Anand

2nd Year Undergraduate

Indian Institute of Engineering Science and Technology , Shibpur