

README

- For running the code, in case the path is not set , run the following command on shell
`$ export HADOOP_CLASSPATH=$(find $HADOOP_HOME -name '*.jar' | xargs echo | tr ' ' ':')`
- Compile the code using *javac* commands
- Make jar files for the code

Arguments required-

- Arguments for KNN Hadoop-
No. of arguments=3
Input file(training data) , output file and test file
- Arguments for KNN Spark-
No. of arguments =5
k(no of neighbors), d(no. Of attributes), testing data file,training data file,output file
- Arguments for Bisecting k Means Hadoop-
Output file, dataset file(dataset.txt),centroids file(centres.txt)
- Arguments for Bisecting k Means Spark-
No of arguments=4
k(no of clusters required),d(no of attributes),dataset file,output file

Note- The code has been written for space separated dataset.