DAY-6(FEB-17) U.Hasini

22A81A0658

SOC\_2

COURSE:ADVANCED PYTHON

**NOTES:TOPICS COVERED**

#Unsupervised Learning Algorithm

#**k means clustering:**

* It is an unsupervised learning algorithm that will attempt to group similar clusters together from your data.
* It is mainly used in clustering similar documents
* clustering customers based on similar features.

#from sklearn.datasets import make\_blobs

* + The make\_blobs function is used to generate synthetic datasets
  + for clustering and classification tasks
  + This function will create clusters of data points with
  + Gaussian distribution

#**creating random dataset:**

data=make\_blobs(n\_samples,n\_features,centers,cluster\_std,random\_state)

* + n\_samples=Total number of points equally divided among clusters.
  + n\_features=it indicates the no of features(columns)
  + centres=it determines number of clusters to be generated
  + cluster\_std=it sets the standers deviation of the clusters.

#High value makes the clusters to spread out

#Total no of null or NaN FIELD

#Convert all col datatype to string, to apply StandardScaler()

#from sklearn.preprocessing import StandardScaler

**#StandardScaler**:

* + It is a preprocessing class that is used to standardise.
  + It normalize the features of dataset.
  + It scales each feature in such way that it has a mean of 0 and std of 11

#Difference between KNN and k means clusterring

1)#KNN is used for classification and regression

#K Means is used for clustering problems

2)#KNN is supervised algorithm

#K Means is unsupervised algorithm

3)#To traing KNN,we need a dataset with all the datpoints having class labels

for traninig k means,we no need any such information

4)#We use Knn TO PREDICT THE CLASS LABEL OR new points

#We use K means to find patterns in a given dataset by grouping data points

into clusters

#Example of standard scalar

#To find the Error Rate