Introduction to K-Nearest Neighbors.

The Idea in KNN methods is to identify 'K' samples in the dataset that are similar or close in the space. Then we use the 'K' samples to estimate the value of the missing data points. Each Samples missing values are Imputed using the mean value of the 'K' neighbors found in the data set.

DISTANCE CALCULATION IN THE PRESENCE OF MISSING VALUE:-

Missing Value in data are Common in real World application. Since the performance of data mining algorithm depends critically on it being given a good metric over the input space. We decide to define a distance function for unlabeled dataset with missing Values. According to this distance the distance between two points without missing attributes value is simply the Mahalanobis distance. Our distance in general can be used as a part of any algorithm that Computes the distance between data point. On these data sets We stimulated missing values and Compared the performne of the KNN Classifier using our distance to